



**HAL**  
open science

# Abord de l'épidémiologie des cancers par des données de registre et des données médico-administratives : complémentarité des deux approches

Julien Simon

## ► To cite this version:

Julien Simon. Abord de l'épidémiologie des cancers par des données de registre et des données médico-administratives : complémentarité des deux approches. Médecine humaine et pathologie. Université Paris-Saclay, 2022. Français. NNT : 2022UPASR004 . tel-03647707

**HAL Id: tel-03647707**

**<https://theses.hal.science/tel-03647707v1>**

Submitted on 20 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Abord de l'épidémiologie des cancers par des données  
de registre et des données médico-administratives :  
complémentarité des deux approches

*Epidemiological studies of cancer using registry data and medico-  
administrative data: complementarity of the two approaches*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 570 : santé publique (EDSP)

Spécialité de doctorat : Santé publique, épidémiologie

Graduate school : Santé publique

Réfèrent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche **CESP** (Université Paris-Saclay, UVSQ, Inserm),  
sous la direction de **Catherine Quantin**, Professeure, Université Paris-Saclay  
et le co-encadrement de **Karine Goueslard**, Docteure, Université de Bourgogne

**Thèse soutenue à Paris-Saclay, le 03 mars 2022, par**

**Julien SIMON**

**Composition du Jury**

**Marie-Christine BOUTRON**

Directrice de recherche, Université  
Paris-Saclay

Présidente

**Peter VON THEOBALD**

Professeur, Université de la  
Réunion

Rapporteur & Examineur

**Emmanuel SIMON**

Professeur, Université de  
Bourgogne

Rapporteur & Examineur

**Jean-Pierre DAURES**

Professeur, Université de  
Montpellier

Examineur

**Catherine QUANTIN**

Professeure, Université Paris-Saclay

Directrice de thèse

**Titre :** Abord de l'épidémiologie des cancers par des données de registre et des données médico-administratives : complémentarité des deux approches.

**Mots clés :** épidémiologie, cancer, registre, base médico-administrative, avantages, limites

**Résumé :** À partir des données des registres, qui couvrent environ 20% de la population, les tendances des estimations d'incidence nationale et régionales sont publiées régulièrement. Les études de survie et les publications de l'incidence selon les facteurs pronostiques contribuent à l'évaluation de l'impact du dépistage ou de nouveaux traitements des cancers.

Les données médico-administratives sont de plus en plus utilisées, à la fois à des fins de recherche et de surveillance sanitaire. L'intérêt potentiel de ces bases de données dans une optique épidémiologique apparaît clairement dans la mesure où elles fournissent des données individuelles médicalisées, structurées et codées de manière standardisée. Toutefois, ayant été construites pour répondre à des objectifs médico-économiques, leur utilisation à des fins épidémiologiques nécessite un travail important de contrôle et de validation des données, et de nombreux algorithmes et études de validation ont été publiés ces dernières années.

Dans une première partie analysant les apports et limites des registres de cancer, nous avons effectué une première étude à partir des données du Registre des cancers du sein et autres cancers gynécologiques de Côte d'Or.

Dans une seconde partie analysant les apports et limites des bases médico-administratives, nous avons effectué une seconde étude à partir des données du Système National des Données de Santé.

Dans une dernière partie, nous avons effectué une revue de la littérature composée d'études utilisant des données issues de registres ou de bases médico-administratives. Cette revue de la littérature et la description de plusieurs exemples de chaînage entre des bases médico-administratives et des registres de cancer nous ont permis d'analyser les différences et la complémentarité qu'il peut exister entre ces deux approches.

**Title:** Epidemiological studies of cancer using registry data and medico-administrative data: complementarity of the two approaches.

**Keywords:** epidemiology, cancer, registry, medico-administrative database, advantages, limits.

**Abstract:** Based on registry data, which cover approximately 20% of the population, trends in national and regional incidence estimates are published regularly. Survival studies and reports of incidence by prognostic factors can contribute to the evaluation of the impact of cancer screening and new treatments.

Medico-administrative data are increasingly used for both research and health surveillance purposes. The potential value of these databases from an epidemiological point of view is clear insofar as they provide individualized medical data that is structured and coded in a standardized way. However, having been built to meet medico-economic objectives, their use for epidemiological purposes requires significant data control and validation work, and many algorithms and validation studies have been published in recent years.

First, we conducted a study using data from the Côte d'Or Breast and other gynaecological cancer registry in order to analyse the contributions and limitations of cancer registries.

We then conducted a second study using data from the "Système National des Données de Santé" (SNDS), which is the French national health data system that contains individual, exhaustive and linkable but anonymous data on health expenditures for about ninety-nine percent of the French population, in order to analyse the contributions and limitations of medico-administrative databases.

The last part of the investigation was a literature review of studies using data from registries or medico-administrative databases. We used this review of the literature and several examples of chaining between medico-administrative databases and cancer registries to analyse the differences between these two approaches and their potential complementarity.

# Remerciements

A Madame le Professeur Catherine QUANTIN, directrice de thèse, qui m'a sans cesse encouragé et guidé dans la rédaction de mes articles et de ma thèse, et qui m'a inspiré par son implication professionnelle, sa pédagogie et sa vision de l'épidémiologie. Je ne doute pas que notre rencontre et nos échanges auront une influence bénéfique tout au long de ma vie professionnelle.

A Madame le Docteur Karine GOUESLARD, co-encadrante de thèse, pour son soutien, pour le temps qu'elle m'a consacré et ses remarques toujours pertinentes lors de ses multiples relectures de mes travaux. Mes remerciements également pour sa précieuse contribution et ses enseignements concernant les données du SNDS.

A Monsieur le Docteur Patrick ARVEUX, qui m'a permis d'entrer dans le monde de la recherche et de l'épidémiologie, qui m'a donné la confiance et le soutien dont j'avais besoin pour débiter et qui m'a guidé par son expertise et sa pédagogie en faisant toujours preuve d'une grande disponibilité.

A Madame le Docteur Sandrine DABAKUYO-YONLI, pour son expertise en épidémiologie et sa maîtrise du registre des cancers du sein et des cancers gynécologiques de Côte d'Or qui ont apporté beaucoup à ce travail.

A Madame le Docteur Marie-Christine BOUTRON, Monsieur le Professeur Emmanuel SIMON, Monsieur le Professeur Peter VON THEOBALD, Monsieur le Professeur Jean-Pierre DAURES qui ont accepté de participer au jury de cette thèse.

A Madame Gwenaëlle PERIARD, pour sa bienveillance et son aide précieuse lors de la soumission des articles.

A l'équipe du Registre des cancers du sein et des cancers gynécologiques de Côte d'Or, pour leur accueil, leur gentillesse et leur grande disponibilité à chacune de mes sollicitations.

A l'équipe du service de Biostatistiques et d'information médicale du CHU de Dijon, pour leur bienveillance et leur soutien.

A mes enfants Laurine et Lucien, à ma famille, mes amis, et tous ceux qui m'ont soutenu durant l'ensemble de mon parcours.

# Liste des principales abréviations

ALD	Affection Longue Durée
AJCC	American Joint Commission on Cancer
ATIH	Agence Technique de l'Information sur l'Hospitalisation
BCMD	Base des Causes Médicales de Décès
CCAM	Classification Commune des Actes Médicaux
CIM	Classification Internationale des Maladies
CIRC	Centre International de Recherche sur le Cancer
CNAM	Caisse Nationale d'Assurance Maladie
CNIL	Commission Nationale de l'Informatique et des Libertés
CP	Cancer du Pancréas
CS	Cancer du Sein
CT	Chimiothérapie
DG	Diabète Gestationnel
EIQ	Écart interquartile
FISH	Fluorescent In Situ Hybridization
HDH	Health Data Hub
HER2	Human Epidermal growth factor Receptor 2
HR	Hazard Ratio
HT	Hormonothérapie
IARC	Association Internationale des Registres du Cancer
IMC	Indice de Masse Corporelle
INCa	Institut National du Cancer
INDS	Institut National des Données de Santé
INSERM	Institut National de la Santé Et de la Recherche Médicale
NIR	Numéro d'Inscription au Répertoire des personnes physiques
PMSI	Programme de Médicalisation des Systèmes d'Information
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RH	Récepteurs Hormonaux
RSS	Résumé de sortie standardisé
SEER	Surveillance, Epidemiology and End Results

SG	Survie Globale
SNDS	Système National des Données de Santé
SNIIRAM	Système National d'Information InterRégimes de l'Assurance Maladie
SSP	Survie Sans Progression
TNM	Tumor-Node-Metastasis

# Valorisations scientifiques

## 1. Articles issus des travaux de thèse

**Survival in patients with HR+/HER2- metastatic breast cancer treated with initial endocrine therapy versus initial chemotherapy. A French population-based study.**

Simon J, Chaix M, Billa O, Kanga AM, Roignot P, Ladoire S, Coutant C, Arveux P, Quantin C, Dabakuyo-Yonli TS.

Br J Cancer. sept 2020;123(7):1071-7. doi: 10.1038/s41416-020-0979-3. Epub 2020 Jul 17. PMID: 32678278; PMCID: PMC7525445.

*(Annexe 1)*

**Increased Risk of Hospitalization for Pancreatic Cancer in the First 8 Years after a Gestational Diabetes Mellitus regardless of Subsequent Type 2 Diabetes: A Nationwide Population-Based Study.**

Simon J, Goueslard K, Arveux P, Bechraoui-Quantin S, Petit J-M, Quantin C.

Cancers (Basel). 15 janv 2021;13(2). doi: 10.3390/cancers13020308. PMID: 33467667; PMCID: PMC7829937.

*(Annexe 2)*

**Is Gestational Diabetes Mellitus a Risk Factor of Maternal Breast Cancer? A Systematic Review of the Literature.**

Simon, J.; Goueslard, K.; Bechraoui-Quantin, S.; Arveux, P.; Quantin, C.

Biomedicines **2021**, *9*, 1174. <https://doi.org/10.3390/biomedicines9091174>

*(Annexe 3)*

## **2. Communications orales sur les travaux de thèse**

### **Congrès EMOIS 2019 à Nancy**

Simon J., Goueslard K., Petit J-M., Arveux P., Quantin C. « Etude de la relation entre cancer du pancréas et diabète gestationnel à partir des données du PMSI »

### **Pancreatic Cancer Symposium 2019 à Toulouse**

Simon J., Goueslard K., Petit J-M., Arveux P., Quantin C. Risk of pancreatic cancer 8 years after gestational

# Sommaire

Remerciements	4
Liste des principales abréviations	6
Valorisations scientifiques	8
Sommaire	10
Sommaire des tableaux	12
Sommaire des figures	13
Introduction	15
Objectifs scientifiques de la thèse	19
1. Registre de cancer	20
1.1. Les données de registre : exemple du registre de cancers du sein de Côte d'Or	20
1.1.1 Le registre de cancers du sein de côte d'Or : présentation et description	20
1.1.2 Contexte et question de recherche	22
1.1.3 Matériel et méthodes	23
1.1.4 Analyses statistiques	25
1.1.5 Résultats	28
1.1.6 Discussion	39
1.1.7 Conclusion	42
1.2. Avantages et inconvénients de données issues du registre des cancers du sein de Côte d'Or	42
1.2.1 Les avantages :	42
1.2.2 Les inconvénients :	44
1.3. Exemples d'études à partir de registres de cancer dans le monde	45
1.3.1 Le registre de cancer de Hong-Kong	47
1.3.2 Le registre du cancer de la région Nord du Portugal	48
1.3.3 Le registre de Cancer d'Osaka au Japon	49
1.3.4 Le registre "Surveillance, Epidemiology, and End Results" (SEER) aux Etats-Unis.	49
1.3.5 Le registre national du cancer du Pays-Bas	51
1.3.6 Synthèse	52
2. Données médico-administratives	53
2.1. L'exemple du SNDS	53
2.1.1 Présentation et description	53
2.1.2 Etude de la relation entre le diabète gestationnel et cancer du pancréas	59
2.2. Avantages et inconvénients des données médico-administratives	77
2.2.1 Les avantages et inconvénients du SNDS en France	77
2.2.2 Exemples d'études à partir de bases médico-administratives dans le monde	81
3. Complémentarité des deux approches	87

3.1. Revue de la littérature concernant la relation entre le diabète gestationnel et le cancer du sein _____	87
3.1.1 Contexte et objectif : étude du lien entre diabète gestationnel et cancer du sein _____	87
3.1.2 Matériel et méthodes _____	89
3.1.3 Résultats _____	93
3.1.4 Discussion _____	105
3.1.5 Conclusion _____	111
3.2. Les registres de cancer en France : utilisation de données complémentaires. ____	112
3.3. Exemples de chaînage entre le SNDS et d'autres sources de données en France.	113
3.4. Exemples de chaînage entre les bases médico-administratives et les registres de cancer dans le monde _____	115
3.3.1 Etats-Unis _____	115
3.3.2 En Ontario _____	118
3.3.3 A Taïwan _____	119
3.3.4 En Angleterre _____	120
3.3.5 Au Japon _____	120
Conclusion _____	122
Références _____	124
Annexes _____	140

# Sommaire des tableaux

<b>Tableau 1. Caractéristiques des patientes et caractéristiques de la maladie dans les groupes hormonothérapie initiale et chimiothérapie initiale. Les valeurs en gras indiquent des résultats statistiquement significatifs. ....</b>	<b>30</b>
<b>Tableau 2. Analyses uni- et multivariées de la survie des patientes atteintes d'un cancer du sein métastatique à récepteurs hormonaux positifs et HER2 négatifs. Les valeurs en gras indiquent des résultats statistiquement significatifs. ....</b>	<b>34</b>
<b>Tableau 3. Distribution des caractéristiques des patientes appariées avec le score de propension selon le traitement systémique anti-métastatique initial. ....</b>	<b>37</b>
<b>Tableau 4. Comparaison des caractéristiques des femmes avec des antécédents de Diabète Gestationnel (DG) ou sans antécédent de DG. ....</b>	<b>68</b>
<b>Tableau 5. Modèles de Cox de la relation entre le diabète gestationnel (DG) et l'hospitalisation pour cancer du pancréas (n=126). Le DG, le diabète de type 2 ultérieur et le tabagisme sont des covariables dépendantes du temps. ....</b>	<b>69</b>
<b>Tableau 6. Caractéristiques des études de cohorte. ....</b>	<b>93</b>
<b>Tableau 7. Caractéristiques des études cas-témoins. ....</b>	<b>98</b>
<b>Tableau 8. Risque de biais et évaluation de la qualité des études de cohorte. ....</b>	<b>102</b>
<b>Tableau 9. Risque de biais et évaluation de la qualité des études cas-témoins. ....</b>	<b>103</b>

# Sommaire des figures

<b>Figure 1. Schéma montrant la censure au début de l'hormonothérapie d'entretien, dans notre seconde analyse de sensibilité, pour les patientes ayant débuté leur schéma thérapeutique par de la chimiothérapie après le diagnostic de leur cancer du sein métastatique .....</b>	<b>27</b>
<b>Figure 2. Diagramme de Flux de la population d'étude.....</b>	<b>30</b>
<b>Figure 3. Survie sans progression et survie globale des patientes atteintes d'un cancer du sein métastatique de sous-type luminal en fonction du traitement systémique anti-métastatique initial. ....</b>	<b>33</b>
<b>Figure 4. Carte présentant la situation mondiale de l'enregistrement des cancers par pays .....</b>	<b>46</b>
<b>Figure 5. Carte des Etats-Unis montrant les zones couvertes par le programme SEER (Surveillance, Epidemiology, and End Results) en 2021 .....</b>	<b>50</b>
<b>Figure 6. Les circuits d'information du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM).....</b>	<b>57</b>
<b>Figure 7. Diagramme de Flux de la population d'étude.....</b>	<b>68</b>
<b>Figure 8. Diagramme de flux du processus de sélection des articles .....</b>	<b>91</b>



# Introduction

## Épidémiologie des cancers :

Le cancer est une pathologie provoquée par la transformation de cellules qui deviennent anormales et prolifèrent de façon excessive. Ces cellules dérégées finissent par former une masse qu'on appelle tumeur maligne. Les cellules cancéreuses ont tendance à envahir les tissus voisins et à se détacher de la tumeur. Elles migrent alors par les vaisseaux sanguins et les vaisseaux lymphatiques pour aller former une autre tumeur appelée métastase. (1)

En 2018, le nombre de nouveaux cas de cancer dans le monde est estimé à 18,1 millions et le nombre de décès à 9,6 millions. L'incidence de risque cumulé indique qu'un homme sur huit et qu'une femme sur dix développera un cancer au cours de sa vie. Les cancers les plus fréquemment diagnostiqués dans le monde sont ceux du poumon (11,6 % du total), du sein (11,6 % du total) et le cancer colorectal (10,2% du total). Les causes les plus fréquentes de décès par cancer sont les cancers du poumon (18,4 % du total), les cancers colorectaux (9.2 % du total) et de l'estomac (8,2 % du total). (2)

En France, en 2018, on décomptait environ 382 000 nouveaux cas de cancer et 157 000 décès dus aux cancers. Les cancers les plus fréquents sont les cancers du sein, du côlon-rectum et du poumon chez la femme, et les cancers de la prostate, du poumon et du colon-rectum chez l'homme. (3)

Afin de surveiller et mieux soigner les personnes atteintes de cancer, il est important d'utiliser des outils efficaces pour mieux estimer les besoins en matière de prévention, de diagnostic et de traitement. L'un de ces outils de veille sanitaire et de recherche est le registre qui inclut des données les plus exhaustives possibles dans une population définie (départementale, régionale ou nationale). Une autre source qui est utilisée aujourd'hui dans l'épidémiologie des cancers est l'exploitation de bases de données médico-administratives, moins exhaustives d'un point de vue épidémiologique mais plus large au niveau populationnel.

## Les registres de cancers en France :

Un registre est défini par le Comité National des Registres comme « un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées ».

En France, à partir de 1975, plusieurs registres de cancer départementaux ont été créés. En 1986, la création du Comité National des Registres a permis d'inscrire les registres dans une politique nationale de santé publique et de recherche. Les registres des cancers généraux et spécialisés qualifiés par le Comité National des Registres (devenu depuis 2014 le Comité d'Evaluation des Registres) se sont alors regroupés en association en 1991 au sein du réseau français des registres de cancer, FRANCIM. Ce réseau comprend 14 registres généraux métropolitains couvrant 19 départements français complétés par 3 registres généraux outremerins, 10 registres spécialisés dans certaines localisations cancéreuses (dont le registre des cancers du sein et gynécologiques de Côte d'Or), Mésonat qui est un registre multicentrique des mésothéliomes pleuraux couvrant 22 départements et le Registre National des Cancers de l'Enfant. Le réseau FRANCIM s'est associé en 2008 à un « programme de travail partenarial » avec le Service de biostatistique des Hospices civils de Lyon, Santé Publique France et l'Institut National du Cancer (INCa) (4).

Cela a permis aux registres des cancer de contribuer à l'orientation, au suivi et à l'évaluation des politiques de lutte contre le cancer. À partir des données des registres, qui couvrent environ 20% de la population, différents types d'indicateurs sont publiés comme, par exemple, des estimations nationales et régionales de prévalence, d'incidence et de la mortalité par cancer. Ces indicateurs permettent de piloter et d'évaluer, au niveau national et régional, les actions de prévention et de prise en charge des patients. Plus précisément, Ils servent aussi à apprécier l'évolution dans le temps de l'impact de certains risques connus. Leur communication au grand public sert à sensibiliser les populations et les professionnels à la prévention primaire et au dépistage. Dans les départements couverts par un registre de cancer, les données permettent de répondre à la majorité des besoins locaux du département (connaissance des risques, évaluation des actions de santé publique et des soins). La géolocalisation des cas de cancer, possible en raison de l'enregistrement de l'adresse des patients au diagnostic, est essentielle pour la surveillance des cancers liés à l'environnement ou

l'investigation de clusters. L'actualisation du statut vital des cas incidents des registres permet la réalisation d'études de survie qui reflète l'impact de prises en charge des patients atteints de cancer en population générale. Enfin, les registres sont des outils essentiels pour contribuer à l'évaluation de l'impact des actions de prévention primaire et de dépistage.

### Les données médico-administratives en France :

Les bases de données médico-administratives, quant à elle, sont plus récentes notamment dans leur utilisation dans un objectif épidémiologique. En effet, c'est la loi du 31 juillet 1991 qui oblige les établissements de santé publics et privés à procéder à l'analyse de leur activité médicale et transmettre aux services de l'État et à l'Assurance maladie « les informations relatives à leurs moyens de fonctionnement et à leur activité » : articles L.6113-7 et L.6113-8 du code de la santé publique.

À cette fin, ils doivent « mettre en œuvre des systèmes d'information qui tiennent compte notamment des pathologies et des modes de prise en charge » : c'est la définition même du programme de médicalisation des systèmes d'information (PMSI). Les informations ainsi produites sont utilisées principalement à deux fins : pour le financement des établissements de santé (tarification à l'activité), et pour l'organisation de l'offre de soins (planification).

Puis, en 1999 est créé le Système National d'Information InterRégimes de l'Assurance Maladie (SNIIRAM) par la loi de financement de la sécurité sociale. C'est une base de données nationale dont les objectifs sont de contribuer à une meilleure gestion de l'assurance maladie et des politiques de santé, d'améliorer la qualité des soins et de transmettre aux professionnels de santé les informations pertinentes sur leur activité. Son périmètre, ses finalités, son alimentation et l'accès aux données sont définis dans un arrêté du ministère des affaires sociales et de la santé transmis pour avis à la Commission Nationale de l'Informatique et des Libertés (CNIL) puis publié au Journal Officiel. La CNIL est une autorité indépendante compétente en matière de protection des données personnelles. La Caisse Nationale d'Assurance Maladie (CNAM) est chargée de la gestion du SNIIRAM, elle est responsable du système d'information au regard de la CNIL.

La loi du 26 janvier 2016 de modernisation de notre système de santé a créé le Système National des Données de Santé (SNDS). Cette création fait suite à la volonté d'ouvrir l'accès aux données de santé afin que « leurs potentialités soient utilisées au mieux dans l'intérêt de la collectivité ». Après un avis de la CNIL, un décret du 26 décembre 2016 ([Délibération n° 2016-316 du 13 octobre 2016 portant avis sur un projet de décret en Conseil d'Etat relatif au Système national des données de santé](#)) (5) encadre les conditions d'accès et prévoit une procédure d'autorisation délivrée par la CNIL. Le SNDS a été mis en œuvre le 10 avril 2017.

Cette nouvelle base, qui regroupe des bases de données existantes doit permettre d'améliorer les connaissances en matière de prise en charge médicale et d'élargir le champ des recherches dans le domaine de la santé.

# Objectifs scientifiques de la thèse

1. Décrire et illustrer à l'aide de plusieurs exemples d'études (issues de nos travaux et de la littérature) les apports respectifs des données médico-administratives et des données de registre dans l'épidémiologie des cancers.
2. Une première partie sera consacrée aux registres de cancer. Dans un premier temps, nous effectuerons une étude à partir des données du registre des cancers du sein et gynécologiques de Côte-d'Or à partir desquelles nous évaluerons l'effet sur la survie sans progression et sur la survie globale du choix de la première ligne de traitement systémique chez les femmes atteintes d'un cancer du sein métastatique positif aux récepteurs hormonaux et négatif aux récepteurs du facteur de croissance HER2 (RH+/HER2-). Dans un second temps, à partir des forces et des limites de cette étude et d'autres travaux issus de la littérature, nous discuterons les avantages et les limites des registres de cancers.
3. Une seconde partie sera consacrée aux bases de données médico-administratives. Dans un premier temps, nous étudierons la relation entre le diabète gestationnel et le risque de cancer du pancréas à partir des données de la base nationale du SNDS. Dans un second temps, à partir des forces et des limites de cette étude et d'autres travaux issus de la littérature, nous discuterons les avantages et les limites des bases médico-administratives.
4. Nous effectuerons, dans une troisième partie, une revue de la littérature concernant la relation entre le diabète gestationnel et le risque de cancer du sein. Cette revue de la littérature, nous permettra de prendre en compte différentes études qui utilisent des données de registres et/ou des données médico-administratives. Les forces et limites méthodologiques de ces études, ainsi que plusieurs exemples de chaînage entre des bases médico-administratives et des registres de cancer, nous permettront d'étayer une synthèse concernant la complémentarité qui peut exister entre les registres de cancer et les données médico-administratives contribuant à faciliter la recherche concernant les cancers.

# **1. Registre de cancer**

## **1.1. Les données de registre : exemple du registre de cancers du sein de Côte d'Or**

### **1.1.1 Le registre de cancers du sein de côte d'Or : présentation et description**

Le Registre des cancers du sein et autres cancers gynécologiques de Côte d'Or et le seul registre spécialisé en gynécologie au niveau national, il est agréé par le Comité d'Evaluation des Registres et bénéficie d'un financement public national (INCa et Santé Publique France). Il collecte pour toutes les patientes résidant en Côte d'Or, des informations sur la maladie et l'ensemble des traitements délivrés. Le registre collabore avec l'ensemble des registres français (FRANCIM), l'INCa, Santé Publique France, l'Association Internationale des Registres du Cancer (IARC) et le Réseau européen des registres du cancer pour la surveillance épidémiologique des cancers gynécologiques à travers la production de statistiques d'incidence, de mortalité et de survie des cancers en France et dans le monde. Il s'agit d'une base de données utilisée également pour évaluer l'efficacité du système de santé, pour répondre à des questions de recherche et fournir des données de « vie réelle » sur les traitements des cancers gynécologiques.

L'enregistrement de tous les cas de cancers du sein et cancers gynécologiques respecte les règles de codage établies au niveau national par le réseau FRANCIM des registres français des cancers, et au niveau international par le Réseau européen des registres du cancer et le Centre International de Recherche sur le Cancer (CIRC).

Le registre des cancers du sein et des cancers gynécologiques de Côte d'Or a été réévalué favorablement en Janvier 2021 par le Comité d'Evaluation des Registres (rang AAA pour les 3 aspects : moyens et méthodes ; intérêt en santé publique ; intérêt en recherche).

Le registre des cancers du sein et autres cancers gynécologiques de Côte d'Or répertorie dans deux bases distinctes les cas de cancers du sein (topographies C50 et D05 selon la

Classification Internationale des Maladies (CIM)) et les cas de tumeurs malignes des organes génitaux de la femme (topographie C51 à C58) diagnostiqués dans le département depuis 1982. Depuis 1998, il est également indiqué dans la base des cas de néoplasies lobulaires bas grade et moyen grade. Il y est également enregistré et codé des tumeurs borderlines ovariennes ainsi que des tumeurs in situ et des lésions pré-cancéreuses (CIN2) du col utérin.

Concernant la Base SEIN, à ce jour, 14 658 tumeurs dont 13 043 tumeurs invasives et 1 474 cancers in situ (ainsi que 141 néoplasies lobulaires -non cancéreuses-) sont répertoriées dans la base pour la période 1982-2017. On compte 128 cancers du sein chez l'homme pour cette période.

Concernant la Base PELVIS, elle enregistre 8 138 tumeurs depuis 1982 dont notamment 4 318 lésions du col, 1 848 du corps de l'utérus et 1 549 cancers ovariens. Depuis 1998, parmi les 5 662 tumeurs enregistrées dans la base Pelvis on compte 2 619 tumeurs infiltrantes (46%).

Le registre des cancers du sein et autres cancers gynécologiques de Côte d'Or est un registre spécialisé qui recueille dans un questionnaire très détaillé un grand nombre de variables (plus de 150) dans le cadre du diagnostic primitif ainsi que les traitements dans le cadre des évolutions. (cf questionnaire cancer du sein en Annexe 4)

Outre les variables obligatoires (identification personnelle, sexe, date de naissance et commune de naissance, adresse complète au moment du diagnostic, date d'incidence, base de diagnostic la plus valide, localisation de la tumeur primitive, type histologique et comportement tumoral, sources d'informations), toutes les variables « recommandées » (date des dernières nouvelles, état aux dernières nouvelles, stade, grade/différenciation, traitement initial, mode de découverte pour les tumeurs bénéficiant de campagne de dépistage organisé) sont enregistrées ainsi que tous les traitements.

## 1.1.2 Contexte et question de recherche

Une première étude a été effectuée dans le cadre de cette thèse à partir des données du registre des cancers du sein et cancers gynécologiques de Côte d'Or. Nous avons choisi de travailler sur une question que se posent les cliniciens concernant le choix de la première ligne de traitement du cancer du sein métastatique.

### 1.1.2.1 Le cancer du sein : épidémiologie

Le cancer du sein est le cancer le plus fréquent, représentant environ un quart de l'ensemble des cas de cancer chez la femme, en Europe et dans le monde. En 2020, il a été estimé 2 261 419 nouveaux cas de cancer du sein et 684 996 décès causés par le cancer du sein au niveau mondial (2). En France, les données de « Santé Publique France » estiment 58 459 nouveaux cas pour 12 146 décès pour 2018, avec une survie de 87% à 5 ans (6). Grâce au dépistage de plus en plus précoce et à l'amélioration de la prise en charge, la survie des cancers du sein non-métastatique et métastatique ne cesse de s'améliorer (7–10).

### 1.1.2.2 Les traitements et les recommandations

Selon les recommandations internationales, les traitements doivent être prescrits sur une base individuelle (médecine personnalisée) et tenir compte du statut des récepteurs hormonaux (RH) et du « récepteur du facteur de croissance épidermique humain 2 » (Human Epidermal growth factor Receptor 2 ou HER2). Le sous-type luminal (RH+/HER2-) représente environ 70% de l'ensemble des cancers du sein.

Concernant le cancer du sein métastatique luminal, même si aucune différence n'a été trouvée entre la chimiothérapie (CT) et l'hormonothérapie (HT) en termes de survie, l'HT est un meilleur choix car il est moins toxique (11,12). En effet, la chimiothérapie augmente le risque de nausées, de vomissements et d'alopécie. Certains essais cliniques ont montré une bonne survie sans progression (SSP) pour l'HT seul et une augmentation de la SSP pour l'HT en association avec des thérapies ciblées (13–15). Selon les recommandations internationales, l'HT devrait être l'option préférée dans le cancer du sein métastatique luminal, même en présence de métastases viscérales, sauf en cas de crise viscérale ou d'hormonorésistance. La crise viscérale est définie comme un dysfonctionnement grave des organes, évalué par des signes, des symptômes, des examens de laboratoire et une progression rapide de la maladie. La

crise viscérale n'est pas la simple présence de métastases viscérales, mais implique une défaillance viscérale importante conduisant à une indication pour un traitement plus rapidement efficace (16–18). Malgré ces données, certaines études récentes ont montré que les recommandations ne sont pas toujours suivies dans la pratique quotidienne et que les pratiques médicales varient considérablement (19–23).

### 1.1.2.3 Objectifs

Cette étude a donc pour objectifs d'évaluer, chez les femmes atteintes d'un cancer du sein métastatique RH+/HER2- en Côte-d'Or, l'effet du choix du traitement systémique métastatique de première ligne sur la SSP comme objectif principal et sur la survie globale (SG) comme objectif secondaire.

## 1.1.3 Matériel et méthodes

### 1.1.3.1 Schéma de l'étude

Cette étude en population a été entreprise à partir des données du registre du cancer du sein et gynécologique de Côte d'Or. Les femmes diagnostiquées avec un cancer du sein métastatique RH+/HER2- de janvier 1998 à décembre 2016 ont été sélectionnées rétrospectivement dans les données du registre et incluses dans cette étude. Les carcinomes canaux ou lobulaires invasifs ont été inclus. Les récepteurs hormonaux et le statut HER2 ont été déterminés par immunohistochimie, complétée par un examen FISH (Fluorescent in situ Hybridization) chez des patientes avec un résultat équivoque. Le statut HER2 a été déterminé rétrospectivement pour les cas diagnostiqués avant les analyses de routine. Les femmes qui ont eu des métastases après 2016, celles qui n'ont pas reçu de traitement anti-métastatique systémique et celles qui ont eu simultanément une CT et une HT en première ligne ont été exclues.

Les comorbidités ont été définies selon l'indice de comorbidité de Charlson (ICC). Cet indice est constitué de 19 troubles de comorbidité : l'infarctus du myocarde, l'insuffisance cardiaque congestive, la maladie vasculaire périphérique, la maladie cérébro-vasculaire, la démence, la maladie pulmonaire chronique, la maladie des tissus conjonctifs, les ulcères, les maladies du foie, le diabète, l'hémiplégie, les maladies du rein, la tumeur, la leucémie, le lymphome, le SIDA (24). Les patientes ont été classées en deux groupes : aucune comorbidité (ICC = 0) et au moins une comorbidité (ICC > 0).

Le stade du cancer a été déterminé en utilisant la cinquième (avant 2008) et la sixième édition (après 2008) de la classification de l'American Joint Commission on Cancer (AJCC) « Tumor-Node-Metastasis » (TNM) des tumeurs malignes.

Conformément aux directives européennes, nous avons défini la résistance primaire à l'hormonothérapie comme une rechute pendant les deux premières années de l'hormonothérapie adjuvante. La résistance secondaire à l'hormonothérapie a été définie comme une rechute pendant l'hormonothérapie adjuvante mais après les deux premières années, ou une rechute dans les douze mois suivant la fin de l'hormonothérapie adjuvante (16).

### 1.1.3.2 Les critères de jugement

Le critère de jugement principal était la SSP définie comme le délai entre le début du premier traitement systémique de la maladie métastatique et la progression de la maladie ou le décès, selon l'événement qui surviendrait le premier. La progression, telle que déterminée par l'initiative DATECAN (25), a été définie comme la progression des métastases initiales ou l'apparition de nouvelles métastases. La définition de la progression fait généralement appel à des techniques d'imagerie (échographies, radiographies, scanners, Imageries par Résonance Magnétique ou Tomographies par Emission de Positron) permettant de détecter une augmentation de la taille de la métastase ou la propagation des lésions aux tissus voisins. Elle peut également être établie à partir d'une progression biochimique sur la base d'une augmentation d'un marqueur tumoral.

Notre critère de jugement secondaire était la SG, qui était définie comme le délai entre le début du premier traitement systémique de la maladie métastatique et le décès quel qu'en soit la cause.

## 1.1.4 Analyses statistiques

Nous avons utilisé des tests de  $\chi^2$  ou des tests exacts de Fisher pour les variables catégorielles et le test de somme de rang de Wilcoxon pour les variables continues pour comparer les caractéristiques de base entre les patientes ayant reçu de la CT initialement et celles ayant reçu de l'HT initialement.

Les résultats de survie ont été estimés en utilisant la méthode de Kaplan-Meier et comparés entre les groupes en utilisant le test du Log-Rank.

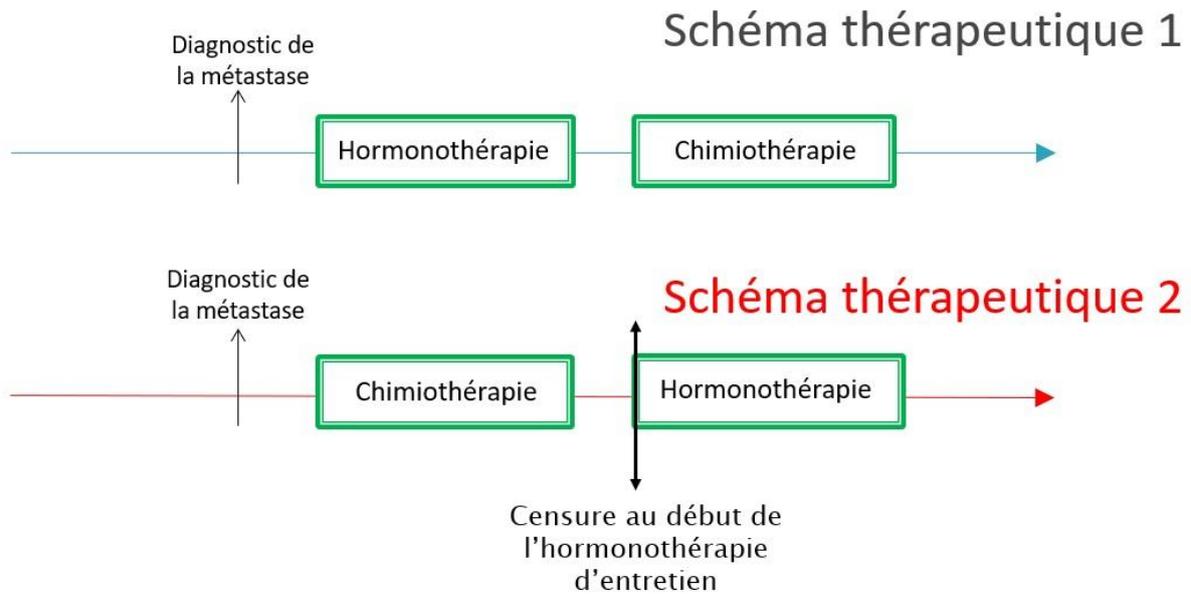
Les différences de SSP et de SG entre les patientes initialement traitées par CT ou HT ont été analysées en utilisant des modèles de Cox avec ajustement pour les covariables avec une valeur de  $p < 0,2$  dans les analyses univariées. L'hypothèse des risques proportionnels a été vérifiée pour chaque variable à l'aide des tests de résidus de Schönfeld, et pour chaque facteur qui ne répondait pas à l'hypothèse, nous avons inclus une interaction avec le temps dans le modèle. Nous avons également testé les corrélations ( $\rho < 0,7$ ) entre les différentes variables.

Pour déterminer l'association entre les facteurs pronostiques et les résultats de survie, nous avons utilisé des modèles de Cox univariés. Les facteurs pronostiques étaient : l'âge au moment du diagnostic du cancer du sein métastatique ( $< 65$  ans /  $\geq 65$  ans), l'ICC ( $ICC = 0$  /  $ICC > 0$ ), le statut ménopausique (non-ménopausée / ménopausée), l'année du diagnostic de la tumeur primaire en utilisant la médiane ( $< 2008$  /  $\geq 2008$ ), le type de la tumeur primaire (canaulaire / lobulaire), si la maladie métastatique était primaire (oui / non), les sites métastatiques initiaux (os, viscéral ou multiple) et la résistance primaire à l'hormonothérapie (oui / non). Les patientes vivantes sans progression de la maladie ont été censurées à la fin de la période d'étude (le 31 décembre 2017).

Dans une première analyse de sensibilité, nous avons utilisé un score de propension pour considérer la subjectivité du choix du médecin (26) et pour ajuster sur les variables qui étaient statistiquement différentes entre les deux groupes de patientes (celles qui ont été traitées par une HT initialement et celles qui ont été traitées par une CT initialement). En effet, le score de propension est la probabilité qu'un patient reçoive un traitement donné en fonction de ses caractéristiques initiales connues. A partir des différents facteurs de confusions potentiels

disponibles, on calcule donc, grâce à des modèles logistiques, pour chaque patiente un score unique traduisant la probabilité qu'avait cette patiente de recevoir de l'HT ou de la CT. On considère alors que les patientes ayant le même SP ont le même profil de facteurs de confusion parmi ceux connus, qu'elles peuvent être comparées comme si elles avaient reçu le traitement par randomisation et donc considérer que le traitement est la cause de la divergence d'évolution. Cette méthode permet de s'approcher de la méthodologie des essais cliniques randomisés, mais seulement pour les facteurs inclus dans le score de propension. Chaque variable incluse dans le modèle logistique multivarié diffère statistiquement entre les patientes avec CT initiale et HT initiale. Tout d'abord, chaque patiente traitée par CT initialement a été appariée avec une patiente qui a subi une HT initialement avec les mêmes caractéristiques en utilisant le score de propension. L'appariement du plus proche voisin a été utilisé, avec une distance de score de propension maximale entre les sujets appariés de 0,1 ; les patientes non appariées ont été exclues.

Nous avons ensuite effectué une seconde analyse de sensibilité pour les données de la SSP dans laquelle les patientes avec une CT initiale qui ont reçu une HT d'entretien avant la progression ont été censurées au début de l'HT d'entretien (cf. figure 1). Semblable à une étude précédente (22), cette analyse a été effectuée pour comparer le groupe de patientes qui ont été exclusivement traitées par HT et les autres exclusivement traitées par CT, dans une tentative de passer outre l'effet de l'hormonothérapie d'entretien. Mais nous supposons que cette approche conduit à un biais à la suite de la censure informative, car le processus de censure peut être lié à l'événement étudié. En effet, l'HT d'entretien peut être associée à la survie car les patientes qui débutent par de la CT (schéma thérapeutique 2) peuvent changer de traitement pour de l'HT d'entretien sans avoir subi l'un des événements présents dans la définition de la SSP (la progression de la maladie ou le décès), ces patientes vont être censurées dans notre étude lorsque l'HT d'entretien débute. Alors que dans le groupe ayant reçu de l'HT initialement (schéma thérapeutique 1), systématiquement c'est la progression de la maladie qui entraîne le changement de traitement vers une CT.



**Figure 1. Schéma montrant la censure au début de l'hormonothérapie d'entretien, dans notre seconde analyse de sensibilité, pour les patientes ayant débuté leur schéma thérapeutique par de la chimiothérapie après le diagnostic de leur cancer du sein métastatique**

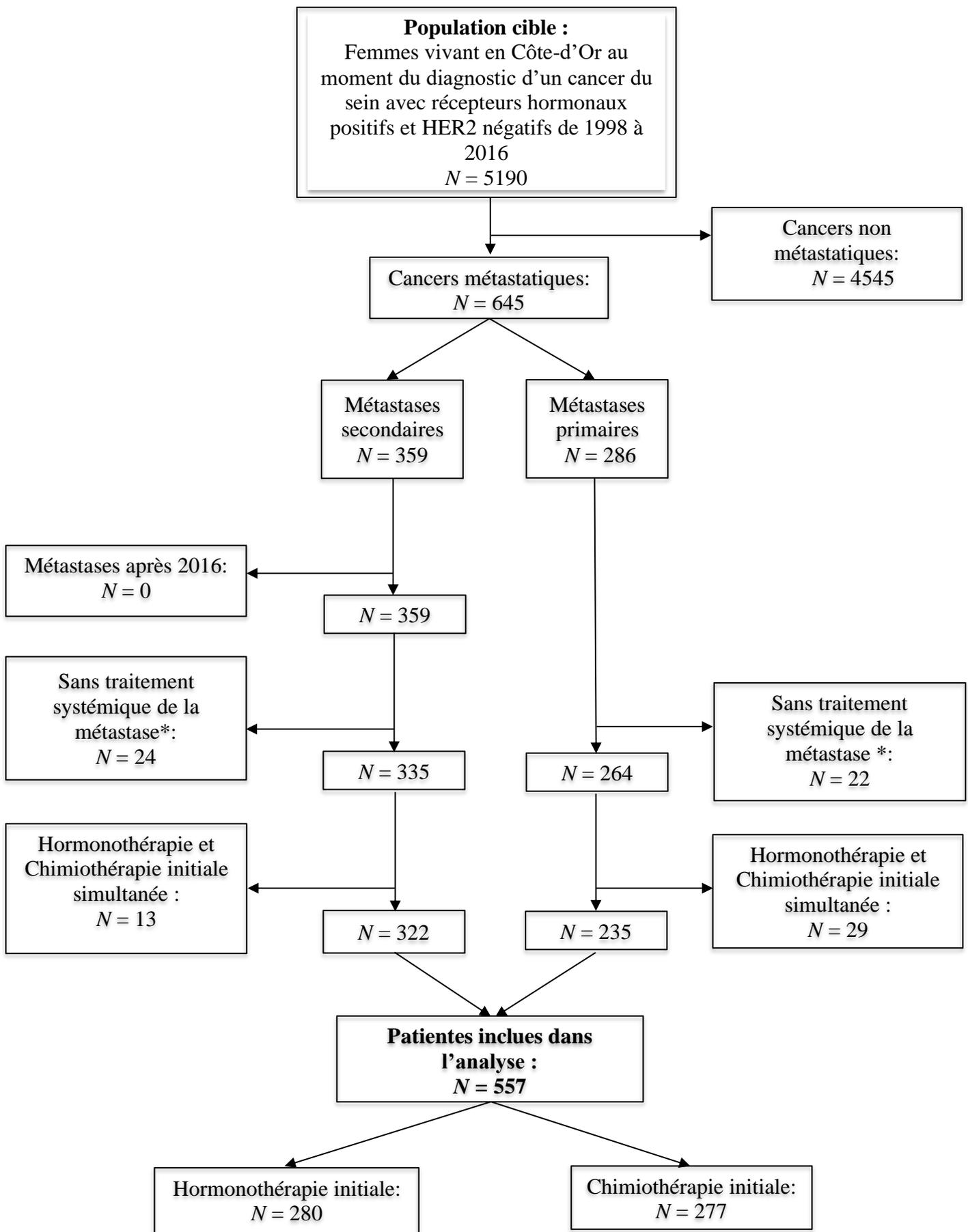
Nous avons ensuite effectué une troisième analyse de sensibilité pour les données de la SSP dans laquelle l'HT était considérée comme une variable dépendante du temps (en tenant compte de l'HT quel que soit le moment où l'HT a été administrée, sans censure sur l'HT d'entretien). Dans le groupe HT initial, l'HT a été commencée le premier jour du suivi et dans le groupe CT initial, l'HT a été débutée lorsque l'HT d'entretien a commencé.

L'analyse statistique a été réalisée à l'aide du logiciel SAS 9.4 (SAS Institute, Cary, NC, USA). La significativité statistique a été fixée à une valeur de  $p < 0,05$ .

## **1.1.5 Résultats**

### **1.1.5.1 Les caractéristiques des patientes**

Au total, 5190 femmes de Côte d'Or ont reçu un diagnostic de cancer du sein à récepteurs hormonaux positifs et HER2 négatif de 1998 à 2016, et 645 cas ont été diagnostiqués ou sont devenus métastatiques au cours du suivi. Parmi ceux-ci, 557 patientes ont reçu un traitement anti-métastatique systémique et ont été analysées (277 CT en première ligne et 280 HT en première ligne) (figure 2).



\* les patientes qui ont reçu exclusivement des traitements de confort ou celles qui ont reçu uniquement une chirurgie et/ou une radiothérapie comme traitements du cancer métastatique.

## **Figure 2. Diagramme de Flux de la population d'étude**

Ces patientes avaient un âge médian de 65 ans (écart interquartile (EIQ) : 55-77 ans) au moment du diagnostic du cancer métastatique. Le suivi médian était de 25 mois (EIQ : 13-45 mois). La SSP médiane était de 13 mois (IC à 95% 12-14 mois) et la SG médiane était de 26 mois (IC à 95% 23-29 mois). À la fin de la période d'étude (31 décembre 2017), pour 95% de toutes les patientes (531/557) leur maladie avait progressé et 83% des patientes (463/557) étaient décédées.

Au total, 50% des patientes ont été traitées par CT initialement et 50% des patientes ont reçu une HT initialement ; leurs caractéristiques sont résumées dans le tableau 1. Il n'y avait aucune différence statistique entre les deux groupes concernant les caractéristiques de la tumeur primaire (année de diagnostic, sous-type histologique, statut RH et statut ganglionnaire). Les patientes traitées par CT initialement étaient plus jeunes, présentaient moins de comorbidités et ont reçu plus de traitements adjuvants que celles traitées par HT initialement. L'HT était plus fréquemment utilisée initialement chez les femmes avec un cancer du sein métastatique primaire ou avec uniquement des métastases osseuses ; il y avait moins de résistance à l'hormonothérapie dans ce groupe.

**Tableau 1. Caractéristiques des patientes et caractéristiques de la maladie dans les groupes hormonothérapie initiale et chimiothérapie initiale. Les valeurs en gras indiquent des résultats statistiquement significatifs.**

Caractéristiques	Chimiothérapie initiale	Hormonothérapie initiale	P
	N = 277 n (%)	N = 280 n (%)	
<b>Âge au diagnostic de la métastase</b>			
Médiane en années (EIQ)	61 (50-69)	73 (61-80)	<b>&lt;0,0001<sup>a</sup></b>
< 65 ans	170 (61)	97 (35)	<b>&lt;0,0001</b>
≥ 65 ans	107 (39)	183 (65)	
<b>Statut ménopausique</b>			<b>&lt;0,0001</b>
Non ménopausée	83 (30)	35 (12)	
Ménopausée	194 (70)	245 (88)	
<b>Indice de comorbidités de Charlson</b>			<b>&lt;0,0001</b>
SCC = 0	178 (64)	130 (46)	
SCC > 0	99 (36)	150 (54)	
<b>Année de diagnostic de la tumeur primaire</b>			0,07
< 2008	97 (35)	119 (42)	
≥ 2008	180 (65)	161 (58)	
<b>Type de tumeur primaire</b>			0,2543
Canalaire	231 (83)	223 (80)	
Lobulaire	46 (17)	57 (20)	
<b>Statut des récepteurs hormonaux</b>			
RO positif	269 (97)	276 (99)	0,2355
RP positif	199 (72)	208 (74)	0,2388
<b>Stade de la tumeur primaire<sup>c</sup></b>			0,1053
T1-T2	187 (67)	172 (61)	
T3-T4	88 (32)	108 (39)	
Inconnu	2 (1)	0	
<b>Statut ganglionnaire<sup>c</sup></b>			0,9667
Négatif	139 (50)	140 (50)	
Positif	132 (48)	132 (48)	
Inconnu	6 (2)	8 (2)	
<b>Thérapie systémique adjuvante<sup>d</sup></b>			
Hormonothérapie	164 (91)	105 (74)	<b>&lt;0,0001</b>
Chimiothérapie	141 (78)	82 (58)	<b>&lt;0,0001</b>
Aucune	6 (3)	24 (17)	<b>&lt;0,0001<sup>b</sup></b>
<b>Radiothérapie adjuvante<sup>d</sup></b>			0,3244
Oui	163 (90)	122 (87)	
Non	18 (10)	19 (13)	
<b>Métastase primaire</b>			<b>0,0002</b>
Oui	96 (35)	139 (50)	
Non	181 (65)	141 (50)	
<b>Intervalle sans métastase<sup>d</sup></b>			0,1423
< 24 mois	26 (14)	29 (21)	
≥ 24 mois	155 (86)	112 (79)	

<b>Site de la métastase initiale</b>			<b>&lt;0,0001</b>
Osseux	62 (22)	136 (49)	
Viscéral	90 (33)	66 (23)	
Multiple	125 (45)	78 (28)	
<b>Hormonorésistance</b>			<b>0,0003</b>
Primaire	50 (18)	35 (13)	
Secondaire	74 (27)	45 (16)	
Hormonosensible	153 (55)	200 (71)	

<sup>a</sup> Test de rang de Wilcoxon.

<sup>b</sup> Test de Fisher exact.

<sup>c</sup> 5<sup>ème</sup> édition avant 2008 et 6<sup>ème</sup> édition depuis 2008 de la classification TNM des tumeurs malignes.

<sup>d</sup> concerne seulement les cancers avec métastases secondaires.

*n*, nombre; EIQ, écart interquartile; RO, récepteurs à œstrogène; RP, récepteurs à progestérone.

### 1.1.5.2 Les traitements

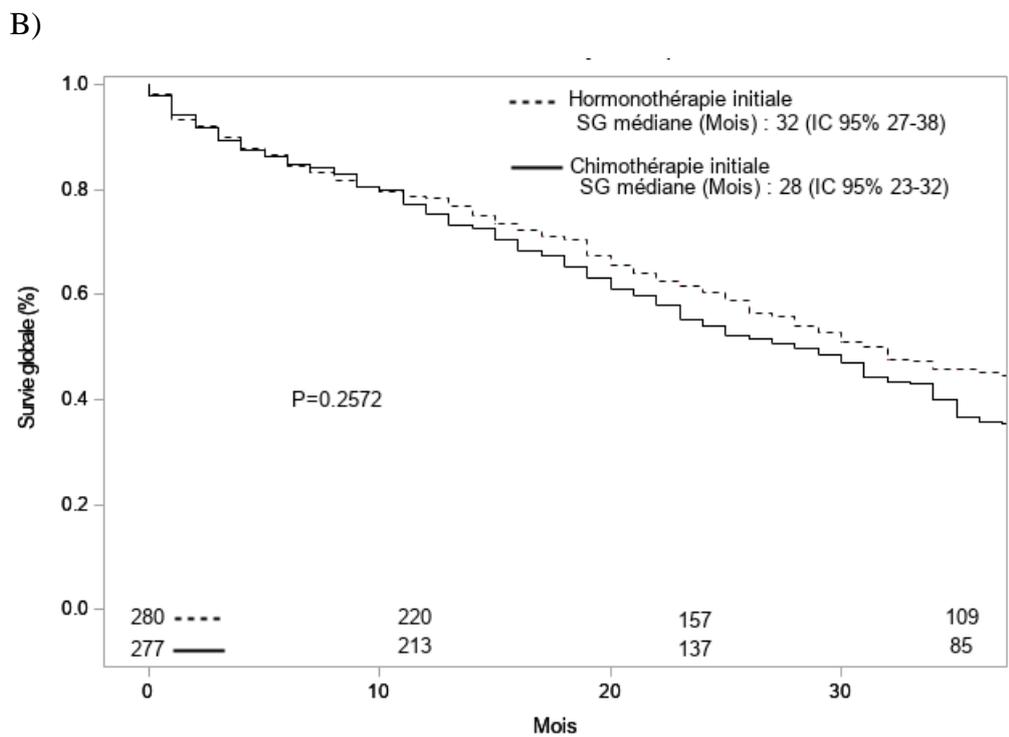
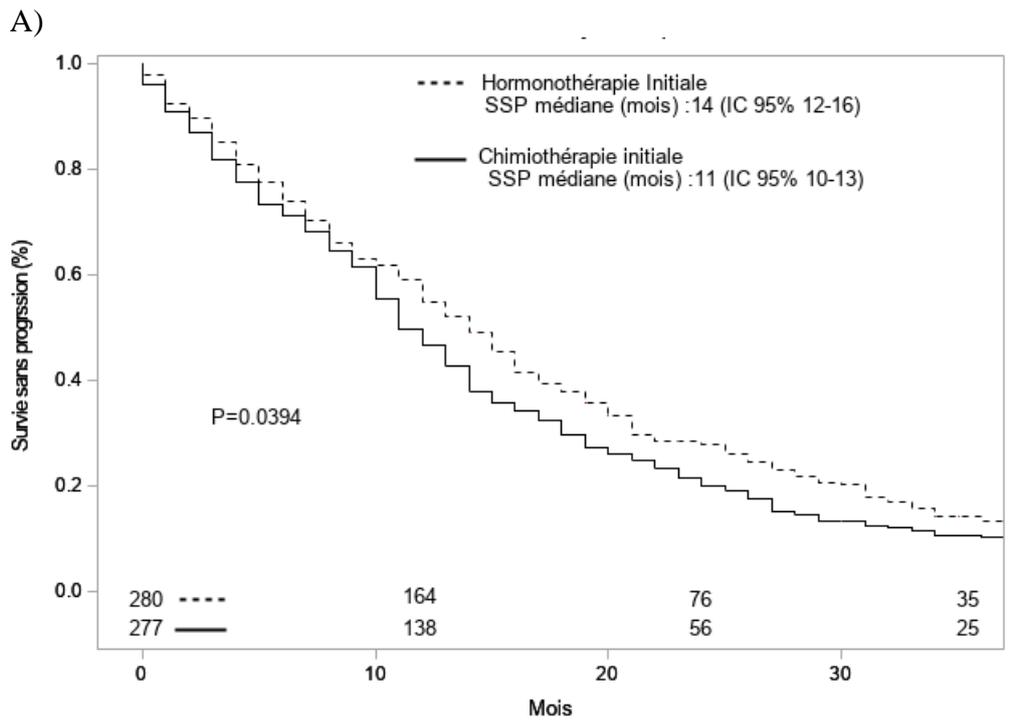
Parmi les patientes traitées initialement par de l'HT, 72% (203/280) ont été traitées avec des inhibiteurs de l'aromatase, 24% (66/280) avec des anti-œstrogènes (tamoxifène, fulvestrant) et 4% (11/280) avec une castration chimique ou chirurgicale.

Parmi les patientes traitées initialement avec une CT, 32% (88/277) ont reçu des traitements à base d'anthracycline et 60% (165/277) ont reçu des traitements à base de taxane. Au total, 53% (146/277) sont passés à une HT d'entretien, et parmi elles, 60% (88/146) l'ont fait dans les 6 mois suivant le début de la CT.

### 1.1.5.3 Les analyses de survie

La SSP et la SG ont été comparées chez les patientes qui ont reçu de l'HT ou de la CT comme traitement anti-métastatique systémique initial. La SSP médiane des patientes traitées par HT initialement était de 14 mois (IC à 95% 12-16 mois) et la SSP médiane de celles traitées par CT initialement était de 11 mois (IC à 95% 10-13 mois) ;  $p = 0,0394$  (figure 3A).

La SG médiane des patientes traitées par HT initialement était de 32 mois (IC à 95% 27-38 mois) et la SG médiane de celles traitées par CT initialement était de 28 mois (IC à 95% 23-32 mois) ;  $p = 0,2572$  (figure 3B).



**Figure 3. Survie sans progression et survie globale des patientes atteintes d'un cancer du sein métastatique de sous-type luminal en fonction du traitement systémique anti-métastatique initial.**

Les variables d'ajustement dans le modèle multivarié pour la SSP étaient le statut de la maladie métastatique (primaire ou secondaire), les sites métastatiques initiaux, le statut ménopausique, l'Indice de Comorbidité de Charlson et la résistance primaire à l'HT. La SSP était significativement meilleure chez les patientes traitées par HT initialement par rapport à celles traitées initialement avec une CT, avec un hazard ratio (HR) de 0,83 (IC à 95% 0,69-0,99). Les facteurs d'ajustement pour le modèle multivarié pour la SG étaient les suivants : l'âge au diagnostic de la maladie métastatique, le statut ménopausique, l'Indice de Comorbidité de Charlson, le statut de la maladie métastatique, les sites métastatiques initiaux et la résistance primaire à l'HT. Étant donné que la covariable « statut de la maladie métastatique » ne répondait pas à l'hypothèse des risques proportionnels, nous avons inclus une interaction entre cette covariable et le temps dans le modèle multivarié. La SG était significativement meilleure chez les patientes traitées avec une HT initialement que chez celles traitées initialement par une CT, avec un HR de 0,71 (IC à 95% 0,58-0,86) (tableau 2).

**Tableau 2. Analyses uni- et multivariées de la survie des patientes atteintes d'un cancer du sein métastatique à récepteurs hormonaux positifs et HER2 négatifs. Les valeurs en gras indiquent des résultats statistiquement significatifs.**

Caractéristiques	Survie Sans Progression				Survie Globale			
	Univariée HR (IC 95%)	<i>P</i>	Multivariée HR (IC 95%)	<i>P</i>	Univariée HR (IC 95%)	<i>P</i>	Multivariée <sup>a</sup> HR (IC 95%)	<i>P</i>
<b>Thérapie systémique initiale</b>		<b>0,0452</b>		<b>0,0472</b>		0,2623		<b>0,0004</b>
Chimiothérapie	1		1		1		1	
Hormonothérapie	0,84 (0,70-0,99)		0,83 (0,69-0,99)		0,90 (0,75-1,08)		0,71 (0,58-0,86)	
<b>Métastase primaire</b>		<b>&lt;0,0001</b>		<b>0,0004</b>		0,0765		-
Non	1		1		1		-	
Oui	0,69 (0,58-0,83)		0,71 (0,59-0,86)		0,85 (0,70-1,02)		-	
<b>Site de la métastase initial</b>		<b>0,0016</b>		<b>0,0158</b>		<b>0,0007</b>		0,1673
Osseux	1		1		1		1	
viscéral ou multiple	1,34 (1,12-1,60)		1,26 (1,04-1,51)		1,40 (1,15-1,69)		1,15(0,94-1,41)	
<b>Statut ménopausique</b>		0,1695		0,2323		<b>0,0001</b>		0,1434
Non ménopausée	1		1		1		1	
Ménopausée	1,16 (0,94-1,43)		1,15 (0,92-1,43)		1,55 (1,29-1,87)		1,23 (0,93-1,63)	
<b>Indice de comorbidités de Charlson</b>		<b>0,0368</b>		<b>0,0297</b>		<b>&lt;0,0001</b>		<b>0,0195</b>
SCC = 0	1		1		1		1	
SCC > 0	1,20 (1,01-1,42)		1,22 (1,02-1,47)		1,51 (1,23-1,84)		1,27 (1,04-1,55)	

<b>Hormonorésistance primaire</b>		<b>0,0215</b>		0,4450		0,0565		0,2596
Non	1		1		1		1	
Oui	1,32 (1,04-1,67)		1,10 (0,86-1,42)		1,27 (0,99-1,63)		1,17 (0,89-1,53)	
<b>Age au diagnostic de la métastase</b>		0,5689		-		<b>&lt;0,0001</b>		<b>0,0147</b>
< 65 ans			-		1		1	
≥ 65 ans	1,05 (0,89-1,25)		-		1,64 (1,36-1,98)		1,32 (1,06-1,65)	
<b>Type de tumeur primaire</b>		0,7039		-		0,8611		-
Canalaire	1		-		1		-	
Lobulaire	1,04 (0,84-1,30)		-		1,02 (0,81-1,29)		-	
<b>Année de diagnostic de la tumeur primaire</b>		0,2060		-		0,8779		-
< 2008	1		-		1		-	
≥ 2008	1,12 (0,94-1,34)		-		1,01 (0,84-1,22)		-	

<sup>a</sup> avec une interaction entre la variable “métastase primaire” et le temps

Dans les analyses avec score de propension (première analyse de sensibilité), les cas appariés (N = 296) ont montré un HR = 0,87 (IC à 95% 0,70-1,07) pour la SSP et un HR de 0,79 (IC à 95% 0,63-0,99) pour la SG. Le tableau 3 montre la comparabilité entre les deux groupes.

**Tableau 3. Distribution des caractéristiques des patientes appariées avec le score de propension selon le traitement systémique anti-métastatique initial.**

<b>Caractéristiques</b>	<b>Chimiothérapie Initiale N = 186 n (%)</b>	<b>Hormonothérapie Initiale N = 186 n (%)</b>	<b>P</b>
<b>Age au diagnostic de la métastase</b>			0,9170
< 65 ans	84 (45)	83 (45)	
≥ 65 ans	102 (55)	103 (55)	
<b>Statut ménopausique</b>			0,7861
Non ménopausée	34 (18)	32 (18)	
Ménopausée	152 (82)	154 (82)	
<b>Indice de comorbidités de Charlson</b>			0,7524
SCC = 0	110 (58)	107 (56)	
SCC > 0	76 (42)	79 (44)	
<b>Type de tumeur primaire</b>			0,7950
Canalaire	150 (81)	148 (80)	
Lobulaire	36 (19)	38 (20)	
<b>Métastase primaire</b>			0,5678
Oui	110 (59)	105 (56)	
Non	76 (41)	81 (44)	
<b>Site de la métastase initial</b>			0,8044
Osseux	60 (32)	62 (33)	
Viscéral	53 (29)	57 (31)	
Multiple	73 (39)	67 (36)	
<b>Hormonorésistance</b>			0,1161
Primaire	108 (58)	125 (67)	
Secondaire	28 (15)	27 (15)	
Hormonosensible	50 (27)	34 (18)	

*n*, nombre;

Dans la deuxième analyse de sensibilité, dans laquelle les patientes du groupe CT avec HT d'entretien étaient censurées, la SSP était significativement meilleure chez les patientes traitées avec HT initialement que chez celles traitées initialement par CT avec un HR de 0,72 (IC à 95% 0,56-0,92).

Dans la troisième analyse de sensibilité, dans laquelle l'HT était considérée comme une variable dépendante du temps, la SSP n'était pas significativement meilleure chez les patientes traitées avec l'HT initiale que chez celles traitées initialement par CT, avec un HR de 0,84 (IC à 95% 0,70-1,01).

### **1.1.6 Discussion**

Les résultats de cette étude en population ont montré que la SSP et la SG chez les patientes initialement traitées par HT étaient significativement améliorées par rapport aux patientes traitées par CT initialement (HR = 0,83 (IC à 95% 0,69-0,99) pour la SSP et HR = 0,71 (IC à 95% 0,58-0,86) pour la SG). Il convient de noter que la SSP dans le groupe HT initiale était comparable à celle des patientes avec CT initiale dans d'autres grands essais cliniques (13,27).

Afin de limiter le biais d'indication pouvant avoir influencé la première analyse, nous avons créé des scores de propension. Avec la méthode d'appariement, la seule différence entre les deux groupes, bien que plus petits, était le traitement initial. Cela a créé un cadre similaire à une étude randomisée, mais uniquement pour les variables sélectionnées pour créer le score de propension. Les résultats de cette analyse ont renforcé nos résultats principaux concernant les résultats de survie, car les groupes appariés étaient comparables et les HR étaient similaires, même si la puissance était inférieure, pour la SSP (HR = 0,87 ; IC à 95% 0,70-1,07) et pour la SG (HR = 0,79 ; IC à 95% 0,63-0,99).

Il existe deux études récentes sur le sujet, l'une réalisée par Lobbezzo et al. aux Pays-Bas (22) et une autre par Jacquet et al. en France (23). Dans l'étude de Lobbezzo et al., les patientes du groupe CT qui ont reçu également une HT d'entretien ont été censurées pour éviter l'effet de l'hormonothérapie d'entretien sur la SSP. Afin de nous comparer à cette étude, nous avons donc effectué une seconde analyse de sensibilité qui a également montré que la SSP était

significativement meilleure chez les patientes initialement traitées par HT (HR = 0,72 ; IC à 95% 0,56-0,92). Cependant, nos résultats ont également soutenu notre hypothèse selon laquelle cette analyse contenait un biais potentiel car la censure sur l'HT d'entretien était informative : dans le calcul de la SSP des patientes ayant reçu de la chimiothérapie initialement, elle était réduite par rapport à celles ayant reçu une HT d'entretien. En conséquence, la censure aurait pu détériorer la SSP dans le groupe ayant reçu de la CT initialement, favorisant ainsi le groupe HT lors de la comparaison de la SSP. Ce biais pourrait être plus fort dans notre étude puisque 53% de nos patientes ayant reçu de la CT initialement ont reçu également une HT d'entretien contre seulement 8% dans Lobbezzo et al. (22). En revanche, la SG médiane et la SSP médiane dans le groupe CT initiale étaient beaucoup plus faibles dans Lobbezzo et al. que dans notre étude. Cela pourrait s'expliquer par des différences dans les caractéristiques de base de nos populations. En effet, la différence de pourcentage de indiquerait un nombre plus important de patientes avec des tumeurs moins agressives et de meilleur pronostic dans notre étude. De plus, Lobbezzo et al. ont inclus des patientes avant l'introduction du bevacizumab contrairement à notre étude; l'utilisation du bevacizumab aurait pu augmenter la SSP dans le groupe CT initial (28).

Pour éviter le biais potentiel dû à la censure informative, nous avons donc également effectué une troisième analyse de sensibilité dans laquelle l'HT était considérée comme une variable dépendante du temps, sans censure sur l'HT d'entretien. Cette analyse nous a donné une estimation de l'effet de l'HT, en prenant en compte le moment de l'administration de l'HT, et a montré des résultats similaires à l'analyse primaire (HR = 0,84 ; IC à 95% 0,70-1,01). Cependant, dans cette analyse, l'introduction d'une nouvelle variable dépendante du temps a réduit la puissance, ce qui peut expliquer pourquoi ce résultat n'était pas statistiquement significatif. Bien sûr, cette analyse n'est pas parfaitement comparable à l'analyse primaire car nous avons non seulement considéré l'HT initiale mais également l'HT d'entretien.

Contrairement à notre étude, Jacquet et al. n'a trouvé aucune différence significative dans la SSP et la SG entre l'HT initiale et la CT initiale. Cependant, les deux populations ne sont pas comparables, en effet elles ne comprenaient que des maladies sensibles aux inhibiteurs de l'aromatase alors que nous incluons tous les patientes, même les hormonorésistantes.

Nos résultats ont également démontré que les traitements administrés aux patientes atteintes d'un cancer du sein métastatique en Côte d'Or étaient considérablement différents des traitements recommandés par les recommandations internationales (16–18). La moitié de notre population étudiée a reçu une CT initialement, même si seulement 15% de ces patientes présentaient une résistance clinique primaire à l'HT. La nature populationnelle de notre étude pourrait expliquer la forte proportion de CT initiale par rapport aux cohortes hospitalières étant donné que nous avons recruté des patientes dans différents types d'établissements médicaux. En outre, les premières recommandations internationales de consensus pour le cancer du sein métastatique ont été publiées relativement récemment, en 2012 (29). Dans notre étude, la proportion de patientes recevant une CT initialement a encore augmenté après 2008 malgré la publication de nouvelles lignes directrices.

Près des deux tiers des patientes qui sont passées à l'HT d'entretien l'ont fait dans les 6 mois suivant le début de la CT. Nous pouvons supposer que cela était en partie lié à la toxicité, mais nous n'avons pas pu obtenir d'informations spécifiques sur la toxicité ou les crises viscérales.

Nos données proviennent du registre du cancer du sein de Côte d'Or, qui respecte une politique stricte de contrôle de la qualité des données comprenant des contrôles réguliers pour assurer un suivi complet de chaque patiente. Nos données de haute qualité ne sont pas soumises au biais de sélection qui peut être observé avec les cohortes hospitalières. Cependant, le registre couvre une zone géographique relativement petite, ce qui nous oblige à étudier les données sur une longue période afin d'obtenir un échantillon suffisamment grand. Les procédures diagnostiques et les thérapies se sont améliorées depuis le début de l'étude. De plus, la classification des tumeurs a changé : nous avons utilisé la 5e édition de la classification AJCC TNM des tumeurs malignes avant 2008 et la 6e édition après 2008. Par conséquent, avant 2008, nous avons considéré quelques tumeurs métastatiques qui auraient été considérées comme N3 après 2008.

## **1.1.7 Conclusion**

En conclusion, dans cette étude en population, nous avons effectué une analyse primaire ainsi que trois analyses de sensibilité (score de propension, censure au début de l'HT d'entretien et considérant l'HT comme une variable dépendante du temps) pour surmonter les biais potentiels. Toutes ces analyses ont montré des résultats très similaires indiquant que le traitement initial par HT améliore la survie des patientes atteintes d'un cancer du sein métastatique RH positif/ HER2 négatif, par rapport au traitement initial par CT. Sachant que l'HT est reconnue comme moins toxique, nous suggérons que les recommandations internationales soient suivies de plus près et diffusées plus largement.

De plus, les résultats des premières études sur de nouvelles thérapies ciblées (inhibiteurs de CDK 4/6), qui sont chères mais moins toxiques, semblent montrer une capacité à prolonger la SSP (13,27). Cependant, les facteurs qui prédisent la réponse aux traitements n'ont pas encore été identifiés. D'autres études doivent être réalisées pour déterminer la meilleure stratégie de traitement des patientes atteintes d'un cancer du sein métastatique RH+/HER2-.

## **1.2. Avantages et inconvénients de données issues du registre des cancers du sein de Côte d'Or**

### **1.2.1 Les avantages :**

A travers cette étude portant sur les traitements du cancer du sein métastatique luminal, nous avons pu utiliser des données populationnelles exhaustives à l'échelle du département de Côte d'Or.

En effet, les données sont enregistrées pour chaque patient vivant dans le département de Côte d'Or au moment du diagnostic grâce à de multiples sources : dossiers médicaux, données du PMSI des établissements de santé prenant en charge les patients, comptes rendus anatomo-pathologiques des laboratoires, réunion de concertation pluridisciplinaire, Affection

Longue Durée (ALD). Les attachés de recherche clinique mènent un véritable travail d'enquête pour collecter les données et éviter les perdus de vue, lors d'un déménagement par exemple.

Ce suivi nécessite une excellente logistique pour éviter de perdre de vue les personnes concernées, mais également pour collecter l'ensemble des nombreuses variables pour chaque patient. En effet, concernant le registre des cancers du sein de Côte d'Or, plus de 150 variables sont collectées dans le cadre du diagnostic primitif en plus des traitements selon les évolutions qui ont pu avoir lieu. On y retrouve l'identité du patient ainsi que sa situation familiale et professionnelle. On y répertorie également l'état de son suivi (perdu de vue ou non), s'il est vivant ou décédé et les causes de décès le cas échéant. Le registre s'intéresse également aux antécédents médicaux, en particulier ceux qui sont connus pour être liés au cancer du sein (statut hormonal, contraception, parité, antécédent mammaire non malin, antécédents familiaux de cancer du sein, facteurs de risque génétiques, autres cancers ou pathologies,...). Il s'intéresse également aux caractéristiques de la tumeur primitive (sous-type, anatomo-pathologie, histologie, stade, base du diagnostic,...) ainsi qu'aux traitements (chirurgie, chimiothérapie, hormonothérapie, radiothérapie) et à l'évolution de la tumeur notamment les rechutes (récidive locale, régionale et/ou métastatique). Les données de biologie moléculaire pour les tumeurs primitives du sein et les rechutes sont également collectées.

Au total, le registre des cancers du sein de Côte d'Or est exhaustif et permanent dans sa prise en compte de l'ensemble des cancers du sein diagnostiqués chez des patients résidant en Côte d'Or. Ces données de grande qualité permettent d'établir précisément l'incidence et la prévalence du cancer du sein en Côte d'Or à un instant donné, et sont particulièrement utiles pour adapter et planifier les besoins en soins

La quantité et la qualité de ces données permettent d'évaluer la prise en charge des patients, du diagnostic du cancer du sein au décès, en ayant des informations détaillées concernant leur état de santé. Le suivi au cours du temps des patients permet également d'avoir des données concernant la durée de la maladie avant évolution, que celle-ci soit favorable (rémission, guérison) ou péjorative (progression de la maladie, décès). Le registre de cancer paraît donc être un outil de grande qualité pour effectuer des travaux concernant l'épidémiologie des cancers.

## 1.2.2 Les inconvénients :

Cependant, il existe plusieurs inconvénients à l'utilisation de bases de données issues de registre de cancer.

Premièrement, les registres de cancer sont dépendants des politiques publiques et des financements qui leur sont alloués. En effet, ces données individuelles, exhaustives et de grande précision nécessitent du personnel qualifié et un financement continu découlant d'une volonté politique d'amélioration de la prévention et de la prise en charge des personnes atteintes de cancer.

Deuxièmement, contrairement aux études prospectives, effectuer une étude sur un registre avec des données déjà collectées oblige les épidémiologistes à adapter la méthodologie de leur étude aux données disponibles dans le registre. Par exemple, le registre de cancer du sein de Côte d'Or, existe depuis 1982, mais n'a des données exhaustives concernant le statut des récepteurs hormonaux et HER2 qu'à partir de 1998. Par conséquent, nous avons limité notre inclusion à cette date.

Troisièmement, les connaissances concernant les cancers ont beaucoup évolué et continuent d'évoluer. Les classifications des cancers, les techniques d'imagerie et de diagnostics ainsi que les traitements évoluent. Lorsqu'une évolution apparaît dans la prise en charge des patients, le registre doit mettre des moyens logistiques et humains supplémentaires afin de récupérer l'information pour tous les patients et de modifier la base de données. Par exemple, notre étude ne prenait pas en compte le traitement ou non par des thérapies ciblées. Ces dernières, apparues au cours de cette décennie, influencent très probablement la survie des patientes atteintes d'un cancer du sein métastatique et devront donc être prises en compte dans les futures études. Ce nouveau traitement n'a pas pu être renseigné de manière exhaustive dans la base au moment de la réalisation de l'étude. Par conséquent, l'étude de cette nouvelle prise en charge par le registre a été retardée.

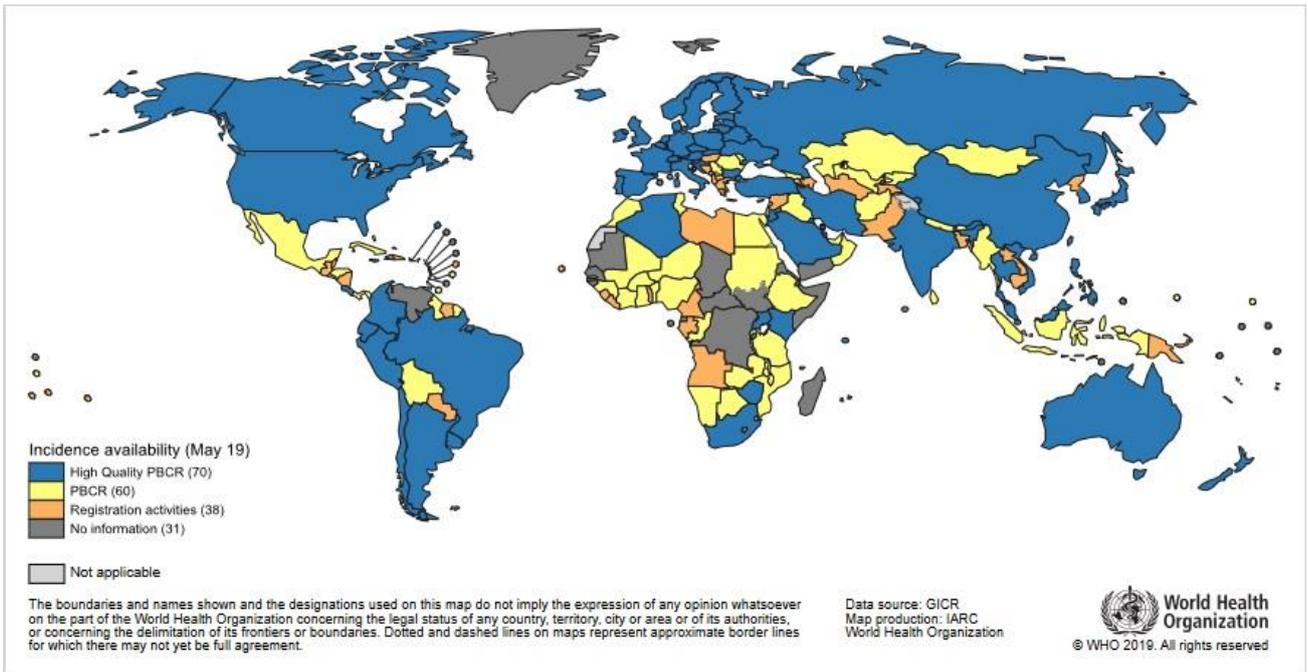
Quatrièmement, en France, les registres de cancer concernant l'adulte ne couvrent pas le pays entier et ne sont exhaustifs que sur une partie de territoire, une région ou un département. Les données épidémiologiques (incidence, prévalence, survie par exemple) concernant les cancers au niveau national ne sont alors que des estimations. Il est également important de noter que cela limite la puissance des études. En effet, le nombre de cas inclus dans les études est parfois limité, nécessitant la collaboration de plusieurs registres départementaux ou territoriaux afin d'avoir suffisamment de cas. Pour notre étude par exemple, nous avons utilisé uniquement

le registre de cancer du sein de Côte d'Or, mais nous avons choisi d'avoir une longue période d'inclusion de 19 années. Par conséquent, dans notre étude le stade du cancer a été déterminé en utilisant la cinquième (avant 2008) et la sixième édition (après 2008) de la classification TNM des tumeurs malignes. L'inclusion des cas sur une période importante a entraîné alors dans notre étude un biais de classification par rapport aux définitions les plus récentes de la maladie.

### **1.3. Exemples d'études à partir de registres de cancer dans le monde**

Ce chapitre a pour objectif d'élargir notre analyse à d'autres registres de cancer que celui que nous avons utilisé pour notre étude. Pour cela, nous avons cherché dans la littérature des exemples d'études récentes utilisant une base de données issue d'un registre de cancer. Nous souhaitons des études faites dans divers pays issus de continents différents afin d'illustrer les similitudes et les divergences pouvant exister entre les registres de cancer à travers le monde.

L'Union Internationale Contre le Cancer publie en 2019 une carte présentant la situation mondiale de l'enregistrement des cancers (cf. figure 4). Cette carte montre pour chaque pays, s'il est couvert nationalement ou régionalement par un registre de cancer en population de bonne qualité. On peut remarquer que, pour la grande majorité, les pays à revenu élevé sont mieux couverts que les pays à faible revenu argumentant le fait que les registres de cancer dépendent des politiques publiques et des financements qui leur sont alloués.



*High Quality PBCR : Registre de Cancer en Population de haute qualité ;*  
*Registration activities : activités d'enregistrement ;*  
*No information : Pas d'information ;*  
*Not Applicable : non applicable.*

**Figure 4. Carte présentant la situation mondiale de l'enregistrement des cancers par pays d'après L'Union Internationale Contre le Cancer publié en 2019**

Selon les règles internationales, les registres de cancers collectent plusieurs variables obligatoires nécessaires pour la surveillance épidémiologique des cancers : identification personnelle, sexe, date de naissance et commune de naissance, adresse complète au moment du diagnostic, date d'incidence, base de diagnostic la plus valide (histologie, paraclinique, ...), topographie (localisation) de la tumeur primitive, morphologie (type histologique), comportement tumoral et sources d'informations. Ils ont également plusieurs variables recommandées : date des dernières nouvelles, état aux dernières nouvelles (vivant/décédé), stade, grade/différenciation, traitement initial, mode de découverte. En ce qui concerne ces variables, tous les registres ne sont pas égaux.

### **1.3.1 Le registre de cancer de Hong-Kong**

Créé en 1963, le registre de cancer de Hong Kong est un registre qui collecte et analyse les données de tous les cas de cancer à Hong Kong, ce qui soutient la planification et l'évaluation des services de cancérologie à Hong Kong. Toutes les données brutes ont été validées par diverses procédures de recoupement via le système d'audit des cas de cancer conçu localement et examinées minutieusement grâce à de multiples processus de contrôle de la qualité conformes aux recommandations de l'IARC. En 2021, Co et al. (30) ont publié une étude concernant la survie à 10 ans des femmes après un diagnostic de cancer du sein canalaire in situ, à partir des données du registre de cancer de Hong Kong. Entre 1997 et 2006, 1 391 patientes atteintes d'un cancer du sein canalaire in situ ont été incluses dans le registre de Hong Kong, parmi elles seules 32 patientes (2.3%) ont été perdues de vue au cours du suivi.

Les forces de cette étude en population résident dans l'exhaustivité des diagnostics de cancer du sein pour l'ensemble du territoire de Hong Kong, leur permettant d'avoir une cohorte relativement importante et un faible taux de perdus de vue au cours du suivi. Cependant les auteurs de l'étude ont noté comme limite le manque de données concernant le décès qui n'est enregistré que pour un décès ayant pour cause un cancer du sein. Par conséquent, ils ont pu étudier la survie spécifique au cancer du sein et n'ont pas pu étudier la survie globale.

### **1.3.2 Le registre du cancer de la région Nord du Portugal**

Le registre du cancer de la région Nord du Portugal, créé en 1988, couvre une région d'environ 3,3 millions d'habitants, soit près d'un tiers de la population portugaise. Tous les cancers incidents survenant dans la région sont enregistrés, soit directement par les principaux hôpitaux publics via une plateforme web, soit sur la base de copies papier des rapports médicaux des hôpitaux privés et des laboratoires de pathologie. L'enregistrement suit les règles du CIRC. En 2017, Morais et al. (31) ont publié une étude portant sur le risque de second cancer primaire chez les patients ayant un premier cancer primaire de l'estomac. Un total de 7 427 cas de premier cancer primaire de l'estomac diagnostiqués entre 2000 et 2006 ont été inclus. Un suivi de ces cas a été effectué à l'aide des données du registre jusqu'au 31 décembre 2010 et a permis de retrouver 331 seconds cancers primaires.

Cette étude utilise un registre exhaustif concernant les diagnostics de cancer au niveau régional qui permet de repérer facilement les cas de second cancer primaire chez les patients ayant un premier cancer primaire de l'estomac. Ils ont pu prendre en compte l'âge, le sexe, la localisation de la tumeur et le statut vital (avec moins de 5% de perdus de vue pour cette donnée). Cependant les auteurs soulignent avoir été limités par les informations disponibles dans le registre. Ainsi, ils n'ont pas pu tenir compte de facteurs tels que les antécédents familiaux, le mode de vie (alimentation, alcool et tabagisme), les comorbidités et les traitements, dont certaines études ont montré qu'ils étaient associés au diagnostic d'un second cancer primaire. Cette étude est également limitée dans l'interprétation car les données du registre sont incomplètes sur le type histologique et le stade des cancers.

### **1.3.3 Le registre de Cancer d'Osaka au Japon**

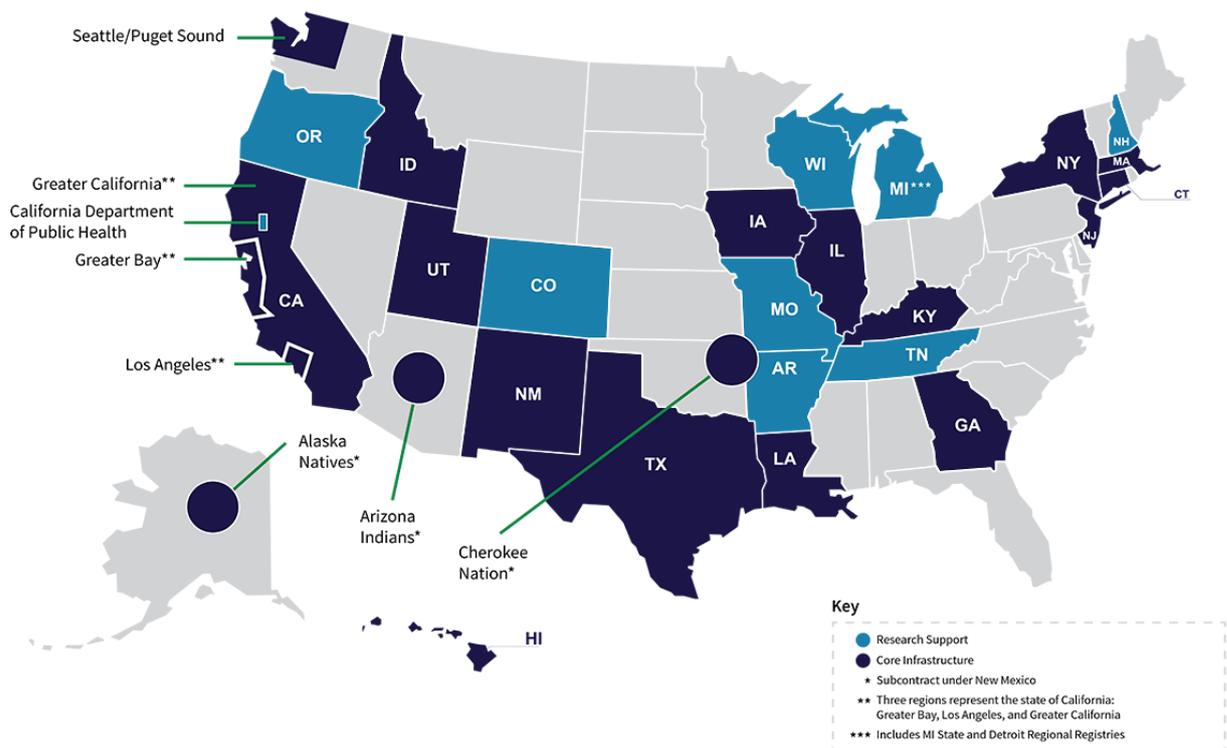
Le registre d'Osaka couvre une population de plus de 8,8 millions de personnes qui a été suivie pendant plus de 50 ans, ce qui en fait l'un des plus grands registres du cancer au Japon. En 2017, Kinoshita et al. (32) ont publié une étude portant sur les tendances de survie des patients diagnostiqués pour un cancer du poumon, en fonction du sexe. Ils ont analysé 79 330 cas du registre du cancer d'Osaka diagnostiqués entre 1975 et 2007 et calculé la survie relative à 5 ans pour six périodes (1975-1980, 1981-1986, 1987-1992, 1993-1997, 1998-2002 et 2003-2007). La proportion de cas enregistrés uniquement sur certificat de décès était de 19,8% dans le registre sur l'ensemble de la période observée. Ils ont exclu de l'analyse les données enregistrées uniquement sur certificat de décès, en raison du manque d'informations pour l'analyse de survie. Ils ont pu prendre en compte l'âge, le sexe, le type histologique et le stade du cancer, et les traitements (radiothérapie, chimiothérapie, chirurgie).

Cependant le registre ne collecte pas d'informations concernant plusieurs facteurs influençant la survie comme le statut tabagique, les caractéristiques socio-économiques ou les mutations génétiques. De plus, le type histologique et le stade du cancer n'étaient pas systématiquement indiqués dans les premières périodes de l'étude. Ces informations leur auraient permis de mieux comprendre les différences de survie entre les hommes et les femmes.

### **1.3.4 Le registre "Surveillance, Epidemiology, and End Results" (SEER) aux Etats-Unis.**

En 1971, le congrès américain a adopté une loi nationale sur le cancer (National Cancer Act), qui prévoyait la collecte, l'analyse et la diffusion de données utiles à la prévention, au diagnostic et au traitement du cancer. Ce mandat a conduit à la création du programme SEER du National Cancer Institute. Le programme SEER fournit des statistiques sur l'incidence du cancer et la survie dans la zone couverte (Figure 5), ainsi que sur la mortalité aux États-Unis, il suit les tendances de l'incidence du cancer dans les groupes de population géographiques et démographiques, y compris les divers groupes ethniques. Il fournit des informations détaillées concernant l'étendue de la maladie au moment du diagnostic, la thérapie et la survie des patients, il répond aux préoccupations et aux demandes de renseignements du public sur le cancer, il développe de nouvelles méthodes statistiques, des modèles et des logiciels pour l'analyse et la

présentation de statistiques nationales et régionales. Le programme SEER collecte et publie des données sur l'incidence du cancer et la survie dans des registres de cancer en population dans 19 zones géographiques des États-Unis. Les registres SEER fournissent une couverture complète pour les régions métropolitaines représentant 28 % de la population nationale. Cela correspond à plus de 11 millions de cas de cancers in situ et invasifs et plus de 516 000 nouveaux cas ajoutés chaque année dans les zones couvertes par le SEER. (33)



**Figure 5. Carte des États-Unis montrant les zones couvertes par le programme SEER (Surveillance, Epidemiology, and End Results) en 2021 (en bleu foncé) disponible sur <https://seer.cancer.gov/registries/>**

En 2020, Lin et al. (34) ont publié une étude portant sur l'avantage de la radiothérapie adjuvante sur la survie globale dans les cas de résection de cancer du pancréas de stade I à II en utilisant la base de données du registre SEER. Ils ont inclus 3 219 cas de résection de cancer du pancréas de stade I à II entre 2010 et 2015. Les auteurs ont pris en compte l'âge, le sexe, l'origine ethnique, le statut d'assurance santé, le site de la tumeur primaire, le stade du cancer, la taille de la tumeur, et les traitements (radiothérapie, chimiothérapie, type de chirurgie).

Le registre SEER permet d'étudier un grand échantillon de population et collecte des données détaillées sur les caractéristiques du cancer. Cependant, cette base de données n'offre pas d'informations complètes et détaillées sur l'expérience thérapeutique (les marges opératoires, ou les types, schémas et moment de la chimiothérapie ou de la radiothérapie par exemple), et sur la qualité de vie des patients qui jouent un rôle important sur la survie globale.

### **1.3.5 Le registre national du cancer du Pays-Bas**

Le registre national du cancer du Pays-Bas comprend des informations sur les patients atteints d'un cancer nouvellement diagnostiqué aux Pays-Bas : le diagnostic du cancer, le stade de la tumeur, le site de la tumeur, son histologie, les comorbidités au moment du diagnostic et les traitements reçus. Le statut vital des patients est obtenu en établissant un lien avec la base de données des registres municipaux, qui enregistrent tous les décès aux Pays-Bas.

En 2020, Rovers et al. (35) ont publié une étude portant sur la comparaison de la survie globale entre la chimiothérapie adjuvante et la surveillance active après résection frontale de métastases colorectales péritonéales isolées et synchrones. Pour ce travail, ils ont utilisé le registre national du cancer du Pays-Bas afin d'inclure 393 patients ayant reçu une résection frontale comme traitement de leur métastases colorectales péritonéales isolées et synchrones.

Cependant, le registre ne permet pas d'avoir de données concernant les morbidités post-opératoires. Cela peut être une source de biais en raison de son association potentielle avec la performance postopératoire, la probabilité de recevoir une chimiothérapie adjuvante et la survie globale. Il ne collecte également pas de données concernant le « peritoneal cancer index » qui évalue l'extension métastatique dans la cavité péritonéale et qui peut être associée à la survie. Le registre ne détaille pas les traitements adjuvants ou de la récurrence ainsi que leur durée ; les auteurs invitent à utiliser des bases de données plus détaillées concernant les traitements pour de futures études sur le sujet.

### **1.3.6 Synthèse**

Au niveau mondial, des registres nationaux et infranationaux se côtoient selon les pays. Malgré les règles internationales, certaines données sont variables d'un registre à l'autre. Elles peuvent être incluses ou exclues en fonction des intérêts locaux particuliers, en tenant compte des considérations sur la disponibilité des informations dans les sources de données. Certaines données directement ou indirectement liées aux cancers sont rarement présentes alors qu'elles sont pertinentes pour les épidémiologistes : les habitudes de vie, les déterminants socio-économiques, les caractéristiques génétiques et la qualité de vie.

## **2. Données médico-administratives**

### **2.1. L'exemple du SNDS**

#### **2.1.1 Présentation et description**

##### **2.1.1.1 Intérêt du SNDS**

Le législateur a souhaité ouvrir l'accès aux données de santé collectées par les personnes publiques afin que « leurs potentialités soient utilisées au mieux dans l'intérêt de la collectivité ». Le SNDS répond à cet objectif en mettant à disposition des données de santé, afin de contribuer :

- à l'information sur la santé, l'offre de soins, la prise en charge médico-sociale et leur qualité ;
- à la définition, à la mise en œuvre et à l'évaluation des politiques de santé et de protection sociale ;
- à la connaissance des dépenses de santé, d'assurance maladie et médico-sociales ;
- à l'information des professionnels, des hôpitaux ou médico-sociaux sur leur activité ;
- à la surveillance, à la veille et à la sécurité sanitaires ;
- à la recherche, aux études, à l'évaluation et à l'innovation dans les domaines de la santé et de la prise en charge médico-sociale.

La Loi interdit l'utilisation des données contenues dans ce fichier à des fins de promotion des produits de santé et à des fins d'exclusion de garanties des contrats d'assurance ou la modification des cotisations et des primes d'assurance. (36)

## 2.1.1.2 Les informations contenues dans le SNDS

Le SNDS rassemble et met à disposition les bases de données qui existaient, jusqu'alors, indépendamment :

- La base SNIIRAM contenant les données de l'assurance maladie ;
- La base PMSI contenant les données issues de l'activité des établissements de santé ;
- La base CepiDC (Centre d'épidémiologie sur les causes médicales de décès), gérée par l'Institut National de la Santé Et de la Recherche Médicale (INSERM), contenant les données sur les causes de décès ;
- Les données liées au handicap issues des maisons départementales des personnes handicapées
- Des données provenant des « complémentaires santé » (mutuelles par exemple) figurent également dans le SNDS.

Le SNDS ne contient aucune donnée directement identifiante concernant les bénéficiaires (pas de noms/prénoms ou numéro de sécurité sociale, ni d'adresse postale).

Les données contenues dans le SNDS sont détaillées à l'article R. 1461-4 du Code de la santé publique. Il s'agit des informations relatives :

### **-aux bénéficiaires de soins :**

- pseudonyme ou code pour chaque bénéficiaire ;
- sexe, mois et année de naissance, rang de naissance et lieu de résidence ;
- informations médico-administratives, y compris le cas échéant celles liées aux ALD et maladies professionnelles ;
- informations liées au décès : date, lieu, causes et circonstances, situation familiale et profession à la date du décès.

### **-aux organismes d'assurance maladie obligatoire et, le cas échéant, complémentaire :**

- identification des organismes ;
- caractéristiques de la prise en charge de ces organismes.

**-aux prestations prises en charge par les organismes, associées à chaque bénéficiaire, notamment :**

- soins de « ville » : nature des actes, actes de pharmacie, dispositifs médicaux, consultations des professionnels de santé, examens biologiques, date de grossesse, etc. ;
- séjours à l'hôpital, y compris accueil aux urgences ;
- montant des actes ou prestations, tarif appliqué et part prise en charge par l'assurance maladie obligatoire, date de remboursement etc. ;
- type de contrat souscrit avec la complémentaire santé (le cas échéant).

**-aux professionnels de santé, notamment :**

- numéro d'identification (ces numéros sont conservés séparément des autres données du SNDS) ;
- profession, sexe, date de naissance ;
- lieu de réalisation de l'acte.

**-aux personnes en situation de handicap** (informations médico-sociales liées au handicap et à sa prise en charge, décisions de la commission des droits et de l'autonomie des personnes handicapées).

**-des informations relatives aux arrêts de travail et aux prestations fournies** (arrêt maladie, maternité, accidents du travail, etc.). (36)

Les données sont conservées pendant 19 ans en plus de l'année au cours de laquelle elles ont été recueillies. Elles sont ensuite archivées pendant 10 ans.

### 2.1.1.3 Les bases composant le SNDS

#### 2.1.1.1.1 Le SNIIRAM

Le SNIIRAM a été créé pour contribuer à la connaissance des dépenses de l'ensemble des régimes d'Assurance Maladie, à la définition, à la mise en œuvre et à l'évaluation des politiques de santé, à l'amélioration de la qualité des soins et à la transmission aux professionnels de santé des informations relatives à leur activité, à leurs recettes et, s'il y a lieu, à leurs prescriptions.

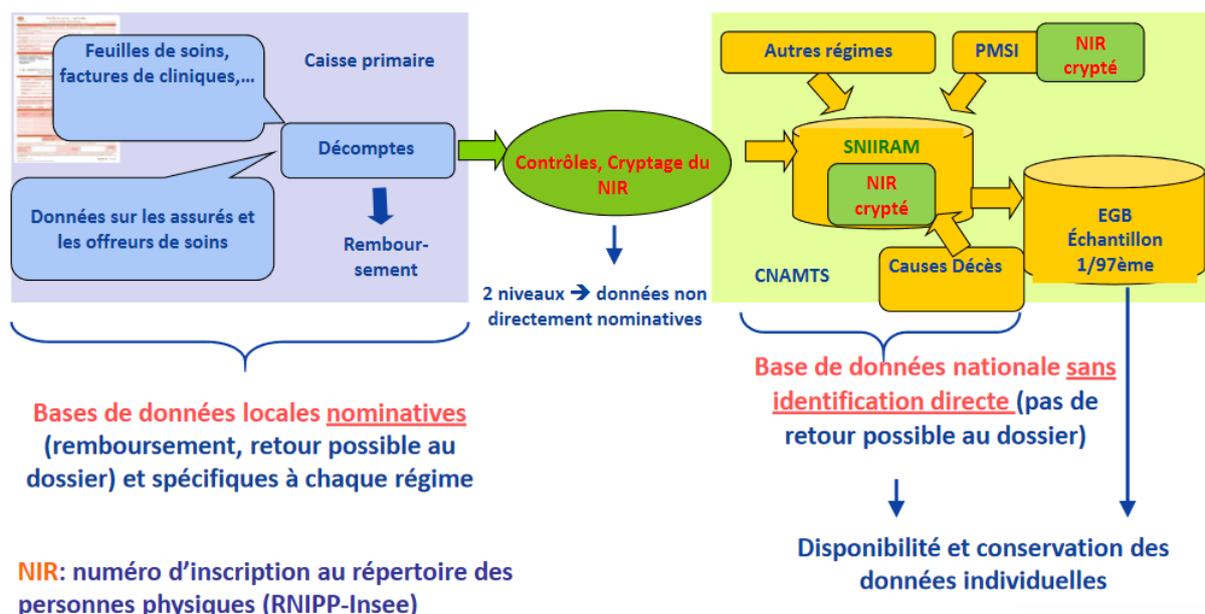
Le SNIIRAM est un entrepôt de données anonymes regroupant les informations issues des remboursements effectués par l'ensemble des régimes d'assurance maladie pour les soins du secteur libéral.

Les informations concernant le bénéficiaire sont l'âge, sexe et lieu de résidence (département et commune), ainsi qu'éventuellement la notion de bénéficiaire de la couverture maladie universelle, d'ALD ou de maladie professionnelle et, le cas échéant, la date de décès. Concernant les professionnels de santé qui ont délivré les soins et éventuellement les prescripteurs le SNIIRAM comporte l'âge, sexe et lieu d'exercice ainsi que la spécialité et le mode d'exercice. Sur la consommation de soins la base contient les dates de soin ainsi que les montants remboursés par l'assurance maladie et payés par le patient. Elle comporte également le codage détaillé des médicaments délivrés, des actes techniques des médecins, des dispositifs médicaux, des prélèvements biologiques.

Les données du SNIIRAM ont été collectées et organisées progressivement depuis 2002. Aujourd'hui, trois ensembles de restitution sont mis en service :

- 15 bases de données thématiques de données agrégées appelées datamarts orientés vers une finalité particulière : suivi des dépenses (Dépenses d'Assurance Maladie Inter-Régimes ou DAMIR), analyse de l'offre de soins libérale, biologie, pharmacie, dispositifs médicaux, établissements privés ;
- Un échantillon général des bénéficiaires (EGB) au 1/97e de la population protégée : l'EGB permet de réaliser des études longitudinales et d'analyser le parcours individuel de près de 660 000 bénéficiaires en ville et à l'hôpital ;
- Une base de données individuelles des bénéficiaires (Datamart de Consommation Inter Régime ou DCIR) pour réaliser des études sur la consommation des soins.

Pour chaque régime national d'assurance maladie, les données sont collectées localement à partir de la carte vitale, des feuilles de soins, des factures des établissements privés et des actes et consultations externes facturés par les hôpitaux publics. À ce niveau, les données sont inévitablement liées au nom du patient pour assurer le remboursement grâce au Numéro d'Inscription au Répertoire des personnes physiques (NIR). Les données sont alors soumises à un cryptage du NIR puis sont centralisées au niveau du SNIIRAM. Les circuits d'information du SNIIRAM sont illustrés dans la figure 6.



**Figure 6. Les circuits d'information du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM). D'après Tuppin et al. (37)**

#### 2.1.1.1.2 Le PMSI

L'objectif principal du PMSI est de procéder à l'analyse de l'activité médicale des établissements hospitaliers à des fins d'allocation budgétaire. Dans le cadre du PMSI, tout séjour dans un établissement de santé, public ou privé, fait l'objet d'un recueil systématique et minimal d'informations administratives et médicales qui sont utilisées principalement pour le financement des établissements de santé (tarification à l'activité) et pour l'organisation de l'offre de soins (planification). C'est l'Agence Technique de l'Information sur l'Hospitalisation (ATIH) qui gère le PMSI.

Les informations recueillies au niveau de l'établissement sont ensuite centralisées au niveau national sur la forme d'un résumé de sortie anonyme. Chaque résumé de sortie anonyme contient des informations médicales (les diagnostics reliés, les actes médicaux réalisés, etc.) ainsi que des informations de type administratif (identification de l'établissement, durée de séjour, mode d'entrée et de sortie dont, éventuellement, le décès) et sur le patient (sexe, âge, code géographique de résidence basé sur le code postal de résidence). Outre les résumés de

sortie anonyme, les établissements doivent produire des informations sur le nombre de consultations et d'actes externes réalisés, sur la consommation de certains médicaments et dispositifs médicaux implantables (prothèses, implants) utilisés, et sur les prélèvements d'organes sur les personnes décédées.

Le recueil systématique couvre l'ensemble des hospitalisations en médecine, chirurgie et obstétrique, en soins de suite et réadaptation, en psychiatrie et les hospitalisations à domicile.

#### 2.1.1.1.3 La Base des Causes Médicales de Décès (BCMD)

Depuis 1968, l'INSERM est chargé d'élaborer annuellement la statistique nationale des causes médicales de décès en collaboration avec l'Institut National de la Statistique et des Etudes Economiques (INSEE). Cette statistique est établie à partir des informations recueillies dans le certificat de décès.

La finalité première de ce recueil est d'élaborer la statistique des causes médicales de décès permettant ainsi d'étudier l'évolution de la mortalité par cause médicale en France et de faire des comparaisons internationales. Cette base est utilisée également pour l'étude des disparités socio-spatiales, pour l'analyse de la mortalité évitable ou de la mortalité prématurée par exemple.

Le contenu du certificat de décès, pour la partie sur les causes de décès, répond à des recommandations internationales de l'Organisation Mondiale de la Santé (OMS). Le certificat comprend les maladies ou affections morbides à l'origine du décès ainsi que d'autres facteurs y ayant contribué. L'identification de la cause initiale de décès, celle qui est à l'origine du décès, est liée à la déclaration du médecin et à des règles internationales de classification.

Outre les causes médicales, d'autres informations complémentaires, comme le lieu du décès, la notion de grossesse, sont également collectées. Des informations relatives au défunt comme l'âge, le sexe, la date de décès, le département et la commune de résidence et de décès complètent la base. (38)

## **2.1.2 Etude de la relation entre le diabète gestationnel et cancer du pancréas**

### **2.1.2.1 Contexte et questions de recherche**

#### **2.1.2.1.1 Le cancer du pancréas : épidémiologie**

En 2020, il a été comptabilisé 495 773 nouveaux cas de cancer du pancréas (CP) et 466 003 décès causés par le CP dans le monde (39). En France, les données de Santé Publique France comptabilisent 14 184 nouveaux cas pour 11 456 décès pour 2018 (40). Le CP est l'un des cancers dont le taux de mortalité est le plus élevé dans les pays à revenu élevé, avec une survie nette de 6 à 9 % à 5 ans (41–43). Ce taux de mortalité élevé est principalement attribué à un diagnostic tardif, car la plupart des patients atteints de CP restent asymptomatiques jusqu'à ce que la maladie atteigne un stade avancé. L'identification des facteurs de risque est donc importante pour la création de politiques de santé publique appropriées.

De nombreuses études ont rapporté que le risque de CP augmente d'environ 80 % en cas de diabète de type 2 (44–48). Plusieurs études récentes ont fourni des hypothèses qui pourraient expliquer la physiopathologie du lien entre le diabète et le risque de CP :

-l'hyperinsulinémie : lorsque les récepteurs de l'insuline et de l'insulin-like growth factor interagissent avec leurs ligands, de multiples voies de signalisation sont activées, conduisant à la phosphorylation de protéines adaptatrices. Ces voies de signalisation peuvent favoriser la prolifération et peuvent affecter l'invasion, les métastases et la protection contre les stimuli apoptotiques,

-l'hyperglycémie : un taux de glucose circulant plus élevé peut favoriser la croissance des cellules malignes. De plus, l'hyperglycémie insulino-résistante peut causer des dommages à l'ADN, stimuler l'activité invasive et migratoire des cellules cancéreuses. Elle peut également atténuer l'activité des enzymes antioxydantes et créer à son tour un état de stress oxydatif.

-l'inflammation chronique : Les cytokines inflammatoires, les radicaux libres et les médiateurs des voies inflammatoires sont associés à l'expression d'oncogènes, à l'extinction de gènes suppresseurs de tumeurs, à l'augmentation du nombre de tumeurs et à l'altération du cycle cellulaire, ce qui peut faciliter la carcinogenèse pancréatique. (49–52)

### 2.1.2.1.2 Le diabète gestationnel : épidémiologie

La prévalence du diabète gestationnel (DG) est d'environ 8% en France en 2012 contre 3.8% en 2004 (53).

Comme le diabète, le DG peut également être associé à des pathologies ultérieures. En effet, plusieurs études ont montré une association entre DG et diabète de type 2 ou maladies cardiovasculaires (54–63).

Non seulement les femmes atteintes de DG présentent un risque accru de diabète de type 2 dans les années qui suivent l'accouchement, mais le DG partage également des mécanismes d'action avec le diabète de type 2, notamment une hyperglycémie due à un dysfonctionnement des cellules  $\beta$  pancréatiques sur fond de résistance chronique à l'insuline.

#### 2.1.2.1.1 Lien entre diabète gestationnel et cancer du pancréas

Plusieurs études ont exploré l'association entre le DG et le cancer (64–70). Toutefois, la relation entre le DG et le CP n'a fait l'objet que de trois études. Deux études ont été réalisées en Israël par Perrin et al. et Sella et al. respectivement en 2007 et 2011 (71,72). La première a montré que le DG augmente le risque de CP ultérieur, avec un HR égal à 7,10 et un IC à 95% large (2,80–18,00). De même, la seconde étude a montré un HR égal à 7,06 et un IC à 95% encore plus large (1,69-29,45). Une troisième étude de Peng et al., datant de 2019, basée sur les données de l'assurance maladie nationale de Taiwan a montré que les antécédents de DG n'étaient pas associés de manière significative à un risque de CP ultérieur avec un HR de 1,07 (IC 95% 0,66 1,76) (73). Une méta-analyse récente (2020), incluant les trois études précédentes, a conclu que le DG n'était pas associé à un risque statistiquement significatif de CP (74). Les résultats des trois études doivent être pris avec précaution, car le diagnostic de DG n'était pas basé sur des critères récents dans les deux études les plus anciennes et le suivi était relativement court pour certaines patientes dans l'étude la plus récente.

#### 2.1.2.1.2 Objectif

Notre objectif était d'étudier la relation entre le DG et la survenue d'une hospitalisation mentionnant un CP chez les femmes ayant accouché en France de janvier 2008 à décembre 2009, à partir des données du SNDS.

## 2.1.2.2 Matériel et méthodes

### 2.1.2.2.1 Schéma d'étude

Il s'agit d'une étude de cohorte rétrospective, en population, qui intéresse les femmes ayant accouché en France entre janvier 2008 et décembre 2009. A partir de la base de données du SNDS, un suivi longitudinal de 8 ans a été effectué pour toutes les femmes.

### 2.1.2.2.2 Population d'étude

➤ Critères d'inclusion :

Les femmes âgées de 14 à 55 ans qui ont accouché entre le 1er janvier 2008 et le 31 décembre 2009 ont été identifiées par les codes diagnostic Z37 comme diagnostic associé et/ou les procédures d'accouchement de la classification commune des actes médicaux (CCAM) dans les résumés de sortie standardisés (RSS). La date d'inclusion dans la cohorte était la date du premier accouchement pendant la période d'inclusion de deux ans.

➤ Critères de non-inclusion et d'exclusion :

Lors de la sélection de la population étudiée, nous avons exclu toutes les femmes ayant des antécédents de cancer du pancréas (codes C25 en diagnostic principal, associé ou relié dans le RSS) pendant la grossesse et dans l'année précédant la grossesse. Nous avons également exclu les femmes ayant un antécédent de diabète de type 1 ou 2, qui ont été identifiées en utilisant : les codes E10-E14 ou O24.0-O24.3 comme diagnostic principal, associé ou relié dans les RSS et/ou s'il a été notifié en ALD et/ou si au moins trois médicaments antidiabétiques ou de l'insuline (classe A10 de la classification anatomique, thérapeutique et chimique [ATC]) ou deux en cas d'au moins une grande quantité ont été remboursés dans l'année précédant la grossesse (75).

Par ailleurs, les femmes ayant eu un diagnostic de CP dans les deux ans suivant la grossesse ont été exclues de la population analysée. Nous avons pris cette précaution en référence à d'autres études explorant le lien entre le diabète et le CP que nous avons appliqué au DG. En effet, le diabète diagnostiqué peu avant le CP peut être indicatif et, d'une certaine manière, un symptôme du CP. Plusieurs hypothèses existent pouvant expliquer cela : soit le diabète est simplement une conséquence de la destruction glandulaire par la tumeur et donc une manifestation tardive du cancer, soit le diabète est induit par le biais de produits sécrétés par la tumeur. La résolution du diabète et l'amélioration de la tolérance au glucose dans certains cas après la résection de la tumeur soutiennent cette dernière hypothèse (76). Le CP peut alors être préexistant au diabète mais diagnostiqué plus tard.

#### 2.1.2.2.3 L'exposition

L'exposition a été définie comme la présence d'un DG pendant une grossesse au cours des huit années de suivi depuis la grossesse d'inclusion. Si une femme a eu plus d'une grossesse, les informations recueillies dans les RSS pour chaque grossesse ont été prises en compte. Si un code de DG (O24.4 ou O24.9) était identifié pour au moins une de ces grossesses, la femme était considérée comme ayant un DG. Pour tenir compte du fait que le DG peut survenir à n'importe quel moment du suivi (soit lors de la première grossesse, soit lors des grossesses suivantes), nous l'avons considéré comme une variable dépendante du temps (c'est-à-dire que l'exposition n'a été prise en compte qu'à partir de la première fois où le DG a été enregistré). Cela nous a permis de considérer plus précisément l'intervalle de temps entre le diagnostic de DG et l'apparition du CP. Les femmes classées dans le groupe "sans DG" pendant toute l'étude étaient celles pour lesquelles un code de DG n'était mentionné dans aucun des RSS au cours des 8 années de suivi.

#### 2.1.2.2.4 Critère de jugement principal

Le critère de jugement principal est établi en recherchant une hospitalisation mentionnant un code de CP par la présence d'un des codes C25 en diagnostic principal, diagnostic associé ou diagnostic relié dans les RSS. Nous n'avons bien sûr retenu que la première hospitalisation mentionnant un CP, afin de prendre en compte plus précisément le délai entre le diagnostic du DG et le diagnostic du CP. Toutefois, les femmes pour lesquelles a été retrouvée une hospitalisation mentionnant un CP dans les deux ans qui ont suivi la grossesse ont été exclues de l'analyse. En effet, comme expliqué dans la description de notre population d'étude, le diabète diagnostiqué peu avant le CP peut être indicatif et, d'une certaine manière, un symptôme de ce CP.

#### 2.1.2.2.5 Variables explicatives

Les variables explicatives suivantes ont été prises en compte : l'âge maternel, le diabète de type 2 ultérieur et la consommation de tabac. L'âge maternel lors de l'accouchement a été catégorisé ainsi : <26, 26-29, 30-33 et  $\geq$ 34 ans. Le diabète de type 2 ultérieur au cours du suivi était défini par la présence d'un des codes CIM-10 E10 à E14 comme diagnostic principal, associé ou relié dans les RSS et/ou si le diabète ultérieur était notifié en ALD et/ou s'il y avait un remboursement d'au moins trois antidiabétiques ou insuline (classe A10 de la classification ATC à l'exception du benfluorex) ou deux remboursements s'il y avait au moins un gros conditionnement (75). La consommation de tabac dans l'année précédant la grossesse, à l'inclusion ou pendant le suivi a été identifiée à l'aide des codes CIM-10 F17 (dépendance au tabac), T65.2, Z58.7, Z72.0 (effets du tabagisme) ou Z71.6 (sevrage tabagique) dans les RSS et/ou d'au moins un remboursement de varénicline. La varénicline est un médicament agoniste partiel des récepteurs nicotiques cérébraux indiqué dans le sevrage tabagique. Ce médicament est remboursé par l'Assurance Maladie (sauf entre 2011 et 2016), et délivré uniquement sur prescription médicale. Cela nous permet de retrouver ces remboursements dans le SNIIRAM par ces codes d'identification de spécialité pharmaceutique.

La durée (en jours) entre l'accouchement (avec ou sans DG) et l'apparition du CP a été déterminée à partir des données d'hospitalisation.

### 2.1.2.3 Analyses statistiques

Les caractéristiques des femmes définies comme des variables catégorielles ont été décrites pour les deux groupes (avec antécédents de DG et sans antécédents de DG) en utilisant des distributions de fréquence. Les deux groupes ont été comparés à l'aide de tests de  $\chi^2$ . Les variables continues, notamment les durées, ont été décrites à l'aide de moyennes et d'écart-types ( $\sigma$ ).

Le risque d'hospitalisation pour CP a été étudié à l'aide de modèles de Cox. Le début du DG et l'hospitalisation mentionnant un diabète de type 2 ultérieur ont été considérés comme des variables dépendantes du temps. Nous avons développé un premier modèle incluant l'âge, le DG et le diabète de type 2 ultérieur.

Nous souhaitons ensuite inclure dans un nouveau modèle la consommation de tabac, qui est un facteur de risque reconnu de CP. Or, dans le SNDS nous ne retrouvons pas précisément les moments auxquels les femmes auraient débuté ou stoppé leur consommation de tabac. Nous retrouvons parfois dans les RSS des mentions de tabagisme lors des hospitalisations. Nous pouvons également supposer qu'une femme à qui l'on prescrit un traitement de sevrage tabagique a fumé du tabac auparavant. Nous avons développé un second modèle qui ajustait les mêmes covariables avec l'ajout de la consommation de tabac comme variable dépendante du temps. Dans ce second modèle, la consommation de tabac n'a été prise en compte qu'à partir de la première fois où l'un des codes diagnostiques correspondants ou un remboursement de varénicline a été enregistré. Cependant, il nous est impossible de savoir depuis combien de temps elle fumait ou si le sevrage a été un succès. Comme cette approche peut être considérée comme discutable au vu des éléments suscités, nous avons effectué également une analyse de sensibilité sur ce modèle, où la consommation de tabac a été considéré comme une variable fixe. Dans cette analyse, la présence de la consommation de tabac a été déterminée à partir du moment de l'inclusion, indépendamment du moment où le tabagisme a été enregistré dans l'année précédant la grossesse et pendant la période de suivi. Cette analyse nous permettait alors de considérer comme fumeuses les femmes identifiées dans le SNDS durant la période d'étude, quel que soit le moment où la consommation de tabac a été identifiée dans le SNDS.

Les HR et les IC à 95% ont été calculés dans chaque modèle. Nous avons également exploré les interactions entre l'âge ou la consommation de tabac et le DG.

Enfin, l'obésité (Indice de Masse Corporelle (IMC) >30), qui serait une variable associée au risque de DG et au risque de CP, n'a pas pu être incluse dans notre modèle. Malheureusement, l'IMC avant la grossesse n'est que peu documenté car cette donnée n'est disponible que s'il existe une hospitalisation avant la grossesse, et nous craignons que l'IMC pendant la grossesse ne puisse pas être considéré comme aussi fiable. Par conséquent, nous avons choisi d'effectuer une analyse évaluant la « e-value » (77). C'est une mesure de la robustesse d'une association aux facteurs de confusion potentiels non contrôlés (dans notre cas : l'obésité). La e-value est définie donc ici comme la force minimale de l'association que l'obésité non mesurée devrait avoir avec le DG et le CP pour expliquer entièrement l'association DG-CP que nous avons retrouvée. Elle est donc calculée à partir du HR retrouvé dans notre étude selon la formule suivante :

$$E\text{-value} = HR + \sqrt{HR \times (HR - 1)}$$

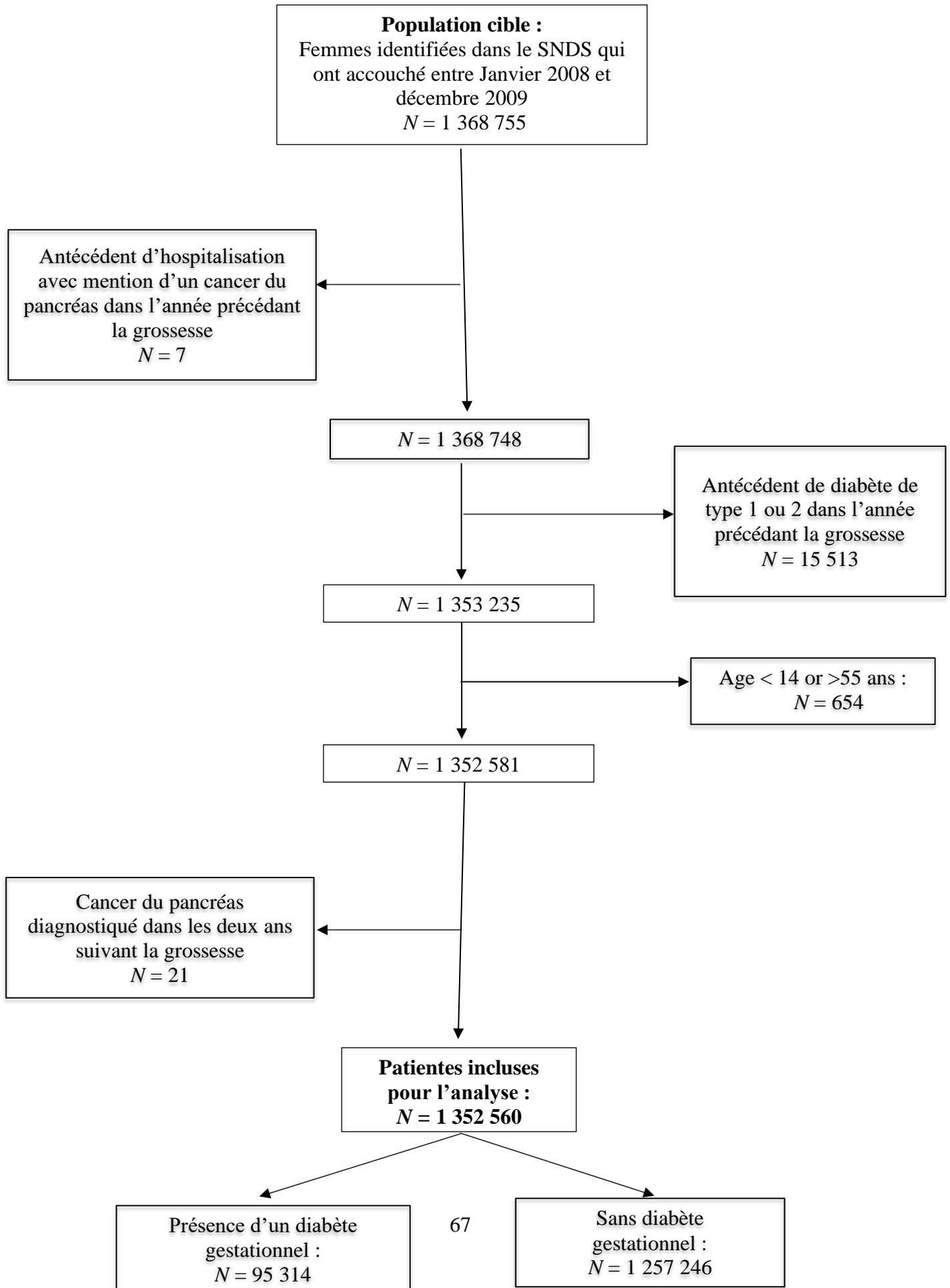
Nous allons ensuite rechercher dans la littérature les valeurs des rapports de risque concernant le DG et le CP entre les patientes obèses et non obèses. Si ceux-ci sont supérieurs à la e-value calculée alors, selon VanderWeele et al., l'obésité non mesurée pourrait expliquer à elle-seule la force d'association que nous aurions trouvée entre le DG et le CP. Au contraire, si les rapports de risque retrouvés dans la littérature sont inférieurs à la e-value calculée, selon VanderWeele et al., l'obésité ne pourrait pas expliquer à elle-seule la force d'association que nous aurions trouvée entre le DG et le CP, montrant ainsi la robustesse de notre résultat malgré le fait que nous n'avons pu inclure l'obésité dans notre modèle.

Le logiciel SAS 9.4 a été utilisé pour les analyses descriptives et multivariées.

Cette étude a été approuvée par le comité national français de protection des données (numéro d'enregistrement 12/11/2019-DR-2019-350). Le consentement écrit individuel n'était pas nécessaire pour cette étude. Les données de la base SNDS ont été fournies par la Caisse nationale d'assurance maladie française (21/02/2019 n° 226847).

#### 2.1.2.4 Résultats

De janvier 2008 à décembre 2009, 1 368 755 femmes hospitalisées pour un accouchement ont été identifiées dans la base de données du SNDS. Sept femmes qui ont été hospitalisées pour un CP avant la grossesse et 21 femmes qui ont été hospitalisées pour un CP dans les 2 ans suivant la grossesse ont été exclues (4 avec un DG et 17 sans) (Figure 7). De même, 15 531 femmes souffrant de diabète de type 2 et 654 femmes ne répondant pas aux critères d'âge ont également été exclues, ce qui a permis d'inclure 1 352 560 femmes dans la population analysée. La durée moyenne entre l'accouchement et l'apparition du CP sur l'ensemble des patientes incluses était de 1910 jours ( $\sigma = 604$ ). Afin de calculer cette durée moyenne à partir de l'exposition au DG, nous avons calculé la durée moyenne entre l'accouchement de la première grossesse avec DG et l'apparition du CP. Il était de 1616 jours ( $\sigma = 715$ ) dans le groupe avec DG et la durée moyenne entre l'accouchement de la première grossesse (grossesse d'inclusion dans l'étude) et l'apparition du CP était de 1923 jours ( $\sigma = 614$ ) dans le groupe sans DG.



### Figure 7. Diagramme de Flux de la population d'étude

Parmi les 1 352 560 femmes incluses dans notre étude, 95 314 ont présenté un DG (7,05%) au cours du suivi et 126 femmes ont été hospitalisées avec un CP (0,01%). Une comparaison des caractéristiques des femmes avec et sans DG est présentée dans le tableau 4.

**Tableau 4. Comparaison des caractéristiques des femmes avec des antécédents de Diabète Gestationnel (DG) ou sans antécédent de DG**

	<b>DG +</b>		<b>DG -</b>		<i>p</i>
	<i>N</i> = 95 314 (7,05%)		<i>N</i> = 1 257 246 (92,95%)		
<b>Age* (ans)</b>					<0,0001
< 26	18 095	19,0%	299 873	23,9%	
26-29	26 013	27,3%	352 910	28,1%	
30-33	24 279	25,5%	313 954	25,0%	
≥ 34	26 927	28,2%	290 509	23,1%	
<b>Hospitalisation avec un cancer du pancréas</b>	18	0,02%	108	0,01%	0,004
<b>Diabète de type 2 ultérieur</b>	8 184	8,6%	6 840	0,5%	<0,0001
<b>Consommation de tabac</b>	7 818	8,2%	86 403	6,9%	<0,0001

*n*, nombre;

\*Age entre 14 et 55 ans.

Les résultats du premier modèle de Cox, qui est ajusté pour l'âge et le diabète de type 2 ultérieur (en tant que variable dépendante du temps), ont montré que le DG (en tant que variable dépendante du temps) était significativement associé à un risque plus élevé d'hospitalisation mentionnant un CP (HR=1,81 ; IC 95% [1,06-3,10]). Les résultats du second modèle de Cox ajusté pour les mêmes variables auxquelles nous avons ajouté la consommation de tabac (en tant que variable dépendante du temps), ont montré que le DG était toujours significativement associé à un risque plus élevé d'hospitalisation mentionnant un CP avec presque le même HR (HR=1,77 IC 95% [1,03-3,03]) (Tableau 5). L'analyse de sensibilité avec la consommation de tabac considérée comme une variable fixe a un HR similaire au modèle 2 (HR=1,81 IC 95% [1,06-3,10]). Nous avons constaté que l'hospitalisation pour un diabète de type 2 ultérieur ou la consommation de tabac (considérée comme une variable dépendante du temps ou comme une variable fixe dans l'analyse de sensibilité) ne modifiait pas la relation entre le DG et l'hospitalisation mentionnant un CP. L'interaction entre la consommation de tabac et le DG était non significative (p=0,60), mais l'interaction entre l'âge et le DG était marginalement significative (p=0,05). Après avoir inclus cette interaction dans le modèle 1, l'effet du DG était encore plus important (HR=1,99 IC 95% [1,29-3,07]).

**Tableau 5. Modèles de Cox de la relation entre le diabète gestationnel (DG) et l'hospitalisation pour cancer du pancréas (n=126). Le DG, le diabète de type 2 ultérieur et le tabagisme sont des covariables dépendantes du temps.**

	Non-ajusté		Modèle 1*		Modèle 2**	
	HR	IC 95%	HR	IC 95%	HR	IC 95%
<b>DG</b>	<b>2,59</b>	<b>1,57-4,26</b>	<b>1,81</b>	<b>1,06-3,10</b>	<b>1,77</b>	<b>1,03-3,03</b>
<b>Age***</b>						
<26	0,60	0,28-1,27	0,61	0,28-1,30	0,56	0,26-1,21
26-29	1		1		1	
30-33	1,85	1,06-3,22	1,83	1,05-3,19	1,88	1,08-3,27
≥ 34	3,77	2,28-6,23	3,60	2,18-5,96	3,67	2,22-6,08
<b>Diabète de type 2</b>	8,06	3,76-17,32	4,85	2,13-11,06	4,74	2,08-10,79
<b>Tabagisme</b>	2,80	1,72-4,56	-	-	3,22	1,97-5,26

*DG : Diabète gestationnel ; HR : Hazard Ratio ; IC : intervalle de confiance.*

*\*Modèle 1 inclut DG, âge and diabète de type 2*

*\*\*Modèle 2 inclut DG, âge, diabète de type 2 and tabagisme*

*\*\*\*Age entre 14 et 55 ans*

Afin de limiter le biais de confusion pouvant être induit par la variable non mesurée « obésité », nous avons effectué une analyse évaluant la e-value. Dans notre étude, notre HR observé est de 1,77, nous calculons la e-value selon la formule :

$$E\text{-value} = HR + \sqrt{(HR \times (HR-1))} = 1,77 + \sqrt{(1,77 \times (1,77-1))} = 2,93$$

Le « HR » est notre hazard ratio observé de 1,77 ; ce qui nous donne une E-value égale à 2,93. Cette e-value représente la force d'association minimale que l'obésité devrait avoir avec le DG et le CP pour expliquer à elle-seule la force d'association que nous avons trouvée entre le DG et le CP. Dans la littérature, nous avons constaté que, par rapport aux femmes non obèses, les femmes obèses ont 3,36 fois plus de risque de souffrir de DG (78) ( ce chiffre est supérieur à 2,93) ; et 1,54 fois plus de risques de souffrir de CP (79), ( ce chiffre est inférieur à 2,93). On peut donc constater que le HR (pour la relation obésité-DG) est supérieur à l'e-value alors que le HR (pour la relation obésité-CP) est inférieur à l'e-value, on ne pourra donc conclure selon l'article princeps de VanderWeele et al., qu'en calculant la valeur B, ici présentée :

$$B = HR_{(\text{obésité-DG})} * HR_{(\text{obésité-CP})} / (HR_{(\text{obésité-DG})} + HR_{(\text{obésité-CP})} - 1) = 3,36 * 1,54 / (3,36 + 1,54 - 1)$$

$$B = 1,33$$

Si cette valeur B était supérieure au HR retrouvé dans notre étude de 1,77, alors la variable « obésité » non mesurée pourrait expliquer à elle-seule la force d'association que nous avons trouvée entre le DG et le CP. Cependant, B est égale à 1,33 et est donc inférieure à 1,77. Selon la démarche proposée par VanderWeele et al., on pourrait donc considérer notre résultat principal comme robuste après analyse de la e-value malgré le fait que la variable obésité n'a pu être mesurée et donc incluse dans notre modèle.

### 2.1.2.5 Discussion

À notre connaissance, il s'agit de la première étude publiée dans une revue internationale qui explore spécifiquement la relation entre le DG et le CP, après ajustement sur le diabète ultérieur, qui est le principal facteur de confusion potentiel. Notre étude en population utilisait une cohorte rétrospective couvrant l'ensemble de la population nationale des femmes qui ont accouché à l'hôpital, environ 99,6% des accouchements avaient lieu à l'hôpital en France durant la période de l'étude (80,81). Grâce à ces données quasi-exhaustives concernant les accouchements en France, nous avons montré que chez les femmes incluses dans notre étude, le DG était significativement associé à un risque plus élevé d'être hospitalisée avec mention d'un CP dans les 8 premières années, indépendamment de l'apparition ultérieure d'un diabète de type 2.

Dans la littérature, les trois études sur ce sujet font état de méthodologies variables et de résultats divergents. Les deux études israéliennes montrent des HR significatifs similaires entre elles, avec un HR égal à 7,10 (IC 95% [2,80–18,00]) dans l'étude de Perrin et al. et un HR égal à 7,06 (IC 95% [1,69-29,45]) pour Sella et al., soit beaucoup plus élevés que celui que nous avons retrouvé, mais avec des intervalles de confiance plus larges que dans notre étude (71,72). Les études israéliennes étaient hospitalières, incluant des patientes de trois hôpitaux pour Perrin et al. et d'un seul pour Sella et al. Elles ont également étudié un nombre limité de cas : cinquante-quatre pour Perrin et al. et neuf pour Sella et al., par rapport à notre étude qui a identifié 126 cas parmi les 1 352 560 femmes incluses. La taille de la population plus importante dans notre étude pourrait bien sûr expliquer l'IC plus étroit. En outre, notre HR plus faible pourrait s'expliquer par un ajustement pour le diabète ultérieur. De plus, notre période d'inclusion (2008-2009) est beaucoup plus récente que celle de Perrin et al. (1964-1975), qui s'appuyait sur des critères de diagnostic du DG se basant sur la glycosurie. Or, ce n'est qu'en octobre 1984, lors de la deuxième « International Workshop on GDM », qu'il a été déterminé que « toutes les femmes enceintes devraient faire l'objet d'un dépistage de l'intolérance au glucose, car le dépistage sélectif basé sur les attributs cliniques ou les antécédents obstétricaux est inadéquat » (82). De plus, plusieurs études ont montré la faible sensibilité de la glycosurie pour dépister le DG (83–85). Cela a pu avoir un impact sur leurs résultats en créant un biais de classification, par sélection des femmes ayant un diabète gestationnel potentiellement moins équilibré. En revanche, l'étude taiwanaise de Peng et al. est davantage comparable à notre étude, puisqu'il s'agit également d'une étude en population utilisant des données médico-

administratives nationales (73). Ils ont inclus presque le même nombre de femmes (1 466 596 femmes) mais ont observé un plus grand nombre de cas de cancer du pancréas (331 cas). Bien qu'ils aient trouvé deux fois plus de cas que dans notre étude, ils n'ont pas pu montrer une augmentation significative du risque de CP chez les femmes ayant des antécédents de DG (HR=1,07 IC 95% [0,65-1,75]). Cependant, ils ont pris en compte tous les CP survenus après la grossesse alors que nous avons exclu les CP survenus moins de deux ans après le DG. De plus, il a été démontré que les populations asiatiques développent le diabète à un plus jeune âge, nécessitent plus souvent un traitement par insuline et ont un risque de cancer augmenté par rapport aux populations occidentales (86,87), ce qui peut expliquer le nombre plus élevé de cas de CP. En outre, contrairement à notre étude, le suivi dans l'étude taïwanaise était de durée variable selon les femmes, certaines ayant été suivies pendant plus de douze ans alors que d'autres n'ont été suivies que pendant un an. Enfin, nous avons inclus le DG comme variable dépendante du temps, ce qui pourrait expliquer pourquoi nous avons trouvé un résultat statistiquement significatif malgré un nombre de cas moins élevé.

Le premier point fort de notre étude est le fait d'avoir pu réaliser un ajustement sur le diabète de type 2. A notre connaissance, notre étude est la seule publiée au niveau international ayant pu procéder à cet ajustement sur ce facteur de confusion potentiel majeur. En effet, plusieurs études ont montré qu'il existe une relation entre le DG et le diabète de type 2, et il a été suggéré que le diabète de type 2 augmente directement, ou indirectement le risque de CP par la dégénérescence des néoplasmes mucineux papillaires intraductaux (88). Par conséquent, il nous a semblé important de prendre en compte ce facteur de confusion dans la méthodologie de notre étude. Tout d'abord, concernant notre base de données, une étude de validation dans une cohorte française par Fuentes et al. a montré que le diabète de type 2 était identifié dans la base de données nationale (SNDS) avec une valeur prédictive positive de 97,2%, une sensibilité de 93,8% et une spécificité de 99,9% (89). Ensuite, nous avons exclu les femmes ayant des antécédents de diabète de type 2 afin d'affiner la définition du DG et de supprimer l'impact des antécédents de diabète de type 2 sur l'analyse. Enfin, le diabète de type 2 survenant au cours du suivi a été inclus dans nos modèles en tant que variable dépendante du temps. En introduisant des variables dépendant du temps dans les modèles de Cox, il a été possible d'introduire des informations recueillies au cours du suivi. Cela nous a permis de prendre en compte un changement de statut, par exemple de non-diabétique à diabétique, mais aussi le temps d'exposition au diabète avant l'hospitalisation mentionnant un CP.

Le temps est également une considération importante pour l'étude du CP. Tout d'abord, concernant la chronologie de la maladie elle-même, certaines études montrent que l'hospitalisation est un bon marqueur de nouveaux cas pour certains cancers agressifs et/ou nécessitant une hospitalisation (90,91). Sachant que nous avons pris en compte tous les types d'hospitalisations (y compris celles concernant le diagnostic, la radiothérapie, la chimiothérapie, la chirurgie et le suivi), nous avons pu exclure les sujets ayant des antécédents d'hospitalisation pour CP un an avant, pendant et deux ans après la grossesse (soit près de quatre ans). Il a été montré dans la littérature que le CP a un taux de survie net à 5 ans de 6-9% (41-43). En excluant les femmes présentant une hospitalisation mentionnant un CP au cours des quatre premières années, nous nous sommes assurés dans cette étude qu'il y avait une très forte probabilité d'inclure uniquement des cas incidents de CP. Deuxièmement, à notre connaissance, notre étude est la seule qui ait établi un délai de deux ans entre la grossesse (avec ou sans DG) et le diagnostic de CP. Il a été suggéré que la sélection des cas de CP ne devrait pas inclure le diabète récemment diagnostiqué, car il est connu qu'un diagnostic de diabète peut être un symptôme de CP. En effet, le diabète d'apparition récente causé par le CP peut être associé à des altérations pro-inflammatoires, à une résistance à l'insuline et à des perturbations des fonctions des cellules  $\beta$  qui entraînent une perte de l'homéostasie du glucose. Les adipocytes et les macrophages libèrent de nombreux facteurs pro-inflammatoires délétères, et l'augmentation de la lipolyse pourrait accroître la production hépatique d'acétyl-coenzyme A, provoquant une insulino-résistance hépatique et une libération excessive de glucose par le foie. (76)

Notre étude présente d'autres points forts. Tout d'abord, notre étude en population comprend 10 ans de données, ce qui nous permet d'avoir 8 ans de suivi pour toutes les femmes incluses. Elle s'est appuyée sur une base de données nationale quasi-exhaustive concernant les accouchements (99,6%), car la quasi-totalité des accouchements en France ont lieu à l'hôpital et tous les séjours hospitaliers sont enregistrés dans cette base. De plus, dans une étude de validation nationale (92), notre équipe a montré que le DG était identifié dans cette base de données avec une valeur prédictive positive de 88,9%.

Nous reconnaissons également que notre étude présente certaines limites. Tout d'abord, notre récente étude de validation a montré que la sensibilité pour le DG n'était que de 79%, ce qui signifie que 21% des femmes atteintes de DG n'ont pas été identifiées comme telles dans notre base de données (93). Par conséquent, ce biais de classification peut avoir conduit à une sous-estimation de l'association. En outre, en France, les directives nationales pour le diagnostic

du DG ont changé pendant la période d'étude. Avant 2010, le dépistage du DG était basé sur une procédure en deux étapes pour toutes les femmes : le premier test était réalisé sur la glycémie veineuse 1 heure après l'ingestion de 50g de glucose et en cas de résultat positif, le second test était réalisé sur la glycémie veineuse 3 heures après l'ingestion de 100g de glucose. Depuis 2010, de nouvelles directives nationales ont été introduites pour les femmes présentant certains facteurs de risque (c'est-à-dire âge maternel  $\geq 35$  ans, IMC  $\geq 25$ , antécédents de diabète au 1er degré, antécédents personnels de DG ou de macrosomie). Le dépistage recommandé pour ces femmes est la glycémie à jeun lors de la première consultation prénatale. Si elle est anormale, prise en charge immédiate du diabète et si elle est normale, un test en une seule étape d'hyperglycémie provoquée orale avec 75g de glucose est prescrit au cours du deuxième trimestre. Les trois valeurs de glycémie (à jeun, à une heure et à deux heures après l'ingestion des 75g de glucose) doivent être considérées comme normales sinon un DG est diagnostiqué. Compte tenu du fait que notre étude concerne les années 2007 à 2017, cette différence dans la manière de diagnostiquer le DG peut modifier l'estimation de la relation. En effet, une revue systématique de Pillay et al. en 2021, a montré que le dépistage en une étape identifiait significativement davantage de DG (11,5%) par rapport au dépistage en deux étapes qui dépistait un DG chez 4,9% des participantes (94).

Deuxièmement, bien que nous disposions d'une durée de suivi plus longue et plus homogène par rapport aux trois études précédentes, 8 ans restent relativement courts pour l'étude de l'apparition des CP, dont la plupart sont diagnostiqués après 65 ans. De plus, pour plusieurs patientes, le DG n'était pas présent lors de la grossesse à l'inclusion mais lors d'une grossesse ultérieure durant le suivi. Il n'y avait donc pas de présence d'un antécédent de DG durant 8 années pour l'ensemble des patientes présentant un DG durant au moins une de ces grossesses. Nous pourrions donc avoir sous-estimé l'impact complet du DG.

Troisièmement, bien que nous ayons inclus la "consommation de tabac" comme variable d'ajustement, cette information est incomplète dans notre base de données administratives entraînant un biais de classification entre les femmes fumeuses et non fumeuses. De plus, la consommation de tabac est un facteur de risque reconnu de CP. On peut donc supposer que, si la consommation de tabac est mal renseignée dans le SNDS, une part de la force d'association entre le DG et le CP soit dû à la consommation de tabac. Afin de limiter ce biais, nous avons réalisé trois modèles multivariés, l'un avec cette variable en tant que variable dépendante du temps, un autre en tant que variable fixe et un autre sans cette variable ; les trois ont donné des

résultats similaires.

Quatrièmement, certains autres facteurs de risque connus n'ont pas pu être inclus dans notre étude. Bien que nous voulions ajouter la "consommation d'alcool" à nos variables d'ajustement, puisqu'il s'agit d'un autre facteur de risque bien connu du CP, il n'a pas été possible de le faire car cette variable est trop peu enregistrée dans notre base de données (95). De plus, la variable " obésité ", dont on sait qu'elle est associée au DG et au CP, n'a pas pu être incluse dans notre modèle. Malheureusement, l'IMC avant la grossesse n'est que peu documenté car cette donnée n'est disponible que s'il existe une hospitalisation avant la grossesse, et nous craignons que l'IMC pendant la grossesse ne puisse pas être considéré comme aussi fiable. C'est pourquoi nous avons effectué une analyse de la e-value, qui a considéré, selon la démarche proposée par VanderWeele et al., que notre résultat principal est robuste malgré la variable obésité non mesurée.

### 2.1.2.6 Conclusion

En conclusion, notre étude montre que le DG est un facteur associé au développement d'un CP dans les 8 ans suivant l'accouchement. Cependant, le DG est encore considéré par les femmes et certains professionnels comme une maladie ayant peu d'impact une fois la grossesse terminée, d'où un suivi post-partum à court et long terme de mauvaise qualité (96–99). Pourtant, la grossesse est une sorte de " stress-test " pour le corps des femmes, qui peut révéler certaines vulnérabilités. En effet, nous avons constaté dans de précédentes études que le DG expose les femmes à un risque accru de diabète de type 2, de syndrome métabolique et de troubles cardiovasculaires (54–63). Les résultats de notre étude suggèrent qu'un risque accru de CP devrait être ajouté à cette liste. Tous ces résultats sont en faveur d'une meilleure adhésion aux recommandations (100,101) pour le suivi des femmes souffrant de DG dans la période post-partum et au-delà. Aux Etats-Unis, comme en France, le dépistage du diabète de type 2 est recommandé pendant la période postnatale, puis tous les un à trois ans, en fonction des facteurs de risque, pendant au moins 25 ans. Après un DG, les professionnels de santé doivent encourager les changements de mode de vie (30 à 60 minutes d'activité physique quotidienne au moins cinq jours par semaine, alimentation équilibrée, arrêt du tabac...) qui sont également utiles pour la prévention du CP (102,103). Pour être efficace, ce suivi nécessiterait la participation de tous les professionnels de santé, notamment les médecins généralistes, les gynécologues, les obstétriciens, les sage-femmes et les endocrinologues.

## **2.2. Avantages et inconvénients des données médico-administratives**

### **2.2.1 Les avantages et inconvénients du SNDS en France**

#### **2.2.1.1 Les avantages**

Les bases médico-administratives possèdent plusieurs avantages. En premier lieu, elles contiennent des données pour de nombreuses variables à l'échelle de larges populations. En effet, le SNDS est inter-régime et regroupe plusieurs bases. Cela permet de suivre la quasi-totalité des français dans leur parcours de soins tout au long de leur vie. Ce parcours de soins comprend, entre autres, leurs soins hospitaliers, leurs soins de ville, leurs traitements prescrits, ou encore leurs handicaps, leurs arrêts de travail pour les salariés ainsi que leurs décès et leurs causes.

Deuxièmement, ce sont des bases quasi-exhaustives pour des maladies, comme le cancer, qui nécessitent une hospitalisation ou une ALD.

Enfin, il y a une quasi-absence de biais d'attrition. La problématique des perdus de vue est cruciale dans les études observationnelles, en particulier lorsque la durée de suivi est longue. Avec peu de données manquantes et une collecte du statut vital des individus auprès de l'état civil de la commune de naissance du patient en France ou par recoupement avec le registre CépiDc, le SNDS est donc particulièrement adapté aux études nécessitant un long suivi.

#### **2.2.1.2 Les inconvénients**

Si le SNDS constitue un intérêt certain en épidémiologie, nous pouvons également citer plusieurs inconvénients.

Tout d'abord, les épidémiologistes se heurtent à quelques contraintes afin de pouvoir disposer des informations contenues dans ces bases. Toute personne ou structure, publique ou privée, à but lucratif ou non lucratif, peut depuis avril 2017 accéder aux données du SNDS sur autorisation de la CNIL, en vue de réaliser une étude, une recherche ou une évaluation présentant un intérêt public. Cependant, conformément à la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé, les accès aux données du SNDS doivent s'effectuer dans « des conditions assurant la confidentialité et l'intégrité des données et la traçabilité des accès et des autres traitements ». En effet, le SNDS assure la mise à disposition des données selon la nature des données et le risque de réidentification des patients, en deux modalités distinctes. Les données pour lesquelles aucune réidentification n'est possible sont en open data. Alors que les données potentiellement ré-identifiantes sont accessibles avec des accès aux données SNDS régulés : soit par projet, soit sur autorisation pérenne. Ces autorisations pérennes concernent l'Etat et ses agences, des établissements publics et des organismes chargés d'une mission de service public (INSERM, INCA, HAS,...) listés dans le décret n°2016-1871 du 26 décembre 2016 et sont restreints selon des niveaux d'agrégation et des historiques (104,105).

Ces bases médico-administratives ont été constituées principalement à deux fins : pour le financement des établissements de santé et pour l'organisation de l'offre de soins. Leur utilisation à des fins de recherche n'a été envisagée que dans un second temps et elles ne sont donc pas spécifiquement adaptées à l'épidémiologie. Par exemple, en l'absence de code spécifique pour certaines maladies ou si l'utilisation seule des codes spécifiques n'est assez sensible et/ou spécifique pour repérer les cas de pathologies, il est alors nécessaire d'utiliser des algorithmes. Ces algorithmes sont conçus à partir des codes-diagnostic et thérapeutiques de la Classification internationale des maladies et CCAM, et/ou de remboursement de médicaments liés à la maladie d'intérêt, et/ou d'ALD si la maladie d'intérêt le permet. Des études de validation vont alors évaluer la capacité des algorithmes à distinguer les malades et les non-malades en rapportant leur sensibilité, c'est-à-dire la capacité d'identifier correctement les personnes ayant la maladie, et leur spécificité, soit la capacité d'identifier correctement chez les personnes indemnes de la maladie.

Si les épidémiologistes ne retrouvent pas d'algorithmes validés dans la littérature, ils peuvent adapter la méthodologie de leur étude afin de limiter les biais, comme nous l'avons fait pour certains facteurs de risque comme par exemple la consommation de tabac. En effet, nous

avons constaté que de définir la consommation de tabac à partir de ces bases de données demandait des compromis, puisque toute l'information concernant le début ou l'arrêt de la consommation n'y est pas captée. Nous avons alors émis certaines hypothèses afin de s'approcher de la réalité du mieux possible en utilisant différentes données disponibles dans le SNDS, comme la prise d'un médicament remboursé pour arrêter de fumer ou la mention de tabagisme lors des hospitalisations. Cela nous a permis de détecter dans notre population des personnes ayant fumer à un moment de leur vie mais nous n'avons pas les données suffisantes pour déterminer la fréquence, la quantité et la durée de cette consommation, ce qui engendre des incertitudes sur les résultats de l'étude.

Le fait que les bases de données médico-administratives ne soient pas au départ conçues pour la recherche engendre un manque de données concernant plusieurs variables pouvant intéresser l'épidémiologie. Par exemple, concernant le statut socio-professionnel, la seule donnée individuelle connue est le fait de bénéficier ou non de la complémentaire santé solidaire, c'est-à-dire les patients ayant de faibles ressources. Les informations sur la commune de résidence des patients sont trop vastes pour permettre une analyse fine de l'environnement et des conditions de vie des individus. On peut citer par exemple la disponibilité dans le SNDS d'un indice de désavantage social, mais c'est une variable écologique, caractérisée au niveau territorial à partir de la commune de résidence. Les résultats des examens cliniques ou paracliniques sont manquants. Les données concernant les prestations non remboursées ne sont pas prises en charge et donc non présentes dans le SNIIRAM engendrant également une absence de données sur l'automédication ainsi que sur les médicaments non remboursés.

### 2.2.1.3 Utilisation du SNDS dans l'épidémiologie des cancers en France

Quelques exemples d'études concernant l'épidémiologie des cancers permettent d'illustrer les différents points que nous avons cités dans les chapitres précédents concernant les avantages et les inconvénients de cette base de données.

Nous avons retrouvé une majorité de travaux qui étudient les traitements du cancer (106–110). Le SNDS est un excellent outil de pharmaco-épidémiologie pour des médicaments remboursés en ville ou des médicaments facturés en sus des tarifs des séjours hospitaliers comme le sont certains traitements innovants et onéreux contre le cancer. C'est d'ailleurs une des forces que relève l'ensemble de ces études.

D'autres aspects de l'épidémiologie des cancers ont été explorés comme l'incidence, la prévalence ou l'étude des caractéristiques des patients atteints de certains cancers (110–112). Pour les auteurs, une des forces majeures de leurs études est le grand échantillon leur permettant d'obtenir une puissance statistique importante. Les auteurs de ces travaux ajoutent également dans la liste des forces : la quasi-exhaustivité de leur échantillon, le SNDS couvrant 99% de la population française.

Concernant les inconvénients, plusieurs de ces études indiquent comme limite une absence ou une information limitée, dans le SNDS, de certaines variables pouvant jouer un rôle sur la relation étudiée.

D'une part, certaines comorbidités ou certains facteurs de risques de cancer sont parmi les variables absentes. Par exemple, Goulabchand et al. (112) étudient la relation entre cancer et le syndrome de Gougerot-Sjögren, qui est une maladie auto-immune caractérisée par une infiltration lymphoïde des glandes salivaires et lacrymales responsable d'une sécheresse buccale et oculaire, et par la production de différents auto-anticorps. Ils ont notifié comme limites de leur étude la difficulté à prendre en compte la consommation d'alcool ou de tabac dans le SNDS. De même, Assié et al. (108) qui étudient la survie des patients atteints d'un cancer du poumon non à petites cellules et traités par Nivolumab indiquent qu'il aurait été intéressant de prendre en compte l'expression du gène PD-L1 ou « certaines comorbidités » qui sont absentes dans le SNDS.

D'autre part, certaines études évoquent également le manque d'information concernant le suivi médical du cancer. Touzeau et al. (106) décrivent un manque de données concernant les lignes de traitement ou la participation à des études cliniques. Champeaux-Depond et al. (107) citent l'absence d'information concernant la qualité de la résection lors du traitement chirurgical de la tumeur. Assié et al. (108) mentionnent également certains

traitements concomitants pouvant influencer la survie globale. Or, les traitements prescrits en hospitalier, non en sus, ne changeant pas le forfait hospitalier, ne sont pas répertoriés dans le SNDS. Thurin et al. (111), comme Scailteux et al. (109), et Tuppin et al. (110) quant à eux, décrivent un manque d'information concernant l'imagerie, les résultats de laboratoire, ou la classification TNM.

## **2.2.2 Exemples d'études à partir de bases médico-administratives dans le monde**

Afin d'élargir notre point de vue concernant les bases médico-administratives, nous allons présenter dans ce chapitre quelques exemples de bases médico-administratives à travers plusieurs études effectuées à l'étranger. Nous avons recherché dans la base Pubmed des études relativement récentes effectuées uniquement à partir de bases médico-administratives. Nous avons fait le choix de présenter des bases médico-administratives sur différents continents dans des pays possédant des systèmes de soins et d'économie de la santé variés. Cela permettra d'identifier certaines similarités et différences en termes de forces et de limites de ces bases de données concernant l'épidémiologie des cancers en comparaison au SNDS en France. A noter que nous avons fait le choix de ne pas inclure d'exemples Nord-Américains (Etats-Unis et Canada) dans ce chapitre car le chaînage entre bases médico-administratives et registres de cancer est courant dans ces pays et nous les développerons donc dans le chapitre 3 traitant de la complémentarité entre ces deux types de bases de données.

### **2.2.2.1 La Catalogne**

En Catalogne, il existe une base de données médico-administratives sur les sorties d'hôpitaux pour maladies aiguës (hospitalisation standard, chirurgie ambulatoire, hôpital de jour et hospitalisation à domicile), qui comprend tous les épisodes de soins de santé dans les centres de santé catalans publics et privés. En 2014, Escriba et al. (113) ont étudié les tendances dans les procédures chirurgicales des femmes atteintes d'un cancer du sein incident en Catalogne sur une période de 7 ans (2005-2011) à partir de cette base de données. Les cas de cancer du sein ont été identifiés sur la base de la 9<sup>ème</sup> édition de la classification internationale des maladies. Les résumés administratifs de sortie d'hôpital de ces cas de cancer du sein ont été utilisés pour

identifier et structurer de manière appropriée les épisodes hospitaliers.

Comme forces de leur étude, ils relèvent notamment l'utilisation d'une base de données médico-administratives qui permet d'avoir un grand échantillon de cas de cancer et qui est proche de l'exhaustivité car le cancer est une maladie qui nécessite très souvent une hospitalisation. Ils soulignent malgré tout plusieurs limites comme une hétérogénéité dans les pratiques de codage entre les professionnels et les hôpitaux (notamment entre le public et le privé). Ils évoquent également que certains algorithmes utilisés, notamment concernant l'identification du stade de la tumeur, ne sont pas validés.

### 2.2.2.2 L'Allemagne

En Allemagne, en 2020, une étude de Bögemann et al. (114) a évalué les schémas de traitement et l'utilisation des ressources de santé chez les patients atteints d'un carcinome rénal métastatique. Elle a utilisé la base de données de la "Bundesverband der Betriebskrankenkassen" comprenant les dossiers d'environ cinq millions d'assurés qui étaient couverts par l'assurance maladie. Les cas de cancer du rein ont été identifiés sur la base de la 10<sup>ème</sup> édition de la classification internationale des maladies. Les données de facturation de routine ont été utilisées pour décrire l'utilisation des ressources de santé des patients. Les traitements prescrits, les traitements ambulatoires, les traitements hospitaliers, les hospitalisations, les arrêts de travail pour cause de maladie ont été évalués chaque trimestre. Les données cliniques comprenaient les durées de traitement, la survie sans progression et la survie globale. Le décès était présumé lorsque l'inscription à l'assurance maladie obligatoire s'arrêtait.

Les auteurs estiment que cette base de données médico-administratives est précieuse pour caractériser les pratiques de soins courants et leurs implications cliniques et financières. Néanmoins, comme ces données sont principalement générées pour les opérations de remboursement et non à des fins de recherche clinique, elles ne contiennent donc qu'un nombre limité de variables cliniques et démographiques et ne sont pas suffisamment détaillées pour interpréter les résultats dans le contexte des profils de risque des patients (propagation des tumeurs, ou par rapport à d'autres considérations cliniques pertinentes, telles que l'occurrence et l'impact des effets secondaires, les modifications de dose et l'adhésion au traitement).

### 2.2.2.3 Corée du Sud

En 2017, Heo et al. (115) a étudié la prévalence de la grippe chez les patientes atteintes d'un cancer du sein en Corée du Sud en utilisant la base de données « Health Insurance Review and Assessment Service ». Cette base de données comprend des données démographiques des patients, les caractéristiques de l'institution, les antécédents médicaux, le diagnostic principal et supplémentaires lors de l'hospitalisation, ainsi que les médicaments prescrits en dehors de l'hôpital. Les cas de cancer du sein ont été identifiés sur la base de la sixième édition de la classification coréenne des maladies, une version de la dixième édition de la Classification internationale des maladies, modifiée pour le système de soins de santé coréens.

Cette base médico-administrative nationale leur a permis d'avoir une grande quantité de cas. Néanmoins, ils évoquent plusieurs limites à leur étude. Tout d'abord, les données cliniques relatives à la gravité de la grippe, les symptômes, les résultats des tests de laboratoire dans les établissements médicaux n'étaient pas disponibles dans la base de données. De plus, les facteurs associés à la gravité du cancer du sein, tels que le stade et le rapport anatomopathologique, n'étaient également pas disponibles. En outre, ils n'ont pas pu confirmer si les patientes avaient été vaccinées contre la grippe, parce que la vaccination contre la grippe n'est pas une donnée de remboursement notifiée par l'assurance maladie en Corée du sud.

### 2.2.2.4 À Taïwan

En 2013, Chen et al. (116) ont évalué les risques de cancer du sein et de l'endomètre chez les femmes diabétiques à Taïwan en utilisant la base médico-administrative du « National Health Insurance Research Database » ou « NHIRD » qui est constituée de toutes les demandes de remboursement de soins hospitaliers et ambulatoires pour environ 99% de la population taïwanaise. Des examens d'experts sur un échantillon aléatoire de 50 à 100 demandes de remboursement de soins ambulatoires et hospitaliers sont effectués chaque trimestre. Ils ont utilisé plusieurs ensembles de données du NHIRD dans cette étude, y compris les demandes de remboursement des visites de soins ambulatoires, les demandes de remboursement des patients hospitalisés, et le registre des bénéficiaires. Les cas de cancer du sein ont été identifiés sur la base de la 9<sup>ème</sup> édition de la classification internationale des maladies. Les informations sur les

antécédents des patientes ont été récupérées à partir des données d'hospitalisation ou en ambulatoire. Ils comprenaient l'hypertension, l'endométriase et les antécédents d'avortement. Ils ont également calculé l'indice de comorbidités de Charlson pour indiquer le niveau de comorbidité des individus et ont ajusté sur la fréquence des consultations externes pour chaque sujet de l'étude afin d'éviter un biais de surveillance des maladies.

Les auteurs ont notifié plusieurs forces parmi lesquelles, l'utilisation d'une base de données nationales, en population, avec très peu de biais de sélection et d'attrition et la puissance statistique de leur étude qui leur a permis d'effectuer des analyses détaillées en sous-groupes d'âge. Ils évoquent également plusieurs limites : la précision d'un diagnostic de diabète était de 74,6 % dans leur base de données selon une étude de validation, cependant il n'y a pas possibilité de différencier le diabète de type 1 du diabète de type 2. De plus, ils n'avaient pas de données concernant l'IMC, la durée et les traitements du diabète, ni d'autres caractéristiques socio-économiques. Certains facteurs de risque de cancer du sein, dont les antécédents familiaux de cancer, le nombre d'accouchements et l'allaitement, ne sont pas non plus disponibles dans la base de données.

### 2.2.2.5 Synthèse

Au total, même si l'on retrouve de nombreuses similarités entre les différentes bases de données médico-administratives présentées ci-dessus, nous pouvons également noter plusieurs différences importantes.

Concernant les similitudes, on peut repérer plusieurs points : ce sont toutes des bases de données ayant été créées dans un but de remboursement des dépenses de soins par les assurances maladies. Cela génère certaines forces et certaines limites similaires pour chaque étude. En effet, chacune évoque des échantillons de grande taille avec peu de biais d'attrition mais des variables manquantes parmi celles qui pourraient jouer un rôle dans la relation étudiée (données cliniques, paracliniques et/ou socio-économiques notamment). En outre, toutes les bases médico-administratives ont la même problématique de validité des données avec la nécessité d'effectuer des études de validation au préalable.

Concernant les différences, on peut également en citer plusieurs en comparant nos exemples. En premier lieu, la représentativité de la population au niveau national (Allemagne

Corée du Sud, France, Taïwan) ou régional (Catalogne) qui peut être proche de l'exhaustivité (99%) comme en France ou à Taïwan ou représenter qu'une partie de la population (5 millions d'habitants couverts par l'assurance maladie dans une population allemande de 83 millions d'habitants). Le fait de ne pas être exhaustif n'est pas forcément un problème si l'échantillon est représentatif de la population de laquelle il est issu. Or, dans la majorité des cas la couverture par l'assurance maladie est dépendante par exemple du revenu ou du statut socio-économique (salariée et/ou retraitée). Par conséquent, la population couverte par l'assurance maladie du pays n'est alors pas représentative de l'ensemble des catégories socio-économiques de ce pays.

Deuxièmement, les bases de données peuvent inclure des données uniquement sur les hospitalisations comme en Catalogne, d'autres vont y ajouter des données de ville comme les consultations chez le médecin généraliste (Taïwan) ou les prescriptions de médicaments (Corée du Sud, Allemagne). En France, l'ensemble de ces éléments est récolté grâce à la fusion des bases du PMSI, du SNIIRAM auxquelles s'ajoutent données concernant les décès et leur cause (CepiDC) entre autres.

Troisièmement, la procédure de validation et de vérification des données n'a pas la même qualité selon les bases de données. En France des équipes d'experts au niveau des départements d'information médicale de chaque établissement hospitalier vérifie l'exactitude des données. Cette vérification sera complétée par les directions régionales et les agences régionales de santé qui sont missionnées pour le contrôle du codage de la tarification à l'activité. En comparaison, à Taïwan, des experts vérifient une faible proportion de dossiers de manière aléatoire.

Quatrièmement, la proportion d'algorithmes et de codages validés par des études de validation diffère également selon les bases de données et peut entraîner des biais plus ou moins importants comme le souligne Escriba et al. (113), qui décrivent l'absence de codes spécifiques pour les techniques chirurgicales les plus récentes entraînant une impossibilité de les discriminer comparé à d'autres pratiques. Également, Chen et al. (116) qui, en l'absence d'algorithmes validés, ont considéré qu'un patient était diabétique s'il avait été codé dans la base comme tel en 2000, puis une seconde fois au cours des 12 mois suivants, pouvant entraîner par exemple un biais de classification pour certains patients n'ayant été codé dans la base qu'une seule fois.

Cinquièmement, on peut également observer que les bases de données utilisent des classifications différentes pour définir les cancers. Par exemple, 9<sup>ème</sup> ou 10<sup>ème</sup> édition de la classification internationale des maladies respectivement en Catalogne et en Allemagne, alors qu'en Corée du Sud, ils ont leur propre classification coréenne des maladies. Cela entraîne des définitions différentes de la maladie, il est alors important de le prendre en considération lorsqu'on veut comparer des résultats d'études faites sur des bases de données différentes.

Enfin, le codage des données dans les bases médico-administratives est dépendant de la couverture de l'assurance maladie. Plus l'assurance maladie couvre d'actes, rembourse de médicaments et discrimine de maladies, plus le codage sera étendu et détaillé. Par exemple, Heo et al. (115) n'ont pas pu confirmer si les patientes avaient été vaccinées contre la grippe, parce que la vaccination contre la grippe n'est pas une donnée de remboursement notifiée par l'assurance maladie en Corée du sud. Enfin, Chen et al. (116) n'ont pas pu différencier le diabète de type 1 du diabète de type 2 dans leur base de données.

## **3. Complémentarité des deux approches**

Maintenant que nous avons développé les avantages et les inconvénients des registres de cancer d'une part et des bases de données médico-administratives d'autre part, nous allons explorer la complémentarité des deux approches. Dans un premier temps nous avons effectué une revue de la littérature concernant la relation entre le diabète gestationnel et le risque de cancer du sein. Celle-ci était composée de plusieurs études utilisant des données issues de registres de cancer ou de bases de données médico-administratives. Cela nous a permis d'analyser si les informations que l'un et l'autre nous apportent sont différentes et/ou complémentaires afin de répondre à une question concernant l'épidémiologie des cancers. Dans un second temps, nous analyserons des exemples d'études dans la littérature utilisant le chaînage entre les bases de données médico-administratives et les registres de cancer afin de savoir si les avantages de l'un peuvent compenser certaines limites de l'autre.

### **3.1. Revue de la littérature concernant la relation entre le diabète gestationnel et le cancer du sein**

#### **3.1.1 Contexte et objectif : étude du lien entre diabète gestationnel et cancer du sein**

Après avoir effectué notre étude concernant la relation entre le DG et le risque de CP, nous souhaitons nous intéresser à la relation entre le DG et le risque de cancer du sein (CS) afin d'effectuer une nouvelle étude à partir du SNDS. En explorant la littérature existante, nous nous sommes rendu compte qu'il existait déjà plus d'une dizaine d'articles sur le sujet avec des

résultats divergents, nous souhaitons donc effectuer une revue systématique afin d'analyser plus finement et qualitativement tous les résultats fournis par la littérature à la lumière des mécanismes physiopathologiques déjà connus du DG, du CS et de leurs facteurs de risque. L'objectif de cette revue de la littérature était donc de compiler les hypothèses pouvant expliquer cette relation complexe.

Un premier postulat était que le DG augmenterait le risque de CS. En effet, les femmes diabétiques présenteraient un risque accru de 15 à 20 % de CS par rapport aux femmes non diabétiques (117–123). Les mécanismes sous-jacents comprennent les effets de l'hyperinsulinémie sur la disponibilité des stéroïdes sexuels (124,125) et la production d'IGF-1 (126,127). Des mécanismes qui ne dépendent pas des hormones ont également été suggérés, dont l'inflammation chronique avec des niveaux élevés de cytokines pro-inflammatoires, l'infiltration des dépôts lipidiques par des macrophages pro-inflammatoires et le stress oxydatif associé (128). Les femmes souffrant de DG présentent un risque accru de diabète de type 2, même dans les premières années suivant l'accouchement (129–131). Mais, le DG partage également des mécanismes d'action avec le diabète de type 2, notamment une hyperglycémie due à un dysfonctionnement des cellules  $\beta$  pancréatiques sur fond de résistance chronique à l'insuline. Il est donc possible que le DG soit associé à un risque accru de cancer (132–140), y compris de CS.

D'autres études ont montré un effet protecteur du DG sur le risque de CS. L'hypothèse émise, qui pourrait expliquer ces résultats, serait qu'une altération du milieu hormonal (comme une augmentation des taux sanguins de progestérone ou de cortisol par exemple) d'une grossesse compliquée par un DG conférerait une protection au tissu mammaire. Par ailleurs, la grossesse est connue pour induire la prolifération et la différenciation fonctionnelle des lobules et des canaux mammaires, ce qui pourrait expliquer la relation mise en évidence dans la littérature entre les antécédents de grossesse et le risque de CS (141–143).

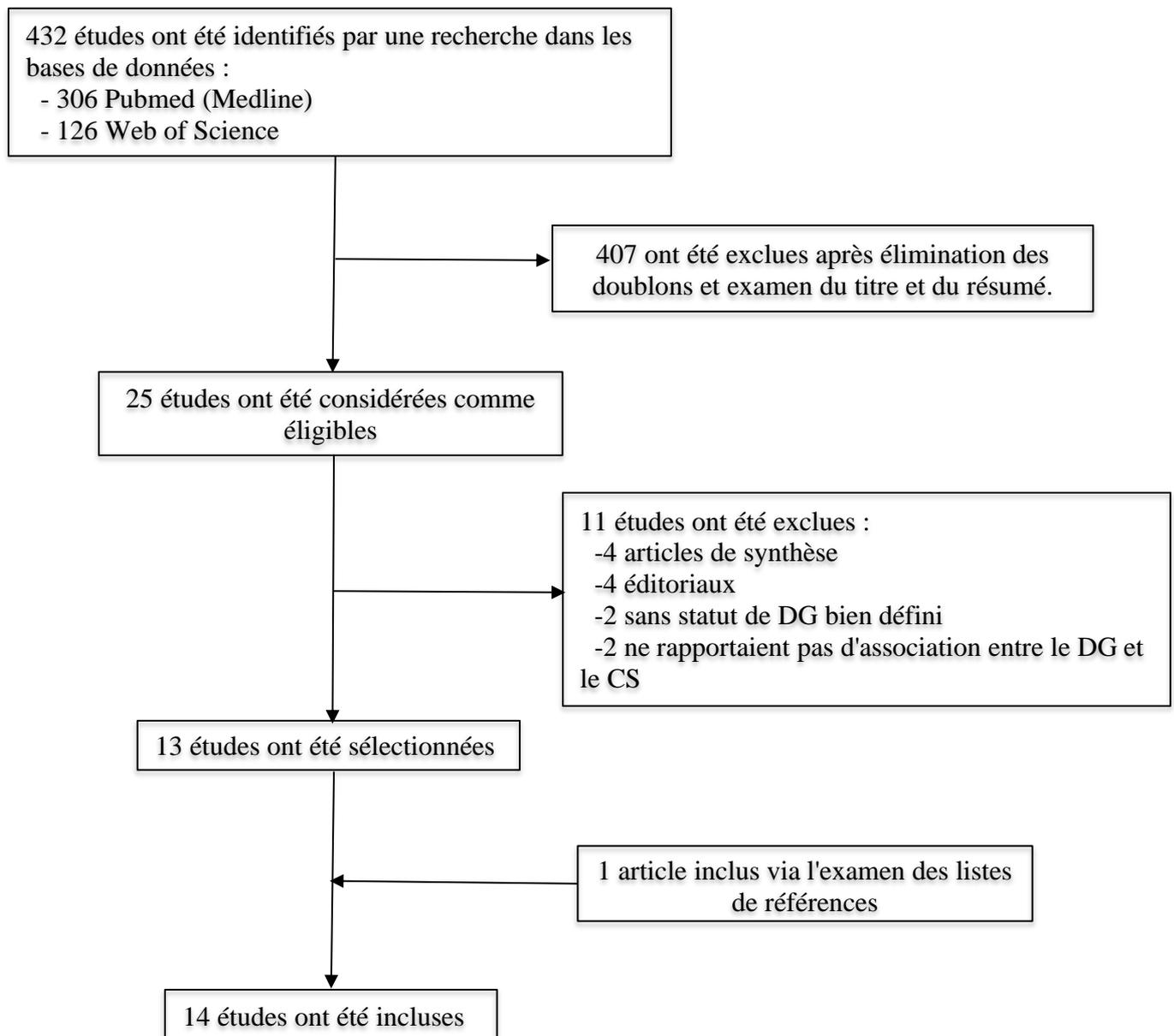
Plusieurs études ont exploré l'association entre le DG et le cancer. Cette association a été analysée quantitativement dans une méta-analyse de Wang et al. en 2020 (144). Cette méta-analyse ne portait pas exclusivement sur le CS et ne permettait pas de conclure sur le sujet de l'association entre DG et CS. De plus, les études sur ce sujet montrent des résultats divergents. Ils ont trouvé trois études rapportant une augmentation statistiquement significative du risque

de CS après un DG, deux études montrant un effet protecteur statistiquement significatif, et dix études présentant des résultats non significatifs. Dans leur méta-analyse, Wang et al. ont trouvé des résultats non significatifs, même s'ils étaient plutôt protecteurs pour les études cas-témoins (OR = 0,88 IC 95% [0,69 ; 1,12]) et péjoratifs pour les études de cohorte (RR de 1,08 IC 95% [0,89 ; 1,32]). De même, la méta-analyse la plus récente portant exclusivement sur le DG et le risque de CS (Xie et al., 2017) n'a trouvé aucune association significative entre le DG et le CS ni dans les études cas-témoins (OR 0,85 IC 95% [0,65 ; 1,10]), ni dans les études de cohorte (RR 1,00 IC 95% [0,80 ; 1,25]) (145). Xie et al. ont exposé plusieurs hypothèses physiopathologiques pour expliquer d'une part l'éventuelle augmentation du risque et d'autre part le possible effet protecteur du DG sur le risque de CS. Cependant, ils ne présentaient que peu d'éléments mettant en relation les forces, les limites et notamment les biais de chacune des études incluses dans leur méta-analyses et les hypothèses pouvant expliquer la divergence de résultats entre ces études. Au total, toutes ces analyses suggèrent que la relation entre le DG et le risque de CS peut être complexe. Dans ce travail, l'objectif de cette revue de la littérature était de compiler les hypothèses étudiées et les méthodes employées dans la littérature qui pourraient aider à mieux comprendre cette relation complexe et d'encourager de futures études sur ce sujet.

### **3.1.2 Matériel et méthodes**

Nous avons réalisé cette revue de la littérature en utilisant les bases de données PubMed/Medline et Web Of Science en utilisant les termes de recherche suivants : (gestational diabetes\* OR gestational diabetes mellitus OR GDM) AND (breast cancer OR breast tumour OR breast neoplasm\*). Les listes de référence des revues précédentes sur le sujet et des études incluses dans l'analyse ont également été recherchées pour toute autre étude pertinente. Toutes les extractions de données ont été effectuées par un seul chercheur. Cette revue de la littérature a été rédigée à l'aide de la check-list « Preferred Reporting Items for Systematic reviews and Meta-Analyses » (PRISMA). C'est une liste de 27 items, dont la présence est recommandée, concernant l'ensemble des parties de la revue systématique. Son objectif est d'aider les auteurs à améliorer le contenu de leur revue systématique (cf. Annexe 5).

Seuls les articles en langue anglaise publiés jusqu'au 1er juin 2021 ont été inclus. Nous avons inclus tous les types d'études cliniques rapportant l'association entre le DG et le CS avec des statistiques disponibles pour l'OR ou le RR et les IC à 95%. Les études sans données originales, y compris les revues, les commentaires, les éditoriaux et les méta-analyses, n'ont pas été incluses dans les résultats mais ont été utilisées dans la discussion pour apporter une perspective supplémentaire. La figure 8 montre le processus de sélection des études.



DG : Diabète Gestationnel ; CS : Cancer du sein

**Figure 8. Diagramme de flux du processus de sélection des articles**

Les données extraites comprenaient le premier auteur, l'année de publication, le pays, les spécificités de la population étudiée, la taille de l'échantillon, la durée du suivi, les critères de diagnostic du DG, le taux de DG, les critères de diagnostic du CS, l'âge ou le statut ménopausique au moment du diagnostic du CS, le taux de CS, les variables d'ajustement, les résultats principaux et les analyses de sensibilité ou supplémentaires. Ces éléments nous ont permis d'évaluer qualitativement chaque étude et de comparer leurs résultats en tenant compte des biais et des forces de chaque étude.

Le risque de biais potentiel et la qualité de chacune des études incluses ont été évalués à l'aide de l'échelle de Newcastle-Ottawa (« Newcastle-Ottawa Scale » ou NOS), qui est une échelle standardisée pour évaluer le biais et la qualité des études non randomisées. Sa conception, son contenu et sa facilité d'utilisation sont axés sur l'intégration des évaluations de la qualité dans l'interprétation des résultats des méta-analyses et des revues systématiques. Un système d'étoiles a été développé dans lequel une étude est jugée selon trois grandes perspectives : la sélection des groupes d'étude ; la comparabilité des groupes ; et la vérification de l'exposition ou du résultat d'intérêt pour les études cas-témoins ou de cohorte respectivement (cf. Annexe 6). Un manuel d'utilisation expliquant plus en détail les réponses à chaque item est également disponible sur le site :

[http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp).

### **3.1.3 Résultats**

#### **3.1.3.1 Recherche documentaire**

Au total, 732 articles ont été examinés par titre et résumé pour cette revue de la littérature. Parmi eux, 25 ont été retenus pour l'examen du texte intégral. Finalement, 14 études répondant aux critères d'inclusion ont été incluses dans l'analyse, dix d'entre elles avaient un design de type cohorte (134–139,146–149) et quatre de type cas-témoin (150–153). Une étude cas-témoins, d'Ardalan et al. (153), a été ajoutée après examen des listes de références parce qu'elle répondait aux critères d'inclusion ; elle a étudié la relation entre l'âge gestationnel à l'accouchement et le risque de CS, mais a également rapporté une mesure de l'association entre le DG et le CS.

#### **3.1.3.2 Caractéristiques des études incluses**

Le tableau 6 présente les caractéristiques détaillées des dix études de cohorte et le tableau 7 les caractéristiques détaillées des quatre études cas-témoins. Les études sont classées par année de publication dans l'ordre croissant.

##### **Tableau 6. Caractéristiques des études de cohorte**

Premier Auteur	Année de publication	Pays	Spécificités de la population étudiée	Taille de l'échantillon	Durée de suivi	Critères diagnostiques du DG	Taux de DG	Critères diagnostiques du CS	Taux de CS	Variables d'ajustement	RR [IC 95%]	Analyses de sensibilité ou supplémentaires
Perrin	2007	Israël	Accouchements entre 1964 et 1976	37 980	Jusqu'en 2015	Dossiers médicaux	1%	Registre du cancer	4,3%	Âge, parité, classe sociale, origine ethnique, éducation, statut d'immigration	1,5 [1,0; 2,1] P=0.03	Âge au moment du diagnostic de CS : < 50 ans RR=1 [0,5; 2,1] ≥ 50 ans RR=1,7 [1,1; 2,5]
Sella	2011	Israël	Agés entre 15 et 50 ans Dépisté entre le 13 mars 1995 et le 27 mai 2009	185 315	Jusqu'au 1 <sup>er</sup> juin 2009 (moyenne de 5 ans)	Dossiers médicaux	6%	Registre du cancer	0,3%	Âge, niveau socio-économique, tabagisme, IMC, gravidité, nombre de visites chez le médecin généraliste.	0,87 [0,63 ; 1,20]	-
Bejaimal	2016	Canada	Âgés de 20 à 50 ans Dépistés entre le 1 <sup>er</sup> janvier 1995 et le 4 juillet 2008	149049 : 49684 femmes atteintes de DG appariées selon l'âge et l'année de l'accouchement dans un rapport de 1 :2 avec des femmes enceintes sans DG.	Jusqu'au 31 décembre 2011	Dossiers médicaux	4,6%	Registre du cancer	-	Nombre de visites chez le médecin	0,86 [0,75; 0,98]	Après ajustement pour le diabète ultérieur, l'association entre le DG et le risque de CS est restée significative (HR 0,85 [0,74 ; 0,99]).

Fuchs	2017	Israël	Accouchements entre 1988 et 2013	104 715	Jusqu'en 2013	Dossiers médicaux	9,4%	Dossiers médicaux	0,5%	Résultats univariés uniquement	1,8 [1,5; 2,1]	-
Powe	2017	Etats-Unis	Uniquement les infirmières qui ont déclaré un accouchement au début de la cohorte en 1989 ou avec un accouchement incident jusqu'en 2001.	86 972	Jusqu'en 2013	Autodéclaré, 94% confirmé par les dossiers médicaux	6%	Dossiers médicaux	2,7%	Âge, IMC à 18 ans, prise de poids depuis l'âge de 18 ans, taille, activité physique, consommation d'alcool, âge des premières règles, registre de naissance, allaitement, utilisation d'un traitement hormonal, antécédents de cancer du sein chez la mère ou la sœur, antécédents de maladie mammaire bénigne, origine ethnique, mammographie au cours des deux dernières années.	0,72 [0,58; 0,89]	DT2 seul : HR=0,69 [0,40; 1,18]. DG+T2D : HR=0,26 [0,10; 0,68] DG et récepteurs hormonaux positifs : HR 0,65 [0,50-0,84], DG et récepteurs hormonaux négatifs : HR 0,96 [0,60-1,52].
Park	2017	Etats-Unis	Âgées de 35 à 74 ans, toutes exemptes de cancer du sein	39 198	Jusqu'au 14 août 2015	Autodéclaré	4,2% (0,9% de gross	Autodéclaré, 81% confirmé	5,4%	Âge, origine ethnique, éducation, âge de la première naissance,	1,10 [0,88; 1,36]	1 grossesse avec DG : HR=0,94 [0,73 ; 1,22], plus d'une

			et sœurs de femmes atteintes d'un cancer du sein entre 2003 et 2009		(moyenne: 7,4 ans)		esses multiples avec DG)	é par les dossiers médicaux		âge des premières règles, poids à l'âge de 10 ans, IMC à 30-39 ans, activité physique pendant l'enfance et l'adolescence.		grossesse avec DG : HR= 1,68 [1,15 ; 2,44].
Han	2018	Corée de Sud	1ère grossesse entre 2004 et 2005 et participer à un "examen national de dépistage sanitaire" au moins 2 ans avant leur premier accouchement	102 900	Jusqu' en 2015	Base de données médico-administrative Codes CIM 10 O 24.4 et O 24.9	4,83 %	Base de données médico-administrative Codes CIM 10	0,7%	Âge, tabagisme, IMC avant la grossesse	1,15 [0,83; 1,58]	-
Peng	2019	Taiwan	Tous les Accouchements entre 2002 et 2012	990 572	Jusqu' au 31 décembre 2013	Base de données médico-administrative Codes 648-250 de la CIM 9	4,8%	Base de données médico-administrative Codes CIM 9	0,47 %	Âge, hypertension, dyslipidémie, maladie du foie, infertilité et maladie rénale.	1,23 [1,09; 1,39]	-
Pace	2020	Canada	34294 accouchements sélectionnés aléatoirement	68 588	Jusqu' en 2012 (Moy	2 codes de DG ou plus	-	Résumés de sortie d'hôpital	1,1%	Hypertension gestationnelle, accouchement prématuré, taille du	0,93 [0,80;1,09]	-

			entre avril 1990 et le 31 décembre 2007 et présentant au moins deux codes de diagnostic de DG et ont été appariés 1:1 à 34294 accouchements sans DG (par année d'accouchement, âge, région).		enne: 13,1 ans)			l utilisant les codes CIM		nourrisson, parité, comorbidité antérieure, indice de privation matérielle, origine ethnique.		
Bertrand	2020	Etats-Unis	Mères afro-américaines entre mars 1995 et mars 2017	41 767	Jusqu' en mars 2017	Autodéclaré	4,9%	Autodéclaré et confirmé dans des registres du cancer ou des dossiers médicaux.	4%	Âge, IMC à l'âge de 18 ans, IMC récent, parité, âge aux ménarches, âge au premier accouchement, durée de la contraception orale et antécédents familiaux de CS.	0,98 [0,77;1,25]	Pas de résultats significatifs après ajustement sur le statut des récepteurs aux œstrogènes

DG : Diabète gestationnel ; CS : Cancer du sein ; RR : Risque relatif ; HR : Hazard Ratio ; IMC : Indice de Masse Corporelle ; DT2 : Diabète de type 2 ; CIM : Classification internationale des maladies ;

**Tableau 7. Caractéristiques des études cas-témoins**

Premier Auteur	Année de publication	Pays	Spécificités de la population étudiée	Taille de l'échantillon	Critères diagnostics du DG	Taux de DG	Critères diagnostics du CS	Variables d'ajustement	OR [IC 95%]	Analyses de sensibilité ou supplémentaires
Troisi	1998	Etats-Unis	Âge compris entre 20 et 44 ans	3174	Autodéclaré	7,5%	Dossiers médicaux	Âge aux ménarches, mammographie, alcool, IMC, site, race, variables combinées représentant la parité et l'âge au premier accouchement.	1,1 [0,7; 1,4]	/
Rollison	2008	Etats-Unis	Femmes blanches non hispaniques, hispaniques, indiennes (âge moyen : 56 ans)	2523	Autodéclaré	3,2%	Registre de cancer	Âge, site de l'étude, statut ménopausique, IMC, IMC à 15 ans, parité, âge aux ménarches, activité physique, antécédents familiaux de CS, antécédents d'allaitement.	0,71 [0,52; 0,98]	Stratification sur l'âge au moment de l'apparition du DG : 0,56 [0,38;0,82] – <35 ans 1,34 [0,72; 2,52] ≥ 35 ans Stratification sur l'âge au moment de l'apparition du DG et le statut des récepteurs hormonaux : <35 ans et RO + 0,52 [0,31 ; 0,85]. <35 ans et RP + 0,53 [0,32 ; 0,89] ≥35 ans et RO -

										2,52 [1,07; 5,53] ≥35 ans et RP - 3,08 [1,45 ; 6,54]
Brasky	2013	Etats-Unis	Âge compris entre 35 et 79 ans	2812	Autodéclaré	3,3%	Dossiers médicaux	Âge, niveau d'éducation, antécédents de maladie mammaire bénigne, antécédents familiaux de CS, âge à la première grossesse, nombre de grossesses, statut ménopausique, âge à la ménopause.	0,79 [0,48; 1,30]	-
Ardalan	2016	Etats-Unis	Femmes chez qui on a diagnostiqué un cancer du sein en préménopause (âgées de 21 à 49 ans) dans les 5 ans suivant l'accouchement d'un bébé vivant entre le 1 <sup>er</sup> janvier 1994 et le 31 décembre 2003.	630	Actes de naissance enregistrés	1,4%	Registre de cancer	Âge de la mère à l'accouchement, race/ethnicité, niveau d'éducation, poids à la naissance, parité, âge gestationnel, prise de poids pendant la grossesse, tabagisme, alcoolisme, déclenchement du travail, hypertension gestationnelle.	1,62 [0,30; 8,68]	-

DG : diabète gestationnel ; CS : Cancer du sein ; OR : Odds Ratio ; IMC : Indice de Masse Corporelle ; RO, récepteur d'œstrogène ; RP, récepteur de progestérone

## Études de cohorte

Sur les dix études de cohorte, cinq ont été menées en Amérique du Nord, trois en Israël et deux en Asie. La première a été publiée en 2008, et au moins une étude a été publiée chaque année entre 2015 et 2020. Les cohortes étaient plutôt de grandes tailles, de 37 980 sujets pour Perrin et al. (2008) (146) à 990 572 sujets pour Peng et al. (2019) (138). Cependant, la grande majorité de ces études ont recruté leur population avant la publication des critères diagnostiques les plus récents du DG en 2010 (154). Les cohortes les plus anciennes sont pour la plupart des cohortes prospectives basées sur des dossiers médicaux et des registres de cancer, tandis que les cohortes les plus récentes étaient pour la plupart rétrospectives et basées sur des données autodéclarées ou médico-administratives.

Il faut également noter que certaines études avaient des critères d'inclusion spécifiques, par exemple, Powe et al. (147) n'incluaient que des infirmières ; Park et al. (148) n'incluaient que les sœurs de femmes ayant un CS diagnostiqué ; Han et al. (136) n'incluaient que les patientes ayant passé un examen de dépistage national dans les deux ans précédant la grossesse, et Bertrand et al. (149) n'incluaient que des femmes afro-américaines. Le suivi était long dans certaines études, jusqu'à plusieurs décennies dans l'étude de Perrin et al. (suivi médian de 34 ans), mais dans d'autres études, le suivi ne durait que quelques semaines pour certains des cas (par exemple, Bertrand et al. (149) et Fuchs et al. (137)). Les taux de DG et de CS différaient entre ces quatre études, allant de 1% à 9,4 % et de 0,3% à 5,4 %, respectivement. Seules trois études de cohorte ont clairement indiqué si le CS était in situ ou invasif (134,147,149) et deux études de cohorte ont fait la différence entre les cancers hormono-dépendants et non-hormono-dépendants (147,149).

Les résultats de ces études différaient probablement aussi en raison du choix des variables d'ajustement. Nous avons trouvé trois études montrant une augmentation statistiquement significative du risque de CS après un DG, deux études montrant un effet protecteur statistiquement significatif et cinq études montrant des résultats non significatifs. Concernant les variables d'ajustement, il faut noter que Fuchs et al. ont publié uniquement des résultats univariés et ont trouvé un effet protecteur statistiquement significatif (137). En revanche, Bejeimal et al. n'ont pris en compte que le nombre de visites chez le médecin dans les 3 premières années de suivi, afin d'ajuster sur une éventuelle différence de fréquence de contact avec le système de santé, dans l'hypothèse que cela influencerait la possibilité de diagnostic d'un éventuel cancer. Ils ont montré un risque accru statistiquement significatif (139). La grande majorité des autres études de cohorte ont ajusté sur l'âge, la parité, l'ethnicité ou l'IMC. Quatre études (134,135,146,148) ont inclus des variables socio-économiques

(par exemple, l'éducation, la classe sociale) et cinq études (134,136,147–149) ont pris en compte les habitudes de vie (tabac, alcool, activité physique, contraception).

### Études cas-témoins

Les quatre études cas-témoins ont été menées aux États-Unis. Elles ont été publiées entre 1998 et 2016 et ont inclus entre 630 et 3 174 sujets. Ces études étaient basées sur des données autodéclarées pour le diagnostic de DG, à l'exception d'Ardalan et al. (153), qui était basée sur des certificats de naissance. Le diagnostic de CS était basé sur des dossiers médicaux, à l'exception de Rollison et al. (151), qui se basait sur un registre du cancer. Une seule étude cas-témoins a clairement précisé la caractéristique in situ ou invasif du CS (151) et deux études cas-témoins ont distingué les cancers hormonodépendants et non hormonodépendants (151,152).

Dans les études de Troisi et al. (150) et Ardalan et al. (153), les sujets étaient plus jeunes que dans les deux autres études et le taux de DG variait de 1,4% à 7,5%. Il convient également de noter que Rollison et al. (151) n'ont inclus que des femmes blanches non hispaniques, hispaniques et amérindiennes.

Parmi les études cas-témoins, nous n'avons trouvé qu'une seule étude montrant un effet protecteur du DG statistiquement significatif et trois études montrant des résultats non significatifs. En ce qui concerne les variables d'ajustement, toutes les études cas-témoins ont inclus l'âge et la parité, trois études ont pris en compte le statut ménopausique et l'origine ethnique, deux ont pris en compte l'IMC, deux ont pris en compte les antécédents familiaux de CS, l'éducation et la consommation d'alcool. Enfin, une étude a ajusté sur l'activité physique et une autre sur le tabagisme.

### 3.1.3.3 Risque de biais dans les études incluses

Le tableau 8 et le tableau 9 montrent, respectivement, le risque de biais pour les études de cohorte et les études cas-témoins, évalué par un classement par étoiles selon l'échelle NOS.

**Tableau 8. Risque de biais et évaluation de la qualité des études de cohorte**

Premier auteur	Année de publication	Sélection				Comparabilité		Critères de jugement			Total
		Représentativité de la cohorte exposée	Sélection de la cohorte non exposée	Détermination de l'exposition	Démonstration que le critère de jugement n'était pas présent au début de l'étude	Contrôles pour le facteur le plus important	Contrôles pour tout facteur supplémentaire	Évaluation des critères de jugement	Le suivi a-t-il été suffisamment long pour que le critère de jugement se produise ?	Adéquation du suivi des cohortes	
Perrin	2008	*	*	*	*	*	*	*	*	*	9
Sella	2011	*	*	*	*	*	*	*	*		8
Bejaimal	2015	*	*	*	*	*	*	*	*	*	9
Fuchs	2016	*	*	*	*	*	*	*			7
Powe	2017		*	*		*	*	*	*		6
Park	2017	*	*		*	*	*		*		6
Han	2018		*	*	*	*	*	*	*	*	8
Peng	2019	*	*	*	*	*	*	*	*	*	9
Pace	2020	*	*	*		*	*	*	*	*	8
Bertrand	2020	*	*			*	*	*			5

Note : Les évaluations sont basées sur l'échelle de Newcastle-Ottawa. Les critères évalués de qualité "élevée" sont identifiés par une "\*".

**Tableau 9. Risque de biais et évaluation de la qualité des études cas-témoins**

Premier auteur	Année de publication	Sélection				Comparabilité		Exposition			Total
		La définition des cas est-elle adéquate ?	Représentativité des cas	Sélection des témoins	Définition des témoins	Contrôles pour le facteur le plus important	Contrôles pour tout facteur supplémentaire	Détermination de l'exposition	Même méthode de détermination pour les cas et les témoins	Taux de non-réponse	
Troisi	1998	*	*	*		*	*		*		6
Rollison	2008		*	*		*	*		*	*	6
Brasky	2013	*	*	*		*	*		*	*	7
Ardalan	2016		*	*	*			*	*	*	6

Note : Les évaluations sont basées sur l'échelle de Newcastle-Ottawa. Les critères évalués de qualité "élevée" sont identifiés par une "\*".

Parmi les dix études de cohorte, trois études avaient neuf étoiles, trois études avaient huit étoiles, une étude avait sept étoiles, trois études avaient six étoiles et une autre avait cinq étoiles. Deux études ont sélectionné un groupe spécifique (infirmières, participants à un dépistage) (136,147). Dans deux études, le diagnostic de DG était autodéclaré (148,149). Dans trois études, il n'a pas été décrit de recherche d'absence d'antécédents de CS chez les patientes incluses (135,147,149). Dans deux études, le suivi n'était pas suffisamment long (137,149). Dans cinq études, le taux de suivi était inférieur à 94% et/ou il n'y avait pas de description des patientes perdues de vue (134,137,147–149). Dans l'étude de Park et al., le diagnostic de CS a été auto-déclaré et confirmé par les dossiers médicaux dans seulement 81% des cas (148).

Parmi les quatre études cas-témoins, trois études ont obtenu six étoiles et une étude a obtenu sept étoiles. Deux études ne disposaient pas d'une définition de cas adéquate, c'est à dire avec une validation indépendante (plus d'un(e) personne/enregistrement/temps/processus pour extraire l'information dans les dossiers médicaux). Le simple couplage des dossiers à l'aide de registres, qui est la méthode utilisée dans ces deux études, n'est pas considéré comme adéquat (151,153). Trois études ne décrivaient pas leur source pour la définition des témoins et les entretiens n'étaient pas réalisés en aveugle pour le statut de cas/témoin (150–152). Dans l'étude de Troisi et al., le taux de réponse à l'entretien était plus faible pour les témoins que pour les cas (150). Dans l'étude d'Ardalan et al., les cas et les témoins n'ont pas été appariés ou ajustés sur les facteurs de confusion connus de la relation entre le DG et le CS dans la conception des groupes, étant donné que l'objectif principal de cette étude était d'étudier la relation entre l'âge gestationnel et le CS (153).

### 3.1.4 Discussion

Dans cette revue de la littérature, nous avons trouvé 14 études examinant la relation entre le DG et le risque de CS. Parmi les 10 cohortes, trois études ont montré une augmentation statistiquement significative du risque de CS après un DG, deux études ont montré un effet protecteur statistiquement significatif, et cinq études ont rapporté des résultats non significatifs. Parmi les quatre études cas-témoins, une étude a montré un effet protecteur statistiquement significatif et trois études ont rapporté des résultats non significatifs. Les 14 études étaient de qualité variable et utilisaient des méthodologies différentes. En utilisant les évaluations de l'échelle NOS, seules 3 des 14 études ont obtenu toutes les étoiles, ce qui pourrait expliquer en partie la diversité des résultats.

Les études cas-témoins ont toutes été menées aux États-Unis. Les études de cohorte montrant soit des résultats non significatifs, soit un effet protecteur statistiquement significatif ont également été réalisées pour la plupart en Amérique du Nord. Toutes les études montrant une augmentation statistiquement significative du risque ont été réalisées sur le continent asiatique (Israël, Chine). En ce qui concerne l'origine ethnique et le risque de CS, Perrin et al. (146) ont montré que les femmes originaires d'Asie occidentale et d'Afrique du Nord ont un risque significativement plus faible de développer un CS. Par conséquent, une première hypothèse serait que l'ethnicité, la culture et/ou le pays de résidence pourraient avoir une influence sur la relation entre le DG et le CS.

Une deuxième hypothèse serait que l'hétérogénéité des résultats pourrait s'expliquer en partie par les différentes définitions de l'exposition (c'est-à-dire le diagnostic de DG), ce qui donnerait lieu à l'obtention de différents taux de DG selon l'étude. Dans les études cas-témoins, les antécédents de DG ont été déclarés par les patientes elles-mêmes, à l'exception de l'étude d'Ardalan et al. (153), qui a utilisé des certificats de naissance, avec un taux de DG de 1,4 %, mais l'exhaustivité de ces certificats est discutable. Dans l'étude de Troisi et al. (150), ils ont trouvé un taux de DG de 7,5%. Toutefois, les femmes étaient jeunes (entre 20 et 44 ans) ; on peut donc supposer que le délai entre le diagnostic de DG et les réponses des sujets était plutôt court, ce qui les aiderait à se souvenir de cet événement. En revanche, pour les deux autres études cas-témoins, le taux de DG était alors plus faible, de l'ordre de 3,3%, probablement car ces études portaient sur des populations plus âgées, ce qui pourrait avoir

entraîné un biais de mémoire. Il faut également tenir compte du fait que les critères de diagnostic du DG ont évolué au cours des 50 dernières années, les recommandations actuelles datant de 2010, permettant de diagnostiquer un plus grand nombre de DG (154). Par exemple, l'étude de Perrin et al. (146), portant sur les grossesses entre 1964 et 1976 a montré un taux de DG de 1%, alors que les études de cohorte plus récentes rapportent un taux de DG entre 4,5% et 6%.

Se pose également la question du traitement et du suivi du DG. Une hypothèse serait que le DG non équilibré pourrait augmenter le risque de CS mais pas le DG équilibré par un régime alimentaire. En effet, en 2004, Dawson et al. ont étudié plusieurs indices glycémiques chez un échantillon de 753 femmes sans antécédent de diabète. Le taux de DG n'est pas mentionné dans l'article. Toutefois, ils ont montré qu'après ajustement sur l'âge, l'IMC et le tabagisme, le risque de CS était de 1,2 à 5,1 fois plus élevé, respectivement, dans le quartile le plus élevé d'indice glycémique par rapport au quartile le plus bas (64). Si le mécanisme physiopathologique d'un risque accru de CS du fait de la combinaison d'une hyperglycémie couplée à une insulino-résistance et une hyperinsulinémie était retenu, alors un DG équilibré par un suivi régulier et un régime alimentaire devrait générer moins d'excès de risque de CS qu'un DG mal suivi ou non équilibré.

Nous pourrions aller plus loin en supposant que l'effet protecteur du DG trouvé dans plusieurs études pourrait être expliqué par l'amélioration des habitudes de vie (activité physique, alimentation) suite à la découverte de ce DG. Cependant, dans notre revue de littérature, la seule étude qui a exploré ces éléments (Powe et al. (147)) rapporte que le DG a bien un effet protecteur, mais qu'il n'y a pas de différences significatives dans l'alimentation et l'activité physique entre les femmes avec et sans antécédents de DG. Néanmoins, il faut également noter que Powe et al. (147) ont étudié un échantillon de femmes exerçant la profession d'infirmière, qui sont d'une part plus susceptibles d'être conscientes de leur santé, mais qui travaillent également souvent de nuit, ce qui est connu comme un facteur de risque de CS (155).

Selon la littérature, certains mécanismes physiopathologiques pourraient expliquer l'effet protecteur retrouvé dans certaines études, comme l'altération du milieu hormonal d'une grossesse compliquée par un DG conférant une protection au tissu mammaire. Nous avons vu plus haut que la grossesse est connue pour induire la prolifération et la différenciation fonctionnelle des lobules et des

canaux mammaires, ce qui pourrait expliquer l'effet protecteur des antécédents de grossesse sur le risque de CS. De plus, les différences de facteurs de croissance ou d'hormones circulant lors des grossesses avec DG pourraient également avoir un impact sur ces processus et sur le risque futur de CS (141–143).

De même, les critères de diagnostic du CS diffèrent d'une étude à l'autre et peuvent conduire à divers biais de classification, comme dans l'étude de Park et al. (148) où le diagnostic de CS était autodéclaré et n'a été confirmé dans les dossiers médicaux que dans 81% des cas. L'utilisation des données des registres de cancer pourrait entraîner un biais de classification s'ils ne sont pas exhaustifs, mais ce risque est très faible car la plupart de ces registres sont décrits comme étant exhaustifs à plus de 90%. La question de l'exhaustivité des bases de données médico-administratives, qui s'appuient sur les admissions hospitalières, se pose notamment pour les cancers qui ne nécessitent pas toujours une hospitalisation, ce qui est le cas du CS. Ainsi, on peut supposer que les patientes du groupe témoin, repérées dans les données hospitalières, peuvent avoir un CS non détecté, contribuant aussi au biais de classification. Par ailleurs, seules quatre études ont clairement précisé le caractère in situ ou invasif du cancer (134,147,149,151) et quatre études seulement différenciaient les cancers hormono-dépendants et non hormono-dépendants (147,149,151,152), alors que ces cancers n'ont pas la même physiopathologie et les mêmes facteurs de risque.

Les études incluses présentaient également des variables d'ajustement très hétérogènes. Parmi les six études de cohorte qui ne se concentraient pas uniquement sur le CS, la plupart n'ont pas ajusté sur plusieurs facteurs de risque connus de CS (par exemple, l'utilisation de contraceptifs oraux, l'allaitement, la consommation d'alcool ou les antécédents familiaux de CS). Seules Powe et al. (147) et Bertrand et al. (149) ont ajusté pour ces facteurs de risque, à l'exception de la consommation d'alcool. Parmi les études cas-témoins, Troisi et al. (150) et Ardalán et al. (153) ont ajusté sur la consommation d'alcool, tandis que Rollison et al. (151) et Brasky et al. (152) ont ajusté sur les antécédents familiaux de CS ; aucune étude cas-témoins n'a ajusté sur l'utilisation de contraceptifs oraux. Il existe également deux facteurs de risque de CS qui sont également des facteurs de risque connus de DG : l'obésité et la sédentarité. On peut donc supposer que ces facteurs peuvent jouer un rôle de facteur de confusion dans la relation entre le CS et le DG (156), ainsi que le diabète de type 2. En effet, plusieurs études ont montré que le DG augmentait le risque de diabète de type 2 (129,131),

et il est reconnu que le diabète de type 2 augmente le risque de CS (157). Bien que la grande majorité des études aient ajusté sur l'IMC, seules trois études ont ajusté sur l'activité physique et aucune n'a ajusté sur le diabète de type 2 dans les analyses principales. Cependant, dans leur analyse secondaire, Bejaimal et al. (139) ont montré qu'après ajustement pour le diabète de type 2, l'association entre le DG et le risque de CS restait significative avec un effet protecteur (HR = 0,85 IC 95% [0,74 ; 0,99]). Après une stratification sur le diabète de type 2, Powe et al. (147) ont également trouvé que le DG avait un effet protecteur statistiquement significatif sur le risque de CS (HR = 0,72 IC 95% [0,58 ; 0,89]). Dans une analyse supplémentaire de leur population d'infirmières, ils ont montré que les femmes qui avaient des antécédents de DG et de diabète de type 2 avaient un risque encore plus faible de CS (HR = 0,26 IC 95% [0,10 ; 0,68]) par rapport à celles n'ayant qu'un antécédent de DG.

Plusieurs études ont également montré que l'âge est un facteur important à prendre en compte. Tout d'abord, en ce qui concerne l'âge au début du DG, Rollison et al. (151), qui ont montré un effet protecteur statistiquement significatif du DG sur le risque de CS dans leur population, ont trouvé un effet encore plus protecteur chez les femmes qui avaient moins de 35 ans au début du DG (OR = 0,56 IC 95% [0,38 ; 0,82]). Deuxièmement, l'âge au moment du diagnostic du CS, probablement en lien avec le statut ménopausique des patientes, influencerait la relation entre le DG et le risque de CS. Par exemple, Perrin et al. (146) ont montré qu'il existe un risque significativement plus élevé de CS diagnostiqué après 50 ans en cas d'antécédents de DG (RR=1,7 IC 95% [1,1 ; 2,5]).

En ce qui concerne les sous-types de CS, trois études ont ajusté ou stratifié sur le statut des récepteurs aux œstrogènes (RO) et/ou à la progestérone (RP). Aucune étude n'a travaillé sur le statut HER2. Par exemple, Bertrand et al. (149) n'ont pas montré d'association statistiquement significative entre le DG et le CS selon le statut des RO. Cependant, Rollison et al. (151) et Powe et al. (147) qui avaient déjà trouvé un effet protecteur statistiquement significatif du DG sur le risque de CS, ont trouvé une relation encore plus protectrice du DG pour les CS à récepteurs hormonaux positifs. Cependant, cette relation statistiquement significative n'a été trouvée par Rollison et al. (151) que

chez les femmes présentant un DG diagnostiqué avant l'âge de 35 ans. Il serait intéressant d'étudier plus avant cette relation entre le DG et les différents sous-types de CS.

Une seule étude s'est penchée sur l'effet d'antécédents multiples de DG sur le risque de CS. Cette étude de Park et al. (148) a été menée dans une population particulière de sœurs de femmes dont le diagnostic de CS a été établi. Ils ont étudié l'effet cumulatif du nombre de grossesses avec DG : ils trouvent un effet non-significatif d'un seul antécédent de grossesse avec DG (RR = 0,94 IC 95% [0,73 ; 1,22]) alors qu'il devient significatif (RR = 1,68 IC 95% [1,15 ; 2,44]) chez les femmes ayant plus d'un antécédent de grossesse avec DG.

Les caractéristiques génomiques n'ont été abordées par aucune des études incluses dans notre revue de la littérature, même si les mutations génétiques peuvent influencer le CS mais aussi le DG. Nous avons trouvé plusieurs études montrant une fréquence accrue de certaines mutations chez les patientes atteintes de CS. Les mutations germinales les plus connues sont BRCA1 et BRCA2, mais d'autres, comme PALB2 ou TP53, peuvent également être retrouvées (158–161). De même, plusieurs hypothèses ont été formulées concernant le lien entre les mutations génomiques, telles que GCK, les antigènes HLA, INSR, IGF2, HNF4A, INS-VNTR, HNF4a, CAPN10, MBL2, KCNJ11, ABCC8, ND1, TCF7L2, ADIPOQ et PAI-1, et le développement du DG (162–164). Chen et al. ont également suggéré que les mutations de l'ADN mitochondrial pouvaient favoriser le développement du DG chez certaines patientes (165). En outre, il existe des mutations génétiques qui pourraient avoir un impact à la fois sur le développement du DG et du CS, telles que les mutations du gène ACE I/D (166,167). La littérature fournit ainsi plusieurs hypothèses sur le lien entre le DG, les mutations génétiques et le risque de cancer. Par exemple, Witczak et al. ont expliqué que le DG est défini comme une intolérance aux glucides entraînant une hyperglycémie diagnostiquée pour la première fois pendant la grossesse. Il est associé à des niveaux accrus de stress oxydatif dus à la surproduction de dérivés réactifs de l'oxygène (DRO). La surproduction de DRO induit l'oxydation des protéines, la peroxydation des lipides et divers types de dommages à l'ADN (168). Toutefois, on ne sait pas si le suivi et les traitements actuellement établis pour le DG pourraient réduire les conséquences ainsi décrites sur l'ADN. Il serait donc intéressant dans de futures études de clarifier le lien entre les génotypes et

l'influence qu'ils pourraient avoir sur le risque de DG et/ou de CS afin d'améliorer la prévention et le traitement.

En ce qui concerne la durée du suivi nécessaire pour étudier la relation entre le DG et le risque de CS, la littérature existante n'a pas clairement établi la durée idéale. Si le suivi est inférieur à 15 ans, comme dans la grande majorité des cohortes de cette revue, il y a un risque d'étudier principalement des CS pré-ménopausiques. Un suivi de plus de 15 ans permettrait au contraire d'évaluer les effets du DG à plus long terme, notamment sur le CS ménopausique. Si on prend l'exemple de l'étude de l'influence du diabète de type 2 ou de l'obésité, à notre connaissance, les études actuelles ont montré chez les femmes atteintes de diabète de type 2 ou d'obésité, soit qu'il n'y avait pas d'association significative, soit qu'il y avait un risque statistiquement plus faible de CS pré-ménopausique (134,146,169). Cependant, les femmes ayant un IMC élevé pourraient avoir un risque significativement plus élevé de CS post-ménopausique (170). Toutefois, ces études manquent souvent de puissance puisque le taux de CS est généralement inférieur à 1% dans les populations étudiées.

Notre revue de la littérature présente plusieurs limites. Tout d'abord, elle a été menée par un seul examinateur. Cependant, nous avons pu nous appuyer sur deux méta-analyses récentes (144,145) pour garantir l'exhaustivité de notre revue. Deuxièmement, il est difficile de faire une comparaison directe entre les études. En effet, les populations étudiées avaient soit un nombre de cas trop faible pour avoir une puissance statistique suffisante, soit elles étaient difficiles à généraliser car souvent réalisées dans des populations particulières (soit réalisées aux Etats-Unis avec des ethnies particulières, soit avec une cohorte d'infirmières, soit basée sur la participation à un examen national de dépistage sanitaire, soit avec une cohorte de sœurs de femmes atteintes de CS). On peut également souligner des biais de sélection potentiels, ainsi qu'un suivi incomplet dans les études de cohorte et de faibles taux de réponse dans les études cas-témoins. Un biais de classification de l'exposition est également possible car plusieurs études incluses dans cette revue systématique ont utilisé des autodéclarations au lieu d'un diagnostic clinique objectif basé sur des tests biochimiques (171). Les variables d'ajustement sont également très hétérogènes et ne correspondent pas toujours aux facteurs de risque connus du DG ou du CS.

### 3.1.5 Conclusion

En conclusion, les résultats des études portant sur le lien entre le DG et le CS semblent divergents et doivent donc être interprétés avec prudence, car la majorité des études manquent de puissance et/ou présentent de nombreux biais. Le lien entre le DG et le CS mériterait donc d'être approfondi lors de nouvelles études permettant une meilleure compréhension de cette association potentielle, mais aussi des mécanismes physiopathologiques sous-jacents, est nécessaire. L'ethnicité et/ou la culture pourraient jouer un rôle dans la relation entre le DG et le CS, les taux de CS et de DG étant différent selon les populations étudiées. Il serait souhaitable que les études soient menées non seulement en Amérique du Nord et en Asie, mais aussi dans le reste du monde, et qu'elles soient représentatives de leur population. En outre, on peut également supposer que la culture (dont le régime alimentaire) et la situation géographique joueraient un rôle sur les facteurs de confusion potentiels de la relation entre le DG et le CS comme le diabète, l'obésité et l'activité physique. Un ajustement sur ces différents facteurs de confusion potentiels et sur les facteurs de risque connus de CS et de DG permettrait également de mieux cibler l'effet propre du DG sur le risque de CS. Il serait également souhaitable que ces futures études aient une définition claire, précise et actualisée du DG, ainsi que des différents types de CS (chez les femmes ménopausées ou non, cancer hormono-dépendant ou non, invasif ou non). De plus, il serait utile de savoir quelles prises en charge ou traitements ont été mis en place après le diagnostic du DG et d'étudier l'effet de plusieurs antécédents de grossesses avec DG.

Au total, il apparaîtrait que la complexité apparente de la relation entre le DG et le CS pourrait être expliquée par le nombre important de facteurs pouvant l'influencer. Une meilleure compréhension de cette relation nécessiterait de mettre en œuvre plusieurs études, de bonne qualité sur des populations différentes, et prenant en compte ces facteurs. Or, on a pu identifier, dans cette revue de la littérature, les difficultés que pouvaient avoir les épidémiologistes à collecter des données aussi diversifiées. En effet, les registres de cancer bien qu'ayant des données très complètes sur un grand nombre de variables, possèdent rarement des données concernant le DG ou ses facteurs de risque propres. Il serait nécessaire de pouvoir interroger directement les patientes, ce que certaines études citées ont fait, mais qui se sont heurtées alors souvent à de faibles taux de réponse ou de forts

taux de perdus de vue impactant ainsi leurs résultats. Une autre solution serait d'utiliser des bases médico-administratives nationales, qui existent aujourd'hui dans plusieurs pays, dont la France. Cela résoudrait très probablement le problème de puissance, mais à ce jour, ces bases de données n'ayant pas été créées dans un objectif épidémiologique, paraissent difficiles à exploiter car elles n'incluent pas tous les facteurs pouvant influencer la relation entre le DG et le CS. Une dernière solution serait d'envisager le chaînage de différentes bases de données, comme par exemple des bases médico-administratives et des registres de cancer, ce qui nécessiterait des autorisations spécifiques. Cette solution est également en cours de réflexion par l'INCA (appariements de données de registres à la plateforme en oncologie de l'INCA).

### **3.2. Les registres de cancer en France : utilisation de données complémentaires.**

Comme nous l'avions vu précédemment, certaines données directement ou indirectement liées aux cancers sont rarement présentes alors qu'elles sont pertinentes pour les épidémiologistes : habitude de vie, déterminants socio-économiques, caractéristiques génétiques, qualité de vie. Pour pallier à ce manque de variables, qui peuvent parfois être nécessaires pour certaines études car ce sont des facteurs importants dans la relation étudiée, les épidémiologistes vont effectuer leur étude sur des données complémentaires aux registres de cancer.

Ils peuvent notamment demander la participation des patientes, soit de manière longitudinale (cohorte), soit de manière transversale, à l'aide de questionnaires. Par exemple, notre équipe du registre des cancers du sein et des cancers gynécologiques de Côte d'Or a pu le faire, en étudiant la qualité de vie ou les déterminants socio-économiques des survivantes des cancers gynécologiques (cancers du col de l'utérus, de l'endomètre ou de l'ovaire). Une première étude a exploré entre autres la qualité de vie (questionnaire SF12), la fonction sexuelle (questionnaire FSFI), le soutien social

(questionnaire de soutien social de Sarason), l'anxiété et la dépression (échelle HAD) (172). Dans une seconde étude, un questionnaire a été conçu par un groupe de travail pluridisciplinaire composé d'épidémiologistes, de sociologues et de psychologues. Ce questionnaire a permis d'étudier l'impact du cancer sur différents domaines de la vie socio-économique et professionnelle. Il permettait de connaître l'impact du cancer sur le travail, l'accès aux prêts et la réinsertion professionnelle. La privation sociale et matérielle a également été évaluée grâce au questionnaire multidimensionnel EPICES (173). Plusieurs variables issues du registre ont permis d'analyser ces résultats par rapport aux caractéristiques des patientes et de leur cancer : l'âge au moment du diagnostic, l'IMC, l'indice de comorbidité de Charlson, le stade et les traitements de la tumeur.

### **3.3. Exemples de chaînage entre le SNDS et d'autres sources de données en France.**

Le chaînage entre le SNDS et d'autres sources de données présentent plusieurs intérêts que nous allons présenter à travers plusieurs exemples.

Tout d'abord, la cohorte CONSTANCES (174) est « une cohorte épidémiologique généraliste constituée d'un échantillon représentatif à l'inclusion en termes d'âge, de sexe et de statut socio-démographique, des 18-69 ans affiliés à la caisse nationale d'assurance maladie ». A terme, cette cohorte prévoit d'inclure environ 200 000 participants volontaires. Les patients inclus ont été invités à rejoindre le centre de dépistage le plus proche afin de bénéficier d'un examen complet puis un examen de santé tous les 5 ans ; en parallèle, un questionnaire annuel auto-administré est envoyé aux patients. Un suivi passif des patients est permis grâce au couplage annuel avec le SNDS ; les causes de décès sont fournies par le CépiDC. Cette cohorte inclut de nombreuses données cliniques, biologiques, sociodémographiques et comportementales, qui ne sont pas disponibles dans le SNDS ; cela permet d'étudier plus précisément l'effet du comportement personnel, de l'état de santé et de l'environnement tout en bénéficiant de l'aide du SNDS pour limiter au maximum le biais d'attrition.

L'étude CANARI (175) réalisée entre 2010 et 2013 étudie le risque de cancer de la prostate de haut grade chez les patients atteints d'une hyperplasie bénigne de la prostate (HBP) symptomatique. Facteur pronostique majeur du cancer de la prostate, le score de Gleason est fourni après évaluation par un pathologiste d'un échantillon de biopsie de la prostate et n'est pas disponible dans le SNDS. Les données de Gleason issues de tous les laboratoires d'anatomopathologie de Bretagne ont été collectées en vue d'un chaînage au SNDS. Le NIR n'étant pas disponible, ce chaînage a été réalisé grâce à deux « clés » : le mois et l'année de naissance de l'homme ; ainsi que la date de prélèvement de la prostate. La réalisation d'un plan similaire à l'échelle nationale aurait été irréaliste en raison des coûts financiers et humains. Cependant, le chaînage au SNDS leur a permis d'avoir une étude en population avec plus de 850 cas identifiés, au niveau régional, et sans biais d'attrition.

Un dernier exemple concerne la description de la procédure de chaînage du SNDS avec un registre national de maladies rénales terminales REIN (176). Ce registre comprend des données sur l'identification du patient (âge, sexe et code postal du lieu de résidence), les comorbidités (par exemple, maladies cardiovasculaires, diabète, cancer...) et sur la première thérapie de remplacement rénal (par exemple, date, dialyse planifiée ou en urgence...). Les patients sont suivis annuellement, et les événements spécifiques sont enregistrés (transplantation, décès...). Une approche itérative et déterministe a été utilisée pour relier les patients dans le registre REIN à la base de données du SNDS. L'algorithme comprend 24 règles de correspondance, ou étapes, avec des conditions progressivement moins strictes, basé sur l'âge, le sexe, le code postal, le centre de traitement +/- la date du décès +/- la date du traitement. Le chaînage va permettre d'effectuer des études avec une augmentation des informations disponibles sur chaque patient (données cliniques et administratives, mortalité liée à la maladie, utilisation des soins de santé...). La mise en relation des données REIN avec les données nationales du SNDS élargit les perspectives de recherche. Par exemple, il est désormais possible d'évaluer rétrospectivement le lien entre la prescription de médicaments spécifiques et les résultats de la thérapie de remplacement rénal. Une autre étude en cours porte sur le lien entre le parcours de soins avant la thérapie de remplacement rénal (consultations avec un néphrologue et un médecin généraliste), qui peut être extrait de la base de données du SNDS, et le début de la dialyse en urgence, qui est enregistré dans le registre REIN.

A notre connaissance, il n'existe actuellement aucun exemple de chaînage entre le SNDS et un registre de cancer. Cependant, comme nous l'avons brièvement déjà décrit dans la conclusion de notre revue de la littérature, l'Alliance pour la Recherche et l'Innovation des Industries de la Santé (ARIIS) et l'INCA mènent conjointement un projet visant à créer une plateforme de données en oncologie. À terme, elle intégrera des données individuelles issues de bases qualifiées (extraites du SNDS) et des données des registres des cancers. Elle pourra également s'enrichir d'autres sources de données comme les essais cliniques, les tumorothèques, les biothèques ou les bases clinicobiologiques. Cette plateforme serait compatible avec le Health Data Hub (HDH). Le HDH est un groupement d'intérêt public qui associe 56 parties prenantes, en grande majorité issues de la puissance publique. Elles sont réparties en neuf collèges (l'Etat, les caisses d'assurance maladie, les complémentaires, la recherche et l'enseignement, les établissements de santé, les professionnels de santé, les usagers, les agences et les industriels) pour une représentation de l'ensemble de l'écosystème. Le HDH garantit l'accès aisé et unifié, transparent et sécurisé, aux données de santé pour améliorer la qualité des soins et l'accompagnement des patients en France. (177)

## **3.4. Exemples de chaînage entre les bases médico-administratives et les registres de cancer dans le monde**

### **3.3.1 Etats-Unis**

Les données SEER-Medicare sont le résultat du couplage de deux grandes sources de données : le programme Surveillance, Epidemiology and End Results (SEER) constitué de données de registres de cancer, ainsi que des demandes de remboursement de Medicare. Le couplage de ces deux sources de données donne lieu à une source d'information unique en population qui peut être utilisée pour évaluer les services de santé et pour la recherche épidémiologique.

Medicare est un programme d'assurance maladie qui aide les personnes âgées à payer les services de santé aux États-Unis. Il est financé et réglementé par le gouvernement fédéral des États-Unis. En 2020, environ 62 millions de personnes étaient couvertes par Medicare.

Medicare est divisé en plusieurs parties qui couvrent les coûts de différents services :

Partie A : Soins hospitaliers non ambulatoires, soins de fin de vie, établissements de soins infirmiers spécialisés (sous certaines conditions)

Partie B : Services fournis par les médecins et autres prestataires de soins de santé, soins ambulatoires, y compris certains équipements médicaux

Partie C : Medicare Advantage (une gamme d'alternatives au programme original de Medicare qui inclut des programmes de mutuelles, de gestion des organismes prestataires privilégiés et de régimes privés de rémunération à l'acte)

Partie D : Médicaments sur ordonnance

Les personnes ont généralement droit à Medicare si :

- Elles sont âgées d'au moins 65 ans
- Elles sont sous dialyse rénale ou ont reçu une transplantation rénale
- Elles sont âgées de moins de 65 ans et atteintes de certains handicaps
- Elles sont atteintes d'une sclérose latérale amyotrophique (maladie de Charcot)
- Les personnes de moins de 65 ans qui remplissent les critères de l'assurance invalidité de la sécurité sociale depuis au moins 24 mois peuvent être admissibles à Medicare. (178)

Parmi les personnes âgées de 65 ans et plus, 97 % ont droit à Medicare. Presque tous les bénéficiaires de Medicare sont couverts par la partie A. Quatre-vingts seize pour cent des bénéficiaires de la partie A choisissent de payer une prime mensuelle pour s'inscrire à la partie B de Medicare. La plupart ont une couverture de la partie C et environ 70 % ont une couverture de la partie D.

Pour relier le SEER aux données de Medicare, les registres envoient des identifiants individuels pour toutes les personnes dans leurs fichiers. Ces identifiants sont mis en correspondance avec les identifiants contenus dans le fichier d'inscription de Medicare. Le couplage a été réalisé pour la première fois en 1991 et a ensuite été mis à jour, le plus récemment, tous les deux ans. Pour chaque mise à jour du couplage, environ 95 % des personnes âgées de plus de 65 ans dans les fichiers SEER ont été mises en correspondance avec le fichier d'inscription de Medicare.

Un exemple utilisant les données de SEER-Medicare est l'étude de Zhou et al. (179) en 2021 portant sur les facteurs de risques de récurrence(s) d'un cancer du sein précoce RH+/HER2- chez les femmes de plus de 65 ans. Les informations sur l'âge au moment du diagnostic, l'ethnicité, le stade, la taille de la tumeur, l'atteinte des ganglions lymphatiques, la chirurgie, la chimiothérapie, la radiothérapie et l'hormonothérapie ont été identifiées à partir des enregistrements du registre SEER. Les comorbidités préexistantes ont été définies par la présence de codes de diagnostic ou de procédure pertinents dans les demandes de remboursement de Medicare et résumées à l'aide de l'indice de comorbidité de Charlson. L'utilisation d'une hormonothérapie adjuvante a été définie selon les données de la partie D de Medicare concernant la délivrance d'inhibiteurs d'aromatase ou de tamoxifène après le diagnostic de cancer.

L'utilisation des données de SEER-Medicare permet à Zhou et al. d'obtenir une grande cohorte en population de femmes atteintes de cancer du sein suivies sur une période de huit ans et présentant des caractéristiques démographiques et cliniques détaillées. Cependant, les récurrences de cancer du sein ne sont pas systématiquement collectées par les registres SEER ; par conséquent, un algorithme basé sur les demandes de remboursement validé pour être utilisé avec la base de données SEER-Medicare a été utilisé. Toutefois, cette méthode exige une inscription continue à Medicare et au moins un an de survie pour distinguer une récurrence incidente du diagnostic prévalent de cancer du sein. Enfin, plusieurs facteurs n'ont pas pu être déterminés à partir des données du registre ou de Medicare, comme l'obésité ou des biomarqueurs cliniques comme le Ki67.

### 3.3.2 En Ontario

En 2021, Koné et al. (180) ont évalué la prévalence de la multimorbidité et l'utilisation des services de santé en Ontario chez les adultes atteints de cancer. La province de l'Ontario offre une couverture universelle des visites chez le médecin et des services hospitaliers pour pratiquement tous les résidents. Les données proviennent de plusieurs bases médico-administratives concernant les résumés de sortie d'hospitalisation, les soins ambulatoires, les facturations de médecins, les remboursements de médicaments. A l'aide d'algorithmes validés pour la plupart, plusieurs comorbidités ont pu être ciblées : l'infarctus aigu du myocarde, l'asthme, l'arythmie cardiaque, la maladie pulmonaire obstructive chronique, l'insuffisance cardiaque congestive, le syndrome coronarien chronique, la démence, le diabète, l'hypertension, les troubles de l'humeur et de l'anxiété non psychotiques, les autres maladies mentales, l'ostéoarthrite, l'ostéoporose, l'insuffisance rénale, la polyarthrite rhumatoïde et les accidents vasculaires cérébraux. Afin de collecter des informations médicales sur le cancer, ils ont pu chaîner ces bases médico-administratives au registre de cancer d'Ontario afin de caractériser le cancer à l'aide de la 3<sup>ème</sup> édition de la classification internationale des maladies oncologiques.

L'utilisation de sources multiples (médico-administratives et registres de cancer) a permis d'avoir une population d'étude conséquente, un suivi prolongé de cinq à quinze ans, des données incluant de multiples maladies chroniques, y compris des indicateurs de santé mentale. Les analyses ont pu être ajustées en fonction de l'âge et du sexe, mais également du type de cancer et du stade du cancer, grâce au registre, et du nombre de visites chez le médecin un an avant le diagnostic grâce aux données médico-administratives. Malgré tout, ils évoquent quelques limites à leur étude notamment que certaines études de validation ont montré une sous-estimation ou une surestimation de certaines comorbidités chez les patients atteints d'un cancer. Ils manquaient également de variables concernant le statut socio-économique, les approches thérapeutiques et la qualité des soins qui aurait pu impacter leurs résultats.

### 3.3.3 A Taïwan

En 2020, une étude de Lin et al. (34) portant sur l'association entre l'utilisation de faibles doses d'aspirine et l'incidence du cancer colorectal a été réalisée à partir du chaînage des données de l'assurance maladie nationale au registre du cancer de Taïwan. C'est un registre de tous les cas de cancer diagnostiqués dans des hôpitaux comptant au moins 50 lits et contenant des informations sur la date du diagnostic ainsi que sur les caractéristiques histologiques et le stade du cancer. À Taïwan, l'aspirine à faible dose est un médicament délivré sur ordonnance, ses prescriptions sont saisies dans la base de données de l'assurance maladie nationale. Les caractéristiques de la cohorte de l'étude comprenaient l'âge, le sexe, les comorbidités, les prises de médicaments concomitantes, l'utilisation des soins de santé (nombre d'hospitalisations, de consultations externes et d'urgence).

L'utilisation des données de l'assurance maladie nationale a permis d'obtenir un grand échantillon et un suivi long de 2002 à 2015. Cette base de données médico-administrative est également un grand atout pour étudier l'utilisation de l'aspirine qui est systématiquement remboursé par l'assurance maladie à Taïwan. C'est une information que l'on peut donc considérer comme quasi-exhaustive dans la base de données. Le chaînage de ces données avec des données de registre de cancer a permis des analyses spécifiques par stade du cancer, présent dans le registre, et des analyses de sensibilité prenant compte la prise de certains autres médicaments (AINS, inhibiteurs de la pompe à protons, ...) présents dans la base médico-administrative, permettant d'analyser plus finement le rôle propre de l'aspirine. Toutefois, à partir de ces données, il n'a pas été possible de prendre en compte plusieurs facteurs de confusion potentiels comme l'IMC, la consommation de tabac ou d'alcool. Les auteurs évoquent également comme limite de leur étude le fait de ne pas avoir eu accès à certains résultats d'examens biologiques.

### **3.3.4 En Angleterre**

En 2019, Pearson et al. (181) étudient la rapidité de diagnostic au sein de la population pour les patients atteints de cancers colorectaux et du cancer du poumon en utilisant des données de registres de cancer et des données médico-administratives. Ils ont pu identifier dans les bases de données médico-administratives nationales 67 741 cancers colorectaux et 74 904 cancers du poumons grâce à la 10<sup>ème</sup> édition de la classification internationale des maladies. Ils ont alors utilisé le chaînage de ces données avec celle d'un registre de cancer afin de calculer différents délais de surveillance du diagnostic du cancer (entre la première consultation, le premier rendez-vous à l'hôpital, la première imagerie, la première coloscopie ou bronchoscopie). Le registre de cancer leur a fourni également des informations concernant l'ethnicité, l'indice de comorbidité de Charlson, le stade du cancer et un indice de pauvreté.

### **3.3.5 Au Japon**

En 2020, Kanemura et al. (182) ont étudié les traitements des tumeurs épithéliales thymiques au Japon en analysant les données des registres du cancer et des données d'assurance maladie. Cette étude rétrospective s'est basée sur un registre de cancer national dont les données proviennent de 427 structures de soins contre le cancer. Ces données couvrent plus de 67% des patients atteints de cancer et a permis d'inclure plus de 1300 tumeurs épithéliales thymiques entre 2012 et 2014 qui sont des tumeurs rares.

Les informations enregistrées dans le registre de cancer comprennent le sexe, la date de naissance, le site de la tumeur, les codes histologiques selon la troisième édition de la classification des maladies en oncologie, les caractéristiques cliniques et pathologiques des tumeurs, la date du diagnostic et les détails du traitement. Après l'enregistrement dans le registre, celui-ci récupère les données de l'assurance maladie durant 2 ans maximum à partir de la date du diagnostic afin de surveiller la qualité et les modèles de soins des différents types d'hôpitaux (hôpitaux municipaux,

centre de cancer ou hôpitaux universitaires) à l'échelle nationale. Cette base de données de l'assurance maladie comprend des détails concernant les soins fournis à chaque patient, y compris les consultations et les prescriptions de médicaments. Ce chaînage a donc permis aux auteurs de décrire précisément le suivi et les schémas de traitements selon le type d'hôpital durant les deux premières années suivant le diagnostic de tumeurs épithéliales thymiques tout en ajustant sur le stade du cancer.

# Conclusion

En conclusion, l'utilisation de bases de données médico-administratives et de registres de cancers offre de multiples possibilités de travaux épidémiologiques en matière de prévention, de diagnostic et de traitement des cancers. Cependant, les épidémiologistes qui ont l'habitude d'utiliser les données des registres de cancer doivent se rappeler que les données des bases de données administratives qu'ils utilisent n'ont pas été recueillies pour répondre à leur question de recherche spécifique. La première étape de la conception de l'étude consiste donc à déterminer si la base de données que l'on souhaite utiliser peut répondre à cette question et si les données disponibles sont suffisantes pour y répondre.

D'une part, la qualité des informations collectées par les registres de cancer peut varier en fonction de la disponibilité des informations contenues dans les sources de données et du financement accordé pour le recueil et la validation de ces données, qui est tributaire de la volonté politique d'amélioration de la prévention et de la prise en charge des personnes atteintes de cancer de chaque pays. Aujourd'hui la France bénéficie de plusieurs registres de cancer départementaux de bonne qualité.

Les données médico-administratives sont quant à elles dépendantes du système de santé du pays et en notamment des soins remboursés par l'assurance maladie. En comparaison aux Etats-Unis par exemple, la France possède une excellente protection sociale du risque maladie. Le SNDS couvre aujourd'hui 99% de la population française et, grâce au SNIIRAM et au PMSI qui le composent, agrège des données issues des hôpitaux, des soins de ville et de prescriptions médicales.

Les registres de cancer comme les bases de données médico-administratives ont leurs avantages et leurs limites respectifs qui parfois se compensent. Nous avons vu dans cette thèse à partir de plusieurs exemples à travers le monde, que cette complémentarité est un grand atout pour l'épidémiologie des cancers. En effet, les données des registres de cancers concernant les antécédents médicaux, les caractéristiques du cancer ainsi que le suivi de la maladie, permettraient de pallier au

défaut des bases médico-administratives qui se limitent aux actes et soins remboursés par l'assurance maladie. D'autre part, la base médico-administrative nationale du SNDS recouvre quasiment l'ensemble de la population française, notamment pour ce qui concerne les actes et les soins remboursés par l'assurance maladie, ce qui est le cas pour la plupart des soins liés au cancer en France. Cependant, en France, pour bénéficier de la complémentarité de ces deux sources de données, il faudrait mettre en œuvre un chaînage entre les données de registre de cancer et les bases du SNDS. Ceci est déjà prévu dans le cadre du projet d'appariement des données du Registre Général des Cancers de Poitou-Charentes (RGCPC) à la plateforme en oncologie de l'INCa, ce qui permettrait d'obtenir davantage d'informations concernant le parcours de soins des patients atteints de cancer. Ce projet de création d'une plateforme de données en oncologie alliant les données de registres et les bases de données médico-administratives s'inscrit dans les projets du Health Data Hub. Il constituera un outil précieux dans l'évaluation des besoins en matière de prévention, de diagnostic et de traitement des cancers.

D'importants progrès ont été effectués ces dernières décennies pour mieux comprendre la physiopathologie des cancers et mieux les traiter, ce qui a contribué à l'amélioration de la survie des patients. Toutefois, le bien-être psychologique, la qualité de vie, les habitudes de vie (régime alimentaire, activités sportives, consommation d'alcool, de tabac ou d'autres drogues, ...) et les données socio-économiques (l'origine ethnique, le niveau d'éducation ou le revenu par exemple) sont très peu renseignées dans ces bases. Il serait intéressant de réfléchir dès maintenant à des solutions pour que ces données puissent également être collectées en routine afin qu'elles puissent être utilisées dans les travaux épidémiologiques.

# Références

1. Définition cancer [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.e-cancer.fr/Dictionnaire/C/cancer>
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6):394-424.
3. Cancers [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers>
4. Source : Réseau FRANCIM [Internet]. [cité 30 juill 2021]. Disponible sur: <https://lesdonnees.e-cancer.fr/Informations/Sources/SOURCE-Reseau-FRANCIM>
5. Délibération n° 2016-316 du 13 octobre 2016 portant avis sur un projet de décret en Conseil d'Etat relatif au Système national des données de santé (demande d'avis n° 16018114).
6. Cancer du sein [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein>
7. Stebbing J, Delaney G, Thompson A. Breast cancer (non-metastatic). *BMJ Clin Evid*. 4 déc 2007;2007:0102.
8. Stebbing J, Ngan S. Breast cancer (metastatic). *BMJ Clin Evid*. 8 sept 2010;2010:0811.
9. Dawood S, Broglio K, Gonzalez-Angulo AM, Buzdar AU, Hortobagyi GN, Giordano SH. Trends in survival over the past two decades among white and black patients with newly diagnosed stage IV breast cancer. *J Clin Oncol*. 20 oct 2008;26(30):4891-8.
10. Chia SK, Speers CH, D'yachkova Y, Kang A, Malfair-Taylor S, Barnett J, et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. *Cancer*. 1 sept 2007;110(5):973-9.
11. Wilcken N, Hornbuckle J, Ghersi D. Chemotherapy alone versus endocrine therapy alone for metastatic breast cancer. *Cochrane Database Syst Rev*. 2003;(2):CD002747.
12. Trédan O, Follana P, Moullet I, Cropet C, Trager-Maury S, Dauba J, et al. A phase III trial of exemestane plus bevacizumab maintenance therapy in patients with metastatic breast cancer after first-line taxane and bevacizumab: a GINECO group study. *Ann Oncol*. juin 2016;27(6):1020-9.
13. Finn RS, Martin M, Rugo HS, Jones S, Im S-A, Gelmon K, et al. Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med*. 17 nov 2016;375(20):1925-36.

14. Robertson JFR, Cheung K-L, Noguchi S, Shao Z, Degboe A, Lichfield J, et al. Health-related quality of life from the FALCON phase III randomised trial of fulvestrant 500 mg versus anastrozole for hormone receptor-positive advanced breast cancer. *Eur J Cancer*. mai 2018;94:206-15.
15. Royce M, Bachelot T, Villanueva C, Özgüroglu M, Azevedo SJ, Cruz FM, et al. Everolimus Plus Endocrine Therapy for Postmenopausal Women With Estrogen Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Advanced Breast Cancer: A Clinical Trial. *JAMA Oncol*. 1 juill 2018;4(7):977-84.
16. Cardoso F, Senkus E, Costa A, Papadopoulos E, Aapro M, André F, et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4)†. *Ann Oncol*. 1 août 2018;29(8):1634-57.
17. Gradishar WJ, Anderson BO, Blair SL, Burstein HJ, Cyr A, Elias AD, et al. Breast cancer version 3.2014. *J Natl Compr Canc Netw*. avr 2014;12(4):542-90.
18. Partridge AH, Rumble RB, Carey LA, Come SE, Davidson NE, Di Leo A, et al. Chemotherapy and targeted therapy for women with human epidermal growth factor receptor 2-negative (or unknown) advanced breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 10 oct 2014;32(29):3307-29.
19. André F, Neven P, Marinsek N, Zhang J, Baladi J-F, Degun R, et al. Disease management patterns for postmenopausal women in Europe with hormone-receptor-positive, human epidermal growth factor receptor-2 negative advanced breast cancer. *Curr Med Res Opin*. juin 2014;30(6):1007-16.
20. Swallow E, Zhang J, Thomason D, Tan R-D, Kageleiry A, Signorovitch J. Real-world patterns of endocrine therapy for metastatic hormone-receptor-positive (HR+)/human epidermal growth factor receptor-2-negative (HER2-) breast cancer patients in the United States: 2002-2012. *Curr Med Res Opin*. août 2014;30(8):1537-45.
21. Jerusalem G, Neven P, Marinsek N, Zhang J, Degun R, Benelli G, et al. Patterns of resource utilization and cost for postmenopausal women with hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer in Europe. *BMC Cancer*. 24 oct 2015;15:787.
22. Lobbezoo DJA, van Kampen RJW, Voogd AC, Dercksen MW, van den Berkmortel F, Smilde TJ, et al. In real life, one-quarter of patients with hormone receptor-positive metastatic breast cancer receive chemotherapy as initial palliative therapy: a study of the Southeast Netherlands Breast Cancer Consortium. *Ann Oncol*. févr 2016;27(2):256-62.
23. Jacquet E, Lardy-Cléaud A, Pistilli B, Franck S, Cottu P, Delalogue S, et al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2-negative metastatic breast cancer patients. *Eur J Cancer*. mai 2018;95:93-101.

24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40(5):373-83.
25. Gourgou-Bourgade S, Cameron D, Poortmans P, Asselain B, Azria D, Cardoso F, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials)†. *Ann Oncol.* mai 2015;26(5):873-9.
26. ROSENBAUM PR, RUBIN DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1 avr 1983;70(1):41-55.
27. Hortobagyi GN, Stemmer SM, Burris HA, Yap Y-S, Sonke GS, Paluch-Shimon S, et al. Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. *N Engl J Med.* 3 nov 2016;375(18):1738-48.
28. Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, Perez EA, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med.* 27 déc 2007;357(26):2666-76.
29. Pritchard KI, Sousa B. ABC 1 (1st International consensus guidelines for advanced breast cancer): a positive step. *Breast.* juin 2012;21(3):225-6.
30. Co M, Ngan RKC, Mang OWK, Tam AHP, Wong KH, Kwong A. Long-term Survival Outcomes of ‘Low Risk’ Ductal Carcinoma in situ from a Territory-wide Cancer Registry. *Clinical Oncology.* 1 janv 2021;33(1):40-5.
31. Morais S, Antunes L, Bento MJ, Lunet N. Risk of second primary cancers among patients with a first primary gastric cancer: A population-based study in North Portugal. *Cancer Epidemiology.* 1 oct 2017;50:85-91.
32. Kinoshita FL, Ito Y, Morishima T, Miyashiro I, Nakayama T. Sex differences in lung cancer survival: long-term trends using population-based cancer registry data in Osaka, Japan. *Japanese Journal of Clinical Oncology.* 1 sept 2017;47(9):863-9.
33. Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource. *Cancer Epidemiol Biomarkers Prev.* déc 1999;8(12):1117-21.
34. Lin H-MD, Vora P, Soriano-Gabarró M, Chan KA. Association Between Low-Dose Aspirin Use and Colorectal Cancer Incidence in Taiwan. *JAMA Netw Open.* 2 nov 2020;3(11):e2026494.
35. Rovers KP, Bakkers C, van Erning FN, Burger JWA, Nienhuijs SW, Simkens GAAM, et al. Adjuvant Systemic Chemotherapy vs Active Surveillance Following Up-front Resection of Isolated Synchronous Colorectal Peritoneal Metastases. *JAMA Oncol.* août 2020;6(8):e202701.

36. SNDS : Système National des Données de Santé | CNIL [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.cnil.fr/fr/snds-systeme-national-des-donnees-de-sante>
37. Tuppin P, Rudant J, Constantinou P, Gastaldi-Ménager C, Rachas A, de Roquefeuil L, et al. Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S149-67.
38. Composantes du SNDS | SNDS [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.snds.gouv.fr/SNDS/Composantes-du-SNDS>
39. Cancer today [Internet]. [cité 30 juill 2021]. Disponible sur: <http://geo.iarc.fr/today/home>
40. SPF. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Tumeurs solides : Étude à partir des registres des cancers du réseau Francim [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.santepubliquefrance.fr/import/estimations-nationales-de-l-incidence-et-de-la-mortalite-par-cancer-en-france-metropolitaine-entre-1990-et-2018-tumeurs-solides-etude-a-partir>
41. Drouillard A, Manfredi S, Lepage C, Bouvier A-M. [Epidemiology of pancreatic cancer]. *Bull Cancer*. janv 2018;105(1):63-9.
42. Cancer of the Pancreas - Cancer Stat Facts [Internet]. SEER. [cité 4 juill 2020]. Disponible sur: <https://seer.cancer.gov/statfacts/html/pancreas.html>
43. Rawla P, Sunkara T, Gaduputi V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J Oncol*. févr 2019;10(1):10-27.
44. Huxley R, Ansary-Moghaddam A, Berrington de González A, Barzi F, Woodward M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br J Cancer*. 6 juin 2005;92(11):2076-83.
45. Elena JW, Steplowski E, Yu K, Hartge P, Tobias GS, Brotzman MJ, et al. Diabetes and risk of pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium. *Cancer Causes Control*. janv 2013;24(1):13-25.
46. Chodick G, Heymann AD, Rosenmann L, Green MS, Flash S, Porath A, et al. Diabetes and risk of incident cancer: a large population-based cohort study in Israel. *Cancer Causes Control*. juin 2010;21(6):879-87.
47. Everhart J, Wright D. Diabetes mellitus as a risk factor for pancreatic cancer. A meta-analysis. *JAMA*. 24 mai 1995;273(20):1605-9.
48. Zhou XH, Qiao Q, Zethelius B, Pyörälä K, Söderberg S, Pajak A, et al. Diabetes, prediabetes

and cancer mortality. *Diabetologia*. sept 2010;53(9):1867-76.

49. Biadgo B, Abebe M. Type 2 Diabetes Mellitus and Its Association with the Risk of Pancreatic Carcinogenesis: A Review. *Korean J Gastroenterol*. 25 avr 2016;67(4):168-77.
50. Rahn S, Zimmermann V, Viol F, Knaack H, Stemmer K, Peters L, et al. Diabetes as risk factor for pancreatic cancer: Hyperglycemia promotes epithelial-mesenchymal-transition and stem cell properties in pancreatic ductal epithelial cells. *Cancer Lett*. 28 févr 2018;415:129-50.
51. Eibl G, Cruz-Monserrate Z, Korc M, Petrov MS, Goodarzi MO, Fisher WE, et al. Diabetes Mellitus and Obesity as Risk Factors for Pancreatic Cancer. *J Acad Nutr Diet*. avr 2018;118(4):555-67.
52. Paternoster S, Falasca M. The intricate relationship between diabetes, obesity and pancreatic cancer. *Biochim Biophys Acta Rev Cancer*. janv 2020;1873(1):188326.
53. Le diabète gestationnel | Fédération Française des Diabétiques [Internet]. [cité 30 juill 2021]. Disponible sur: <https://www.federationdesdiabetiques.org/information/diabete-gestationnel>
54. Bellamy L, Casas J-P, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet*. 23 mai 2009;373(9677):1773-9.
55. Sullivan SD, Umans JG, Ratner R. Gestational diabetes: implications for cardiovascular health. *Curr Diab Rep*. févr 2012;12(1):43-52.
56. Kessous R, Shoham-Vardi I, Pariente G, Sherf M, Sheiner E. An association between gestational diabetes mellitus and long-term maternal cardiovascular morbidity. *Heart*. août 2013;99(15):1118-21.
57. Goueslard K, Cottenet J, Mariet A-S, Giroud M, Cottin Y, Petit J-M, et al. Early cardiovascular events in women with a history of gestational diabetes mellitus. *Cardiovasc Diabetol*. 27 janv 2016;15:15.
58. Carr DB, Utzschneider KM, Hull RL, Tong J, Wallace TM, Kodama K, et al. Gestational diabetes mellitus increases the risk of cardiovascular disease in women with a family history of type 2 diabetes. *Diabetes Care*. sept 2006;29(9):2078-83.
59. Shostrom DCV, Sun Y, Oleson JJ, Snetselaar LG, Bao W. History of Gestational Diabetes Mellitus in Relation to Cardiovascular Disease and Cardiovascular Risk Factors in US Women. *Front Endocrinol (Lausanne)*. 2017;8:144.
60. Kaul P, Savu A, Nerenberg KA, Donovan LE, Chik CL, Ryan EA, et al. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: a population-level analysis. *Diabet Med*. févr 2015;32(2):164-73.

61. Shah BR, Retnakaran R, Booth GL. Increased risk of cardiovascular disease in young women following gestational diabetes mellitus. *Diabetes Care*. août 2008;31(8):1668-9.
62. Retnakaran R, Shah BR. Mild glucose intolerance in pregnancy and risk of cardiovascular disease: a population-based cohort study. *CMAJ*. 15 sept 2009;181(6-7):371-6.
63. Retnakaran R, Shah BR. Role of Type 2 Diabetes in Determining Retinal, Renal, and Cardiovascular Outcomes in Women With Previous Gestational Diabetes Mellitus. *Diabetes Care*. janv 2017;40(1):101-8.
64. Dawson SI. Long-term risk of malignant neoplasm associated with gestational glucose intolerance. *Cancer*. 1 janv 2004;100(1):149-55.
65. Perrin MC, Terry MB, Kleinhaus K, Deutsch L, Yanetz R, Tiram E, et al. Gestational diabetes and the risk of breast cancer among women in the Jerusalem Perinatal Study. *Breast Cancer Res Treat*. mars 2008;108(1):129-35.
66. Bejaimal S a. D, Wu CF, Lowe J, Feig DS, Shah BR, Lipscombe LL. Short-term risk of cancer among women with previous gestational diabetes: a population-based study. *Diabet Med*. janv 2016;33(1):39-46.
67. Park Y-MM, O'Brien KM, Zhao S, Weinberg CR, Baird DD, Sandler DP. Gestational diabetes mellitus may be associated with increased risk of breast cancer. *Br J Cancer*. 28 mars 2017;116(7):960-3.
68. Wartko PD, Beck TL, Reed SD, Mueller BA, Hawes SE. Association of endometrial hyperplasia and cancer with a history of gestational diabetes. *Cancer Causes Control*. août 2017;28(8):819-28.
69. Fuchs O, Sheiner E, Meirovitz M, Davidson E, Sergienko R, Kessous R. The association between a history of gestational diabetes mellitus and future risk for female malignancies. *Arch Gynecol Obstet*. mars 2017;295(3):731-6.
70. Brasky TM, Li Y, Jaworowicz DJ, Potischman N, Ambrosone CB, Hutson AD, et al. Pregnancy-related characteristics and breast cancer risk. *Cancer Causes Control*. sept 2013;24(9):1675-85.
71. Perrin MC, Terry MB, Kleinhaus K, Deutsch L, Yanetz R, Tiram E, et al. Gestational diabetes as a risk factor for pancreatic cancer: a prospective cohort study. *BMC Med*. 16 août 2007;5:25.
72. Sella T, Chodick G, Barchana M, Heymann AD, Porath A, Kokia E, et al. Gestational diabetes and risk of incident primary cancer: a large historical cohort study in Israel. *Cancer Causes Control*. nov 2011;22(11):1513-20.
73. Peng Y-S, Lin J-R, Cheng B-H, Ho C, Lin Y-H, Shen C-H, et al. Incidence and relative risk

for developing cancers in women with gestational diabetes mellitus: a nationwide cohort study in Taiwan. *BMJ Open*. 21 févr 2019;9(2):e024583.

74. Wang Y, Yan P, Fu T, Yuan J, Yang G, Liu Y, et al. The association between gestational diabetes mellitus and cancer in women: A systematic review and meta-analysis of observational studies. *Diabetes Metab*. nov 2020;46(6):461-71.
75. Fosse-Edorh S, Rigou A, Morin S, Fezeu L, Mandereau-Bruno L, Fagot-Campagna A. [Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes]. *Rev Epidemiol Sante Publique*. oct 2017;65 Suppl 4:S168-73.
76. Pannala R, Leirness JB, Bamlet WR, Basu A, Petersen GM, Chari ST. Prevalence and clinical profile of pancreatic cancer-associated diabetes mellitus. *Gastroenterology*. avr 2008;134(4):981-7.
77. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med*. 15 août 2017;167(4):268-74.
78. Torloni MR, Betrán AP, Horta BL, Nakamura MU, Atallah AN, Moron AF, et al. Prepregnancy BMI and the risk of gestational diabetes: a systematic review of the literature with meta-analysis. *Obes Rev*. mars 2009;10(2):194-203.
79. Johansen D, Stocks T, Jonsson H, Lindkvist B, Björge T, Concini H, et al. Metabolic factors and the risk of pancreatic cancer: a prospective analysis of almost 580,000 men and women in the Metabolic Syndrome and Cancer Project. *Cancer Epidemiol Biomarkers Prev*. sept 2010;19(9):2307-17.
80. Quantin C, Cottenet J, Vuagnat A, Prunet C, Mouquet M-C, Fresson J, et al. [Quality of perinatal statistics from hospital discharge data: comparison with civil registration and the 2010 National Perinatal Survey]. *J Gynecol Obstet Biol Reprod (Paris)*. nov 2014;43(9):680-90.
81. Blondel B, Drewniak N, Pilkington H, Zeitlin J. Out-of-hospital births and the supply of maternity units in France. *Health Place*. sept 2011;17(5):1170-3.
82. Proceedings of the Second International Workshop-Conference on Gestational Diabetes Mellitus. October 25-27, 1984, Chicago, Illinois. *Diabetes*. juin 1985;34 Suppl 2:1-130.
83. Watson WJ. Screening for glycosuria during pregnancy. *South Med J*. févr 1990;83(2):156-8.
84. Buhling KJ, Elze L, Henrich W, Starr E, Stein U, Siebert G, et al. The usefulness of glycosuria and the influence of maternal blood pressure in screening for gestational diabetes. *Eur J Obstet Gynecol Reprod Biol*. 15 avr 2004;113(2):145-8.
85. Agbozo F, Abubakari A, Narh C, Jahn A. Accuracy of glycosuria, random blood glucose and

risk factors as selective screening tools for gestational diabetes mellitus in comparison with universal diagnosing. *BMJ Open Diabetes Res Care*. 2018;6(1):e000493.

86. Ma RCW, Chan JCN. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. *Ann N Y Acad Sci*. avr 2013;1281:64-91.
87. Noto H, Tsujimoto T, Noda M. Significantly increased risk of cancer in diabetes mellitus patients: A meta-analysis of epidemiological evidence in Asians and non-Asians. *J Diabetes Investig*. 20 févr 2012;3(1):24-33.
88. Pergolini I, Schorn S, Jäger C, Göß R, Novotny A, Friess H, et al. Diabetes mellitus in intraductal papillary mucinous neoplasms: A systematic review and meta-analysis. *Surgery*. févr 2021;169(2):411-8.
89. Fuentes S, Cosson E, Mandereau-Bruno L, Fagot-Campagna A, Bernillon P, Goldberg M, et al. Identifying diabetes cases in health administrative databases: a validation study based on a large French cohort. *Int J Public Health*. avr 2019;64(3):441-50.
90. Bousquet P-J, Caillet P, Coeuret-Pellicer M, Goulard H, Kudjawu YC, Le Bihan C, 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*. oct 2017;65 Suppl 4:S236-42.
91. Wu JW, Azoulay L, Huang A, Paterson M, Wu F, Secrest MH, et al. Identification of incident pancreatic cancer in Ontario administrative health data: A validation study. *Pharmacoepidemiol Drug Saf*. janv 2020;29 Suppl 1:78-85.
92. Pierron A, Revert M, Goueslard K, Vuagnat A, Cottenet J, Benzenine E, et al. [Evaluation of the metrological quality of the medico-administrative data for perinatal indicators: A pilot study in 3 university hospitals]. *Rev Epidemiol Sante Publique*. août 2015;63(4):237-46.
93. Goueslard K, Cottenet J, Benzenine E, Tubert-Bitter P, Quantin C. Validation study: evaluation of the metrological quality of French hospital data for perinatal algorithms. *BMJ Open*. 12 mai 2020;10(5):e035218.
94. Pillay J, Donovan L, Guitard S, Zakher B, Gates M, Gates A, et al. Screening for Gestational Diabetes: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA*. 10 août 2021;326(6):539-62.
95. Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology*. juin 2013;144(6):1252-61.
96. Goueslard K, Cottenet J, Mariet A-S, Sagot P, Petit J-M, Quantin C. Early screening for type 2 diabetes following gestational diabetes mellitus in France: hardly any impact of the 2010 guidelines. *Acta Diabetol*. juill 2017;54(7):645-51.

97. Quaresima P, Visconti F, Chiefari E, Puccio L, Foti DP, Venturella R, et al. Barriers to Postpartum Glucose Intolerance Screening in an Italian Population. *Int J Environ Res Public Health*. 14 déc 2018;15(12):E2853.
98. Boyle DIR, Versace VL, Dunbar JA, Scheil W, Janus E, Oats JJN, et al. Results of the first recorded evaluation of a national gestational diabetes mellitus register: Challenges in screening, registration, and follow-up for diabetes risk. *PLoS One*. 2018;13(8):e0200832.
99. Bernstein J, Quinn E, Ameli O, Craig M, Heeren T, Iverson R, et al. Onset of T2DM after gestational diabetes: What the prevention paradox tells us about risk. *Prev Med*. août 2018;113:1-6.
100. Collège national des gynécologues et obstétriciens français, Société francophone du diabète. [Gestational diabetes]. *J Gynecol Obstet Biol Reprod (Paris)*. déc 2010;39(8 Suppl 2):S139, S338-342.
101. American Diabetes Association. 14. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes-2020. *Diabetes Care*. janv 2020;43(Suppl 1):S183-92.
102. Capasso M, Franceschi M, Rodriguez-Castro KI, Crafa P, Cambiè G, Miraglia C, et al. Epidemiology and risk factors of pancreatic cancer. *Acta Biomed*. 17 déc 2018;89(9-S):141-6.
103. Midha S, Chawla S, Garg PK. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. *Cancer Lett*. 10 oct 2016;381(1):269-77.
104. Processus d'accès aux données | SNDS [Internet]. [cité 1 nov 2021]. Disponible sur: <https://www.snds.gouv.fr/SNDS/Processus-d-acces-aux-donnees>
105. Guide accès permanents | Documentation du SNDS [Internet]. [cité 13 déc 2021]. Disponible sur: [https://documentation-snds.health-data-hub.fr/formation\\_snds/documents\\_cnam/guides\\_pedagogiques\\_snds/guide\\_pedagogique\\_acces\\_permanents.html#qu-est-ce-que-le-systeme-national-des-donnees-de-sante-snds](https://documentation-snds.health-data-hub.fr/formation_snds/documents_cnam/guides_pedagogiques_snds/guide_pedagogique_acces_permanents.html#qu-est-ce-que-le-systeme-national-des-donnees-de-sante-snds)
106. Touzeau C, Quignot N, Meng J, Jiang H, Khachatryan A, Singh M, et al. Survival and treatment patterns of patients with relapsed or refractory multiple myeloma in France — a cohort study using the French National Healthcare database (SNDS). *Ann Hematol*. 2021;100(7):1825-36.
107. Champeaux-Depond C, Weller J. Tamoxifen. A treatment for meningioma? *Cancer Treatment and Research Communications*. 1 janv 2021;27:100343.
108. Assié J-B, Corre R, Levra MG, Calvet CY, Gaudin A-F, Grumberg V, et al. Nivolumab treatment in advanced non-small cell lung cancer: real-world long-term outcomes within overall and special populations (the UNIVOC study). *Ther Adv Med Oncol*. 2020;12:1758835920967237.

109. Scailteux L-M, Campillo-Gimenez B, Kerbrat S, Despas F, Mathieu R, Vincendeau S, et al. Overall Survival Among Chemotherapy-Naive Patients With Castration-Resistant Prostate Cancer Under Abiraterone Versus Enzalutamide: A Direct Comparison Based on a 2014-2018 French Population Study (the SPEAR Cohort). *Am J Epidemiol*. 1 févr 2021;190(3):413-22.
110. Tuppin P, Paita M, Gastaldi-Menager C, Le Bihan-Benjamin C, Jean Bousquet P, Fagot Campagna A. Patient characteristics and treatments one year after diagnosis of renal cancer in 2015 according to the presence or absence of metastasis and a history of another tumour: A national observational study. *Bulletin du Cancer*. 1 juin 2019;106(6):538-49.
111. Thurin NH, Rouyer M, Gross-Goupil M, Rebillard X, Soulié M, Haaser T, et al. Epidemiology of metastatic castration-resistant prostate cancer: A first estimate of incidence and prevalence using the French nationwide healthcare database. *Cancer Epidemiology*. 1 déc 2020;69:101833.
112. Goulabchand R, Malafaye N, Jacot W, Witkowski Durand Viel P, Morel J, Lukas C, et al. Cancer incidence in primary Sjögren's syndrome: Data from the French hospitalization database. *Autoimmunity Reviews*. 28 oct 2021;102987.
113. Escribà JM, Pareja L, Esteban L, Gálvez J, Melià A, Roca L, et al. Trends in the surgical procedures of women with incident breast cancer in Catalonia, Spain, over a 7-year period (2005-2011). *BMC Res Notes*. 1 sept 2014;7:587.
114. Bögemann M, Zagorska A, Akumo D, Hadad LE, Pignot M. Using Data from a Sickness Fund Claims Database to Assess the Treatment Patterns and Healthcare Resource Utilization among Patients with Metastatic Renal Cell Carcinoma in Germany. *Urol Int*. nov 2020;104(11-12):982-93.
115. HEO J, CHUN M, OH Y-T, KYU NOH O, KIM L. Influenza Among Breast Cancer Survivors in South Korea: A Nationwide Population-Based Study. *In Vivo*. 3 sept 2017;31(5):967-72.
116. Chen H-F, Liu M-D, Chen P, Chen L-H, Chang Y-H, Wen P-C, et al. Risks of Breast and Endometrial Cancer in Women with Diabetes: A Population-Based Cohort Study. *PLoS ONE*. 2013;8(6):e67420.
117. Michels KB, Solomon CG, Hu FB, Rosner BA, Hankinson SE, Colditz GA, et al. Type 2 diabetes and subsequent incidence of breast cancer in the Nurses' Health Study. *Diabetes Care*. juin 2003;26(6):1752-8.
118. Xue F, Michels KB. Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence. *Am J Clin Nutr*. sept 2007;86(3):s823-835.
119. Larsson SC, Mantzoros CS, Wolk A. Diabetes mellitus and risk of breast cancer: a meta-analysis. *Int J Cancer*. 15 août 2007;121(4):856-62.

120. Liao S, Li J, Wei W, Wang L, Zhang Y, Li J, et al. Association between diabetes mellitus and breast cancer risk: a meta-analysis of the literature. *Asian Pac J Cancer Prev.* 2011;12(4):1061-5.
121. Boyle P, Boniol M, Koechlin A, Robertson C, Valentini F, Coppens K, et al. Diabetes and breast cancer risk: a meta-analysis. *Br J Cancer.* 23 oct 2012;107(9):1608-17.
122. Starup-Linde J, Karlstad O, Eriksen SA, Vestergaard P, Bronsveld HK, de Vries F, et al. CARING (CAncer Risk and INSulin analogues): the association of diabetes mellitus and cancer risk with focus on possible determinants - a systematic review and a meta-analysis. *Curr Drug Saf.* nov 2013;8(5):296-332.
123. Dankner R, Boffetta P, Balicer RD, Boker LK, Sadeh M, Berlin A, et al. Time-Dependent Risk of Cancer After a Diabetes Diagnosis in a Cohort of 2.3 Million Adults. *Am J Epidemiol.* 15 juin 2016;183(12):1098-106.
124. Nyholm H, Djursing H, Hagen C, Agner T, Bennett P, Svenstrup B. Androgens and estrogens in postmenopausal insulin-treated diabetic women. *J Clin Endocrinol Metab.* nov 1989;69(5):946-9.
125. Lipworth L, Adami HO, Trichopoulos D, Carlström K, Mantzoros C. Serum steroid hormone levels, sex hormone-binding globulin, and body mass index in the etiology of postmenopausal breast cancer. *Epidemiology.* janv 1996;7(1):96-100.
126. Lawlor DA, Smith GD, Ebrahim S. Hyperinsulinaemia and increased risk of breast cancer: findings from the British Women's Heart and Health Study. *Cancer Causes Control.* avr 2004;15(3):267-75.
127. Chaudhuri PK, Chaudhuri B, Patel N. Modulation of estrogen receptor by insulin and its biologic significance. *Arch Surg.* nov 1986;121(11):1322-5.
128. Ferroni P, Riondino S, Buonomo O, Palmirotta R, Guadagni F, Roselli M. Type 2 Diabetes and Breast Cancer: The Interplay between Impaired Glucose Metabolism and Oxidant Stress. *Oxid Med Cell Longev.* 2015;2015:183928.
129. Bellamy L, Casas J-P, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet.* 23 mai 2009;373(9677):1773-9.
130. Kaul P, Savu A, Nerenberg KA, Donovan LE, Chik CL, Ryan EA, et al. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: a population-level analysis. *Diabet Med.* févr 2015;32(2):164-73.
131. Retnakaran R, Shah BR. Role of Type 2 Diabetes in Determining Retinal, Renal, and Cardiovascular Outcomes in Women With Previous Gestational Diabetes Mellitus. *Diabetes Care.* janv 2017;40(1):101-8.

132. Wartko PD, Beck TL, Reed SD, Mueller BA, Hawes SE. Association of endometrial hyperplasia and cancer with a history of gestational diabetes. *Cancer Causes Control*. août 2017;28(8):819-28.
133. Perrin MC, Terry MB, Kleinhaus K, Deutsch L, Yanetz R, Tiram E, et al. Gestational diabetes as a risk factor for pancreatic cancer: a prospective cohort study. *BMC Med*. 16 août 2007;5:25.
134. Sella T, Chodick G, Barchana M, Heymann AD, Porath A, Kokia E, et al. Gestational diabetes and risk of incident primary cancer: a large historical cohort study in Israel. *Cancer Causes Control*. nov 2011;22(11):1513-20.
135. Pace R, Rahme E, Dasgupta K. Gestational diabetes mellitus and risk of incident primary cancer: A population-based retrospective cohort study. *J Diabetes*. janv 2020;12(1):87-90.
136. Han K-T, Cho GJ, Kim EH. Evaluation of the Association between Gestational Diabetes Mellitus at First Pregnancy and Cancer within 10 Years Postpartum Using National Health Insurance Data in South Korea. *Int J Environ Res Public Health*. 26 nov 2018;15(12):E2646.
137. Fuchs O, Sheiner E, Meirovitz M, Davidson E, Sergienko R, Kessous R. The association between a history of gestational diabetes mellitus and future risk for female malignancies. *Arch Gynecol Obstet*. mars 2017;295(3):731-6.
138. Peng Y-S, Lin J-R, Cheng B-H, Ho C, Lin Y-H, Shen C-H, et al. Incidence and relative risk for developing cancers in women with gestational diabetes mellitus: a nationwide cohort study in Taiwan. *BMJ Open*. 21 févr 2019;9(2):e024583.
139. Bejaimal S a. D, Wu CF, Lowe J, Feig DS, Shah BR, Lipscombe LL. Short-term risk of cancer among women with previous gestational diabetes: a population-based study. *Diabet Med*. janv 2016;33(1):39-46.
140. Simon J, Goueslard K, Arveux P, Bechraoui-Quantin S, Petit J-M, Quantin C. Increased Risk of Hospitalization for Pancreatic Cancer in the First 8 Years after a Gestational Diabetes Mellitus regardless of Subsequent Type 2 Diabetes: A Nationwide Population-Based Study. *Cancers (Basel)*. 15 janv 2021;13(2):308.
141. Russo IH, Russo J. Pregnancy-induced changes in breast cancer risk. *J Mammary Gland Biol Neoplasia*. sept 2011;16(3):221-33.
142. Meier-Abt F, Bentires-Alj M. How pregnancy at early age protects against breast cancer. *Trends Mol Med*. mars 2014;20(3):143-53.
143. Schedin P. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer*. avr 2006;6(4):281-91.
144. Wang Y, Yan P, Fu T, Yuan J, Yang G, Liu Y, et al. The association between gestational

diabetes mellitus and cancer in women: A systematic review and meta-analysis of observational studies. *Diabetes Metab.* nov 2020;46(6):461-71.

145. Xie C, Wang W, Li X, Shao N, Li W. Gestational diabetes mellitus and maternal breast cancer risk: a meta-analysis of the literature. *J Matern Fetal Neonatal Med.* mars 2019;32(6):1022-32.
146. Perrin MC, Terry MB, Kleinhaus K, Deutsch L, Yanetz R, Tiram E, et al. Gestational diabetes and the risk of breast cancer among women in the Jerusalem Perinatal Study. *Breast Cancer Res Treat.* mars 2008;108(1):129-35.
147. Powe CE, Tobias DK, Michels KB, Chen WY, Eliassen AH, Manson JE, et al. History of Gestational Diabetes Mellitus and Risk of Incident Invasive Breast Cancer among Parous Women in the Nurses' Health Study II Prospective Cohort. *Cancer Epidemiol Biomarkers Prev.* mars 2017;26(3):321-7.
148. Park Y-MM, O'Brien KM, Zhao S, Weinberg CR, Baird DD, Sandler DP. Gestational diabetes mellitus may be associated with increased risk of breast cancer. *Br J Cancer.* 28 mars 2017;116(7):960-3.
149. Bertrand KA, Castro-Webb N, Cozier YC, Li S, O'Brien KM, Rosenberg L, et al. Gestational Diabetes and Risk of Breast Cancer in African American Women. *Cancer Epidemiol Biomarkers Prev.* juill 2020;29(7):1509-11.
150. Troisi R, Weiss HA, Hoover RN, Potischman N, Swanson CA, Brogan DR, et al. Pregnancy characteristics and maternal risk of breast cancer. *Epidemiology.* nov 1998;9(6):641-7.
151. Rollison DE, Giuliano AR, Sellers TA, Laronga C, Sweeney C, Risendal B, et al. Population-based case-control study of diabetes and breast cancer risk in Hispanic and non-Hispanic White women living in US southwestern states. *Am J Epidemiol.* 15 févr 2008;167(4):447-56.
152. Brasky TM, Li Y, Jaworowicz DJ, Potischman N, Ambrosone CB, Hutson AD, et al. Pregnancy-related characteristics and breast cancer risk. *Cancer Causes Control.* sept 2013;24(9):1675-85.
153. Ardalan A, Bungum T. Gestational Age and the Risk of Maternal Breast Cancer: A Population-Based Case-Control Study. *Breast J.* nov 2016;22(6):657-61.
154. Karagiannis T, Bekiari E, Manolopoulos K, Paletas K, Tsapas A. Gestational diabetes mellitus: why screen and how to diagnose. *Hippokratia.* juill 2010;14(3):151-4.
155. Cordina-Duverger E, Menegaux F, Popa A, Rabstein S, Harth V, Pesch B, et al. Night shift work and breast cancer: a pooled analysis of population-based case-control studies with complete work history. *Eur J Epidemiol.* avr 2018;33(4):369-79.

156. Tankó LB, Bagger YZ, Christiansen C. Long-term risk of malignant neoplasm associated with gestational glucose intolerance. *Cancer*. 15 juin 2004;100(12):2680-1; author reply 2681.
157. Chodick G, Zucker I. Diabetes, gestational diabetes and the risk of cancer in women: epidemiologic evidence and possible biologic mechanisms. *Womens Health (Lond)*. mars 2011;7(2):227-37.
158. Tariq H, Gul A, Zubair M, Jaffer SR, Zafar N, Sadaf G. Targeted Next-generation Sequencing for Reliable Detection of Genetic Status in Breast Cancer. *J Coll Physicians Surg Pak*. juill 2021;30(7):837-40.
159. Guglielmi C, Scarpitta R, Gambino G, Conti E, Bellè F, Tancredi M, et al. Detection of Germline Variants in 450 Breast/Ovarian Cancer Families with a Multi-Gene Panel Including Coding and Regulatory Regions. *Int J Mol Sci*. 19 juill 2021;22(14):7693.
160. Ren M, Orozco A, Shao K, Albanez A, Ortiz J, Cao B, et al. Germline variants in hereditary breast cancer genes are associated with early age at diagnosis and family history in Guatemalan breast cancer. *Breast Cancer Res Treat*. sept 2021;189(2):533-9.
161. Rogoża-Janiszewska E, Malińska K, Górski B, Scott RJ, Cybulski C, Kluźniak W, et al. Prevalence of germline TP53 variants among early-onset breast cancer patients from Polish population. *Breast Cancer*. janv 2021;28(1):226-35.
162. Stuebe AM, Wise A, Nguyen T, Herring A, North KE, Siega-Riz AM. Maternal genotype and gestational diabetes. *Am J Perinatol*. janv 2014;31(1):69-76.
163. Lambrinoudaki I, Vlachou SA, Creatsas G. Genetics in gestational diabetes mellitus: association with incidence, severity, pregnancy outcome and response to treatment. *Curr Diabetes Rev*. nov 2010;6(6):393-9.
164. Robitaille J, Grant AM. The genetics of gestational diabetes mellitus: evidence for relationship with type 2 diabetes mellitus. *Genet Med*. avr 2008;10(4):240-50.
165. Chen Y, Liao WX, Roy AC, Loganath A, Ng SC. Mitochondrial gene mutations in gestational diabetes mellitus. *Diabetes Res Clin Pract*. avr 2000;48(1):29-35.
166. Khan IA, Jahan P, Hasan Q, Rao P. Angiotensin-converting enzyme gene insertion/deletion polymorphism studies in Asian Indian pregnant women biochemically identifies gestational diabetes mellitus. *J Renin Angiotensin Aldosterone Syst*. déc 2014;15(4):566-71.
167. Singh A, Srivastava N, Amit S, Prasad SN, Misra MP, Ateeq B. Association of AGTR1 (A1166C) and ACE (I/D) Polymorphisms with Breast Cancer Risk in North Indian Population. *Transl Oncol*. avr 2018;11(2):233-42.
168. Witczak M, Wilczyński J, Gulczyńska E, Talar T, Mordalska A, Łopaczyńska D, et al. What

is the impact of gestational diabetes mellitus on frequency of structural chromosome aberrations in pregnant women and their offspring? *Mutat Res.* juin 2017;818:27-30.

169. Cnattingius S, Torráng A, Ekblom A, Granath F, Petersson G, Lambe M. Pregnancy characteristics and maternal risk of breast cancer. *JAMA.* 16 nov 2005;294(19):2474-80.
170. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol.* 15 sept 2000;152(6):514-27.
171. Xie S-H, Yang S, Chen J, Cheng J-Q. Gestational diabetes mellitus and subsequent cancer risk: shared risk factors, causality or confounding? *Asian Pac J Cancer Prev.* 2014;15(24):11001.
172. Mamguem Kamga A, Dumas A, Joly F, Billa O, Simon J, Poillot M-L, et al. Long-Term Gynecological Cancer Survivors in Côte d'Or: Health-Related Quality of Life and Living Conditions. *Oncologist.* juill 2019;24(7):e490-500.
173. Mamguem Kamga A, Dumas A, Joly F, Simon J, Billa O, Poillot M-L, et al. Socio-economic and occupational outcomes of long-term survivors of gynaecological cancer: A French population-based study. *Eur J Cancer Care (Engl).* juill 2020;29(4):e13235.
174. Goldberg M, Carton M, Descatha A, Leclerc A, Roquelaure Y, Santin G, et al. CONSTANCES: a general prospective population-based cohort for occupational and environmental epidemiology: cohort profile. *Occup Environ Med.* 1 janv 2017;74(1):66-71.
175. Scailteux L-M, Balusson F, Vincendeau S, Rioux-Leclercq N, Nowak E. Rationale and design of the CANARI study: a case-control study investigating the association between prostate cancer and 5-alpha-reductase inhibitors for symptomatic benign prostate hypertrophy by linking SNIIRAM and pathology laboratories in a specific region in France. *Fundamental & Clinical Pharmacology.* 2018;32(1):120-9.
176. Raffray M, Bayat S, Lassalle M, Couchoud C. Linking disease registries and nationwide healthcare administrative databases: the French renal epidemiology and information network (REIN) insight. *BMC Nephrol.* 28 janv 2020;21:25.
177. Page d'accueil [Internet]. Health Data Hub. [cité 3 nov 2021]. Disponible sur: <https://www.health-data-hub.fr/>
178. Medicare - La santé des personnes âgées [Internet]. Manuels MSD pour le grand public. [cité 13 nov 2021]. Disponible sur: <https://www.msmanuals.com/fr/accueil/la-sant%C3%A9-des-personnes-%C3%A2g%C3%A9es/couverture-de-sant%C3%A9-pour-les-personnes-%C3%A2g%C3%A9es/medicare>
179. Zhou J, Cueto J, Ko NY, Hoskins KF, Nabulsi NA, Asfaw AA, et al. Population-based recurrence rates among older women with HR-positive, HER2-negative early breast cancer:

Clinical risk factors, frailty status, and differences by race. *Breast*. 13 août 2021;59:367-75.

180. Koné AP, Scharf D. Prevalence of multimorbidity in adults with cancer, and associated health service utilization in Ontario, Canada: a population-based retrospective cohort study. *BMC Cancer*. 14 avr 2021;21(1):406.
181. Pearson C, Fraser J, Peake M, Valori R, Poirier V, Coupland VH, et al. Establishing population-based surveillance of diagnostic timeliness using linked cancer registry and administrative data for patients with colorectal and lung cancer. *Cancer Epidemiol*. août 2019;61:111-8.
182. Kanemura H, Tamura T, Nishimura N, Kobayashi D, Higashi T. Thymic epithelial tumor treatment in Japan: analysis of hospital cancer registry and insurance claims data, 2012-2014. *Jpn J Clin Oncol*. 9 mars 2020;50(3):310-7.

# Annexes

# **Annexe 1**



**ARTICLE**

Clinical Study

# Survival in patients with HR+/HER2– metastatic breast cancer treated with initial endocrine therapy versus initial chemotherapy. A French population-based study

Julien Simon<sup>1,2</sup>, Marie Chaix<sup>3</sup>, Oumar Billa<sup>1</sup>, Ariane Mamguem Kamga<sup>1</sup>, Patrick Roignot<sup>4</sup>, Sylvain Ladoire<sup>3</sup>, Charles Coutant<sup>5</sup>, Patrick Arveux<sup>1,6</sup>, Catherine Quantin<sup>2,7,8</sup> and Tienhan Sandrine Dabakuyo-Yonli<sup>1,9</sup>

**BACKGROUND:** According to international guidelines, endocrine therapy (ET) is the preferred option for hormone receptor-positive (HR+) HER2-negative (HER2–) metastatic breast cancer. In spite of clear recommendations, these are not strictly followed in daily practice. The objectives of this study were to investigate the effect of the first anti-metastatic treatment therapy choice on progression-free survival (PFS) and overall survival (OS).

**METHODS:** In this population-based study, we included patients with HR+/HER2– metastatic breast cancer recorded in the Côte d'Or Breast Cancer Registry. Differences in PFS and OS between patients initially treated with chemotherapy (CT) or ET were analysed in Cox proportional hazards models. In a sensitivity analysis, we used a propensity score (PS) to limit the indication bias.

**RESULTS:** Altogether, 557 cases were included, 280 received initial ET and 277 received initial CT. PFS and OS in patients initially treated with ET was improved significantly when compared to patients with initial CT (respectively, HR = 0.83 (95% CI 0.69–0.99) and HR = 0.71 (95% CI 0.58–0.86)). The results of the sensitivity analysis supported these findings.

**CONCLUSION:** This study shows that treating patients with HR+/HER2– metastatic breast cancer with initial ET could provide a survival advantage in comparison with initial CT.

*British Journal of Cancer* (2020) 123:1071–1077; <https://doi.org/10.1038/s41416-020-0979-3>

**BACKGROUND**

Breast cancer is the most common malignancy among women, representing almost a quarter of all cancer cases in females in Europe and worldwide<sup>1</sup> and causing >12 thousand deaths in France in 2018.<sup>2</sup>

Thanks to detection at an earlier stage and the development of better treatments over the past decades, survival has continued to improve in both non-metastatic<sup>3</sup> and metastatic breast cancers (MBC)<sup>4–6</sup> which nonetheless remain incurable.

Treatment recommendations ought to be made on an individual basis and consider hormone receptor (HR) and human epidermal growth factor receptor 2 (HER2) status. Indeed, the luminal subtype (HR+/HER2–) represents about 70% of all breast cancers. Some studies have shown that, even though no difference was found between chemotherapy (CT) or endocrine therapy (ET) in terms of survival, ET is a better choice since it is less toxic.<sup>7,8</sup> Some clinical trials showed good progression-free survival (PFS) for ET alone and an increase in PFS for ET in association with targeted therapies.<sup>9–11</sup> According to international guidelines, ET

should be the preferred option in luminal MBC, even in the presence of visceral disease, unless there is a visceral crisis or proof of endocrine resistance.<sup>12–14</sup>

Despite these data, some recent studies have shown that guidelines are not always followed in daily practice and that medical practices vary considerably.<sup>15–19</sup> However, to date, none of these studies has been done using population-based data.

In this population-based study, which takes into account all of the women living in Côte d'Or (Côte d'Or Breast and Gynaecological Cancer Registry), we investigated the effect of the first-line treatment (ET or CT) on PFS and overall survival (OS) in luminal MBC patients.

**METHODS**

Study population and data collection

A population-based study was undertaken using data from the Côte d'Or Breast and Gynaecological Cancer Registry. This registry records all cases of breast and gynaecological cancers in women

<sup>1</sup>Breast and Gynecologic Cancer Registry of Côte d'Or, Georges-François Leclerc Comprehensive Cancer Care Centre, 1 rue Professeur Marion, Dijon, France; <sup>2</sup>Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases, Inserm U1181, Villejuif, France; <sup>3</sup>Medical Oncology Department, Georges-François Leclerc Comprehensive Cancer Care Centre, 1 rue Professeur Marion, Dijon, France; <sup>4</sup>Pathology Centre, 33 rue Nicolas Bornier, Dijon, France; <sup>5</sup>Surgical Oncology Department, Georges-François Leclerc Comprehensive Cancer Care Centre, 1 rue Professeur Marion, Dijon, France; <sup>6</sup>Center for Research in Epidemiology and Population Health, Inserm U1018, Villejuif, France; <sup>7</sup>CHRU Dijon, Service de Biostatistiques et d'Information Médicale (DIM), Dijon 21000, France; <sup>8</sup>INSERM, CIC 1432, Univ Bourgogne Franche Comte, Dijon, France and <sup>9</sup>Inserm U1231 Lipids, Nutrition, Cancer Research Center, Dijon, France  
Correspondence: Tienhan Sandrine Dabakuyo-Yonli (sdabakuyo@cgfl.fr)

Received: 14 February 2020 Revised: 12 June 2020 Accepted: 23 June 2020  
Published online: 17 July 2020

living in the department of Côte d'Or at the time of diagnosis of the primary cancer, and it is the only registry in France focusing on these cancers. Women diagnosed with HR+/-HER2-, MBC from January 1998 to December 2016 were retrospectively selected in the registry data and included in this study. Only invasive ductal or lobular carcinomas were included. HR and HER2 status were determined using immunohistochemistry, completed by a fluorescent *in situ* hybridisation exam in patients with an equivocal result. Pathologists in Côte d'Or usually classify a patient as RH+ when one of the 2 receptors (oestrogen receptor or progesterone receptor) is >10%. HER2 status was retrospectively determined for cases diagnosed before routine analyses. Women who had metastasis after 2016, those who did not receive a systemic anti-metastatic treatment and those who had concurrent CT and ET in the first line were excluded.

Comorbidities were defined according to the Charlson Comorbidity Index (CCI)<sup>20</sup> and were classified into two groups: no comorbidities (CCI = 0) and at least one comorbidity (CCI > 0). In accordance with European guidelines, we defined primary endocrine resistance as: relapse while on the first 2 years of adjuvant ET. Secondary endocrine resistance was defined as: relapse while on adjuvant ET but after the first 2 years or relapse within 12 months of completing adjuvant ET.<sup>12</sup> Furthermore, we described the characteristics of the initial systemic therapies.

Only anonymised data were used for this study. The registry has the necessary regulatory agreements from the French Data Protection Authority (Commission nationale de l'informatique et des libertés (CNIL)) to use patients' data. The CNIL ensures that data privacy laws are respected (CNIL authorisation number DR-2012-038).

#### Primary and secondary outcomes

The primary outcome was PFS defined as the time from initiation of the first systemic therapy for metastatic disease to disease progression or death, whichever occurred first. Progression, as determined by the DATECAN initiative,<sup>21</sup> was defined as progression of the initial metastases or the occurrence of new metastases.

Our secondary outcome was OS, which was defined as the time from initiation of the first systemic anti-metastatic therapy (CT or ET) to death.

Women alive without progression were censored at the end of the study period (December 31, 2017).

#### Statistical analyses

We used chi-square tests or Fisher's exact tests for categorical variables and Wilcoxon rank-sum test for continuous variables to compare baseline characteristics (patient characteristics, primary tumour characteristics, adjuvant therapies (radiotherapy, CT and ET before the diagnosis of metastasis) and metastasis characteristics) between patients with initial CT and those with initial ET.

Survival outcomes were estimated using the Kaplan-Meier method and compared between groups using the log-rank statistic.

Differences in PFS and OS between patients initially treated with CT or ET were analysed using Cox proportional hazards models after adjustment for factors with a *P* value <0.2 in univariate analyses. We did not take adjuvant treatments into account in the multivariate analyses because these variables only concern secondary metastases. Moreover, because we only studied metastatic cancers, we did not take the tumour-node stage of the initial tumour into account in the multivariate analyses.

The proportional hazards assumption was checked for each variable using Schoenfeld Residuals tests, and for each factor that did not meet the assumption, we included an interaction with time in the model. We also tested interactions and correlations ( $\rho < 0.7$ ) between the different variables.

To determine the association between prognostic factors and survival outcomes, we fitted univariate Cox proportional hazards models. Prognostic factors were age at diagnosis of MBC (<65 years/ $\geq 65$  years), CCI (CCI = 0/CCI > 0), menopausal status (premenopausal/postmenopausal), primary tumour type (invasive ductal/lobular carcinomas), primary metastatic disease (yes/no), initial metastatic sites (bone only, visceral or multiple), primary endocrine resistance (yes/no) and year of primary tumour diagnosis (<2008/ $\geq 2008$ ). We chose 2008 as a cut-off because changes in practices and in the definition of metastatic disease by the registry occurred in 2008. More specifically, HR/HER2 status became a more important factor in therapeutic decision-making, and before 2008, cancer stage was determined in the registry using the fifth edition of the American Joint Committee for Cancer (AJCC) tumour-node-metastasis (TNM) classification of malignant tumours while the sixth edition was used after 2008.

In the first sensitivity analysis, we used a propensity score (PS) to consider the subjectivity of the physician's choice<sup>22</sup> and to adjust on the variables that were different between the two groups of patients (those who underwent initial ET and those who underwent initial CT). Each variable that differed between patients with initial CT and initial ET was included in the multivariate logistic model. First, each patient treated with initial CT was matched with a patient who underwent initial ET with the same characteristics using the PS. Nearest-neighbour matching was used, with a maximal PS distance between matched subjects (calliper) of 0.1; unpaired patients were excluded.

We then performed a second sensitivity analysis in which patients with initial CT who were given maintenance ET before progression were censored at the start of maintenance ET. Similar to a previous study,<sup>18</sup> this analysis was done to compare the group of patients who were exclusively treated by ET and the others exclusively treated by CT, in an attempt to override the effect of maintenance ET.

We then performed a third sensitivity analysis in which ET was considered a time-dependent variable (taking into account ET no matter when ET was administered, without censoring on maintenance ET). In the initial ET group, ET was started the first day of the follow-up, and in the initial CT group, ET was started when the maintenance ET started.

Statistical analysis was performed using the SAS 9.4 software (SAS Institute, Cary, NC, USA). Statistical significance was set at a *P* value of <0.05.

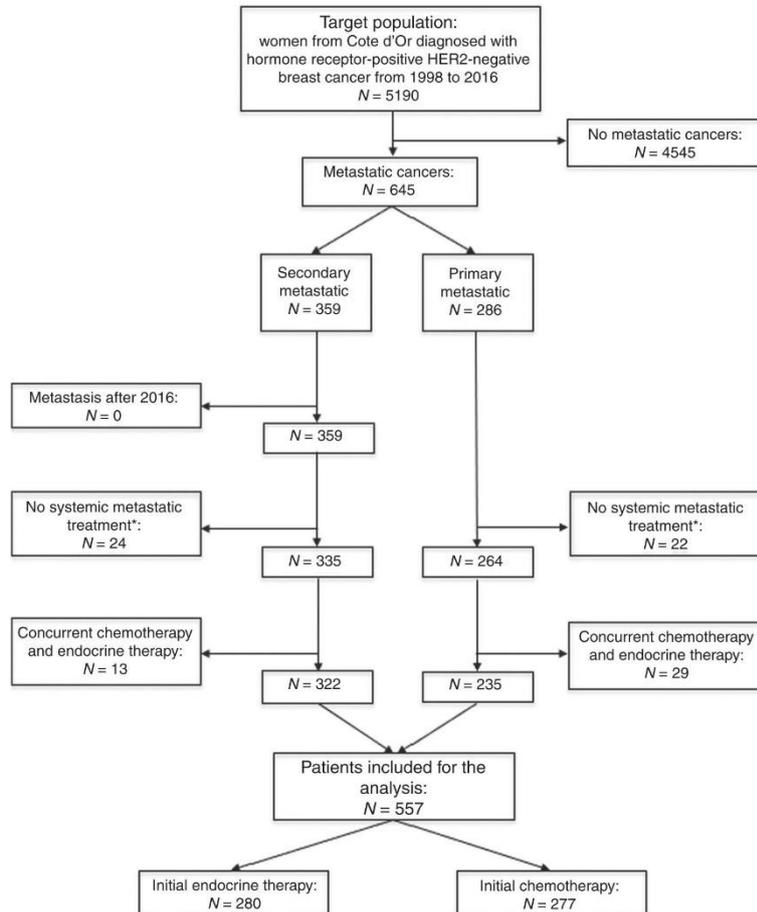
## RESULTS

### Patient characteristics

Altogether, 5190 women from Côte d'Or were diagnosed with HR-positive, HER2-negative breast cancer from 1998 to 2016, and 645 cases were or became MBC. Among these, 557 patients received a systemic anti-metastatic treatment and were analysed (277 first-line CT and 280 first-line ET) (Fig. 1).

These patients had a median age of 65 years [interquartile range (IQR): 55–77 years] when MBC was diagnosed. The median follow-up was 25 months (IQR: 13–45 months). The median PFS was 13 months (95% confidence interval (CI) 12–14) and the median OS was 26 months (95% CI 23–29). At the end of the study period (December 31, 2017), 95% of all patients (531/557) had progressed and 83% (463/557) had died.

Altogether, 50% of patients were treated with initial CT and 50% of patients received initial ET; their characteristics are summarised in Table 1. There were no statistical differences between the two groups concerning primary tumour characteristics (diagnostic year, histological subtype, HR status and node status). Patients treated with initial CT were younger, had fewer comorbidities and were given more adjuvant therapies than those treated with initial ET. Initial ET was more frequently used in women with primary MBC and bone metastases only; there was less endocrine resistance in this group.



**Fig. 1 Flowchart of the study population.** Asterisk (\*): patients who received exclusive comfort treatment or those who received only surgery and/or radiotherapy as treatment for metastatic cancer.

#### Treatments

Among patients with initial ET, 72% (203/280) were treated with aromatase inhibitors, 24% (66/280) with anti-oestrogens (tamoxifen, fulvestrant) and 4% (11/280) with chemical or surgical castration.

Among patients with initial CT, 32% (88/277) were given anthracycline-based treatments and 60% (165/277) were given taxane-based treatments. Altogether, 53% (146/277) changed to maintenance ET, and among these 60% (88/146) changed within 6 months of starting CT.

#### Analysis of survival outcomes

PFS and OS were compared in patients who received ET or CT as the initial systemic anti-metastatic therapy. The median PFS of patients treated with initial ET was 14 months (95% CI 12–16) and the median PFS of those treated with initial CT was 11 months (95% CI 10–13;  $P = 0.0394$ ) (Fig. 2a). The median OS of patients treated with initial ET was 32 months (95% CI 27–38) and the median OS of those treated with initial CT was 28 months (95% CI 23–32;  $P = 0.2572$ ) (Fig. 2b).

Adjustment factors for the multivariate model for PFS were primary metastatic disease, initial metastatic sites, menopausal

status, CCI and primary endocrine resistance. PFS was significantly better in patients treated with initial ET than in those with initial CT, with a hazard ratio (HR) of 0.83 (95% CI 0.69–0.99). Adjustment factors for the multivariate model for OS were: age at MBC diagnosis, menopausal status, CCI, primary metastatic disease, initial metastatic sites, and primary endocrine resistance. Because “primary metastatic disease” did not meet the proportional hazards assumption, we included an interaction between “primary metastatic disease” and time in the multivariate model. OS was significantly better in patients treated with initial ET than in those with initial CT, with an HR of 0.71 (95% CI 0.58–0.86) (Table 2).

In the analyses with PS (first sensitivity analysis), matched cases ( $N = 296$ ) showed an HR = 0.87 (95% CI 0.70–1.07) for PFS and an HR of 0.79 (95% CI 0.63–0.99) for OS.

In the second sensitivity analysis, in which patients of the CT group with maintenance ET were censored, PFS and OS were significantly better in patients treated with initial ET than in those with initial CT with an HR of 0.72 (95% CI 0.56–0.92) and 0.61 (95% CI 0.51–0.74), respectively.

In the third sensitivity analysis, in which ET was considered a time-dependent variable, PFS was not significantly better in

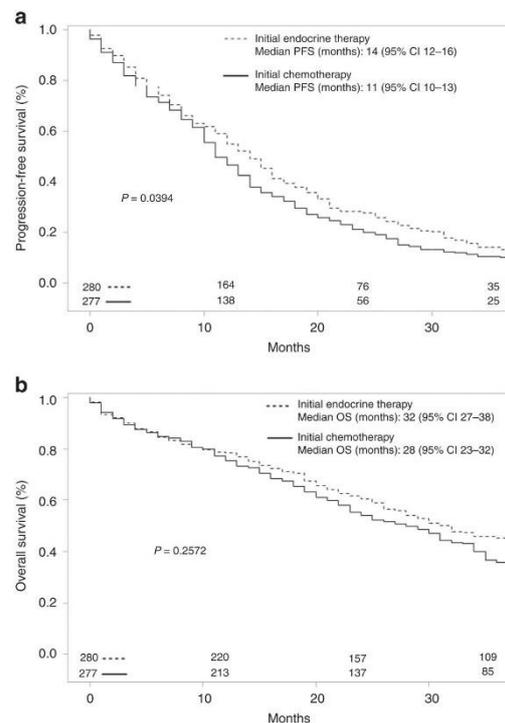
**Table 1.** Patient demographics and disease characteristics of the endocrine and chemotherapy groups.

Characteristics	Initial chemotherapy, N = 277, n (%)	Initial endocrine therapy, N = 280, n (%)	P
<b>Age at MBC diagnosis</b>			
Median in years (IQR)	61 (50–69)	73 (61–80)	<b>&lt;0.0001<sup>a</sup></b>
<65 years	170 (61)	97 (35)	<b>&lt;0.0001</b>
≥65 years	107 (39)	183 (65)	
<b>Menopausal status</b>			
Premenopausal	83 (30)	35 (12)	<b>&lt;0.0001</b>
Postmenopausal	194 (70)	245 (88)	
<b>Charlson Comorbidity Index (CCI)</b>			
CCI = 0	178 (64)	130 (46)	<b>&lt;0.0001</b>
CCI > 0	99 (36)	150 (54)	
<b>Primary tumour diagnostic year</b>			
<2008	97 (35)	119 (42)	0.07
≥2008	180 (65)	161 (58)	
<b>Primary tumour type</b>			
Ductal	231 (83)	223 (80)	0.2543
Lobular	46 (17)	57 (20)	
<b>Hormone receptor status</b>			
ER positive	269 (97)	276 (99)	0.2355
PR positive	199 (72)	208 (74)	0.2388
<b>Primary tumour stage<sup>b</sup></b>			
T1–T2	187 (67)	172 (61)	0.1053
T3–T4	88 (32)	108 (39)	
Unknown	2 (1)	0	
<b>Node status<sup>b</sup></b>			
Node negative	139 (50)	140 (50)	0.9667
Node positive	132 (48)	132 (48)	
Unknown	6 (2)	8 (2)	
<b>Adjuvant systemic therapy<sup>c</sup></b>			
Endocrine therapy	164 (91)	105 (74)	<b>&lt;0.0001</b>
Chemotherapy	141 (78)	82 (58)	<b>&lt;0.0001</b>
None	6 (3)	24 (17)	<b>&lt;0.0001<sup>d</sup></b>
<b>Adjuvant radiotherapy<sup>c</sup></b>			
Yes	163 (90)	122 (87)	0.3244
No	18 (10)	19 (13)	
<b>Primary metastatic disease</b>			
Yes	96 (35)	139 (50)	<b>0.0002</b>
No	181 (65)	141 (50)	
<b>Metastasis-free interval<sup>c</sup></b>			
<24 months	26 (14)	29 (21)	0.1423
≥24 months	155 (86)	112 (79)	
<b>Initial metastatic sites</b>			
Bone only	62 (22)	136 (49)	<b>&lt;0.0001</b>

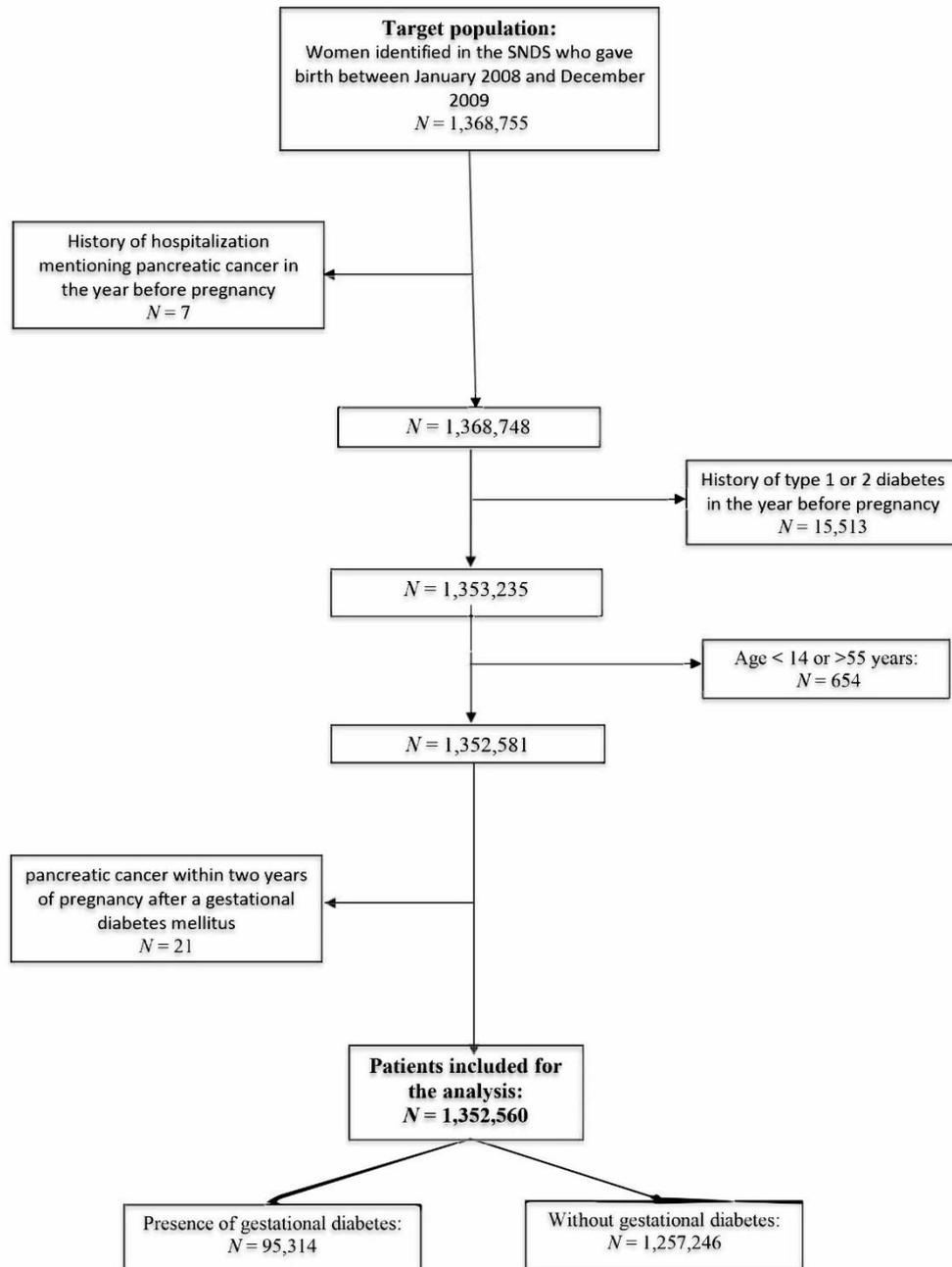
**Table 1 continued**

Characteristics	Initial chemotherapy, N = 277, n (%)	Initial endocrine therapy, N = 280, n (%)	P
Visceral	90 (33)	66 (23)	
Multiple	125 (45)	78 (28)	
<b>Endocrine resistance</b>			
Primary	50 (18)	35 (13)	<b>0.0003</b>
Secondary	74 (27)	45 (16)	
Hormone responsive	153 (55)	200 (71)	

Bold values indicate statistically significant results.  
n number, IQR interquartile range, MBC metastatic breast cancer, ER oestrogen receptor, PR progesterone receptor.  
<sup>a</sup>Wilcoxon rank-sum test.  
<sup>b</sup>Fifth edition before 2008 and sixth edition since 2008 of the TNM classification of malignant tumours.  
<sup>c</sup>Only on secondary metastatic diseases.  
<sup>d</sup>Fisher's exact test.



**Fig. 2 a) Progression-free survival and b) overall survival according to initial systemic anti-metastatic therapy of patients with hormone receptor-positive metastatic breast cancer.** CI confidence interval, PFS progression-free survival, OS overall.



**Figure 1.** Flowchart of the study population.

Similarly, 15,531 women with type 2 diabetes and 654 women who did not meet the age criteria were also excluded, leading to 1,352,560 women who were included in the analysis population. The mean duration time from the delivery to the onset of PC over all included patients was 1910 days (sd = 604). We also found that the mean duration time

**Table 3.** Distribution of characteristics of patients matched with the propensity score according to the initial systemic anti-metastatic therapy.

Characteristics	Initial chemotherapy, N = 186, n (%)	Initial endocrine therapy, N = 186, n (%)	P
Age at diagnostic MBC			0.9170
<65 years	84 (45)	83 (45)	
≥65 years	102 (55)	103 (55)	
Hormonal status			0.7861
Premenopausal	34 (18)	32 (18)	
Postmenopausal	152 (82)	154 (82)	
Charlson Comorbidity Index (CCI)			0.7524
CCI = 0	110 (58)	107 (56)	
CCI > 0	76 (42)	79 (44)	
Primary tumour type			0.7950
Ductal	150 (81)	148 (80)	
Lobular	36 (19)	38 (20)	
Primary metastatic disease			0.5678
Yes	110 (59)	105 (56)	
No	76 (41)	81 (44)	
Initial metastatic sites			0.8044
Bone only	60 (32)	62 (33)	
Visceral	53 (29)	57 (31)	
Multiple	73 (39)	67 (36)	
Endocrine resistance			0.1161
Primary	108 (58)	125 (67)	
Secondary	28 (15)	27 (15)	
Hormone responsive	50 (27)	34 (18)	

n number, MBC metastatic breast cancer.

Lobbezto et al. included patients prior to introduction of daily bevacizumab, whereas our study included patients subsequently; the use of bevacizumab could have increased PFS in the initial CT group.<sup>24</sup> To avoid the potential bias due to informative censoring, we also performed a third sensitivity analysis in which ET was considered a time-dependent variable, without censoring on maintenance ET. This analysis gave us an estimate of the effect of ET no matter when ET was administered and showed results similar to the primary analysis (HR = 0.84, 95% CI 0.70–1.01 for PFS and HR = 0.69, 95% CI 0.57–0.84 for OS). However, in this analysis, the introduction of a new variable (time-dependent variable) reduced the power, which may explain why this result was not statistically significant for PFS. Of course, this analysis is not perfectly comparable to the primary analysis as we have not only considered initial ET but also maintenance ET.

Contrary to our study, Jacquet et al. found no significant differences in PFS and OS between initial ET and initial CT. However, the two populations are not comparable; indeed they included only diseases sensitive to aromatase inhibitors, whereas we included all patients, even the hormone-resistant ones.

Our results also evidenced that the treatments administered to patients in Côte d'Or were considerably different from the treatments recommended by international guidelines.<sup>12–14</sup> Half of our study population received initial CT even though only 15%

of these patients had clinical primary endocrine resistance. The population-based nature of our study could explain the large proportion of initial CT in comparison with hospital cohorts given that we recruited patients from different types of medical establishments. Furthermore, the first international consensus guidelines for advanced breast cancer were published relatively recently, in 2012.<sup>25</sup> In our study, the proportion of patients receiving initial CT continued to increase after 2008 despite the publication of new guidelines.

Almost two-thirds of patients who changed to maintenance ET did so within 6 months of starting CT. We could suppose that this was partly related to toxicity but were not able to obtain specific information about toxicity or visceral crises.

Our data were obtained from the Côte d'Or Breast and Gynaecological Cancer Registry, which respects a strict data quality control policy including regular checks to ensure complete follow-up for every patient. Our high-quality data is not subject to the selection bias that can be observed with hospital cohorts. However, the registry covers a relatively small geographic area, obliging us to study the data over a long period in order to obtain a large enough sample. Diagnostic procedures and therapies have improved since the start of the study. In addition, the tumour classification changed: we used the fifth edition of the AJCC TNM classification of malignant tumours before 2008, and the sixth edition after 2008. Consequently, before 2008 we considered a few metastatic tumours that would have been considered N3 after 2008. Furthermore, the level of intensity of HRs was poorly documented in the registry in the early years of the study because clinicians did not take this information into account when choosing the first systemic anti-metastatic treatment. Likewise, we were unable to include the performance status because this information was not routinely collected in our register. However, we did take into account the CCI, which provides a good indicator of general health.

Our main results, which show a statistically significant difference in favour of ET, may have several explanations. First, the patients included in our study are all HR+ and who were therefore more sensitive to ET. In addition, HR+/HER2- MBC are slower replicating cancers, which are therefore less sensitive to cytotoxic treatment than quicker replicating cancers.

Moreover, the results of the first studies on new targeted therapies (cyclin-dependent kinase 4/6 inhibitors), which are expensive but less toxic, showed an ability to prolong PFS.<sup>9,23</sup> However, the factors predicting response have not yet been identified. In our study, very few patients received targeted therapies because they are fairly recent relative to our follow-up period. Further studies will be needed to determine the best strategy for treating patients with HR+/HER2- MBC.

## CONCLUSION

In conclusion, in this population-based study, we have performed a primary analysis along with three sensitivity analyses (PS, censoring at the start of the maintenance ET and considering ET as a time-dependent variable) to overcome potential biases. All these analyses showed very similar results indicating that initial treatment with ET improves survival in patients with HR-positive HER2-negative MBC, as compared with initial treatment with CT. Knowing that ET causes fewer side effects, we suggest that international guidelines should be followed more closely, and more extensively disseminated, with a clarification of the definition of visceral crisis.

## ACKNOWLEDGEMENTS

We thank L. Bouzigues, M-L Poillot and G. Laporte for data collection and L. Collet, P. Bastable, S. Rankin and I. Gregoire for reviewing the manuscript.

**AUTHOR CONTRIBUTIONS**

J.S. designed the study, acquired the data, controlled the quality of data and algorithms, analysed and interpreted the data, did the data analyses, prepared the manuscript, edited the manuscript and reviewed the manuscript. M.C. acquired the data, controlled the quality of data, prepared the manuscript, edited the manuscript and reviewed the manuscript. O.B. acquired the data, controlled the quality of data and algorithms and reviewed the manuscript. A.M.K. and P.R. acquired the data, controlled the quality of data and algorithms and reviewed the manuscript. S.L. acquired the data, analysed and interpreted the data, controlled the quality of data and algorithms and reviewed the manuscript. C.C. reviewed the manuscript. P.A. and C.Q. designed the study, analysed and interpreted the data, did the data analyses, prepared the manuscript, edited the manuscript and reviewed the manuscript. T.S.D.-Y. designed the study, analysed and interpreted the data, did the data analyses, prepared the manuscript and reviewed the manuscript.

**ADDITIONAL INFORMATION**

**Ethics approval and consent to participate** Patient consent is not required for observational studies. In addition, patients included in the registry are informed during their treatment that their data can be used for research purposes and they can then refuse to have all or part of their data used. As a cancer registry, the Côte d'Or Breast and Gynaecological Cancer Registry has obtained the approval of the French national data protection authority (Commission nationale de l'informatique et des libertés CNIL N1539818) for data recording, and this study was approved by the French national data protection authority (Commission nationale de l'informatique et des libertés MR003 N\_1989764 v. 0)

**Consent to publish** Not applicable.

**Data availability** The data used for this study were extracted from the Côte d'Or Breast and Gynaecological Cancer Registry database.

**Competing interests** The authors declare no competing interests.

**Funding information** The Côte d'Or Breast and Gynaecological Cancer Registry is funded by the National Cancer Institute (INCA) [<http://www.e-cancer.fr/>; Grant number 2015-027]. This research did not receive any additional grants from funding agencies in the public, private or not-for-profit sectors. The INCA provides financial support for the operation of the Côte d'Or Breast Cancer Registry.

**Note** This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution 4.0 International (CC BY 4.0).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**REFERENCES**

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. 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).
- Defossez, G., Le Guyader-Peyrou, S., Uhry, Z., Grosclaude, P., Remontet, L., Colonna, M. et al. *Estimations Nationales de l'incidence et de la Mortalité par Cancer en France Métropolitaine Entre 1990 et 2018. Volume 1 – Tumeurs Solides* (Santé Publique France, Saint-Maurice (Fra), 2019).
- Stebbing, J., Delaney, G. & Thompson, A. Breast cancer (non-metastatic). *BMJ Clin. Evid.* **2007**, 0102 (2007).
- Stebbing, J. & Ngan, S. Breast cancer (metastatic). *BMJ Clin. Evid.* **2010**, 0811 (2010).
- Dawood, S., Broglio, K., Gonzalez-Angulo, A. M., Buzdar, A. U., Hortobagyi, G. N. & Giordano, S. H. Trends in survival over the past two decades among white and black patients with newly diagnosed stage IV breast cancer. *J. Clin. Oncol.* **26**, 4891–4898 (2008).
- Chia, S. K., Speers, C. H., D'yachkova, Y., Kang, A., Malfair-Taylor, S., Barnett, J. et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. *Cancer* **110**, 973–979 (2007).
- Wilcken, N., Hornbuckle, J. & Gherzi, D. Chemotherapy alone versus endocrine therapy alone for metastatic breast cancer. *Cochrane Database Syst. Rev.* **2**, CD002747 (2003).
- Trédan, O., Follana, P., Moullet, I., Cropet, C., Trager-Maury, S., Dauba, J. et al. A phase III trial of exemestane plus bevacizumab maintenance therapy in patients with metastatic breast cancer after first-line taxane and bevacizumab: a GINECO group study. *Ann. Oncol.* **27**, 1020–1029 (2016).
- Finn, R. S., Martin, M., Rugo, H. S., Jones, S., Im, S.-A., Gelmon, K. et al. Palbociclib and letrozole in advanced breast cancer. *N. Engl. J. Med.* **375**, 1925–1936 (2016).
- Robertson, J. F. R., Cheung, K.-L., Noguchi, S., Shao, Z., Degboe, A., Lichfield, J. et al. Health-related quality of life from the FALCON phase III randomised trial of fulvestrant 500 mg versus anastrozole for hormone receptor-positive advanced breast cancer. *Eur. J. Cancer* **94**, 206–215 (2018).
- Royce, M., Bachelot, T., Villanueva, C., Özgüroglu, M., Azevedo, S. J., Cruz, F. M. et al. Everolimus plus endocrine therapy for postmenopausal women with estrogen receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer: a clinical trial. *JAMA Oncol.* **4**, 977–984 (2018).
- Cardoso, F., Senkus, E., Costa, A., Papadopoulos, E., Aapro, M., André, F. et al. 4th ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 4). *Ann. Oncol.* **29**, 1634–1657 (2018).
- Gradishar, W. J., Anderson, B. O., Blair, S. L., Burstein, H. J., Cyr, A., Elias, A. D. et al. Breast cancer version 3.2014. *J. Natl. Compr. Cancer Netw.* **12**, 542–590 (2014).
- Partridge, A. H., Rumble, R. B., Carey, L. A., Come, S. E., Davidson, N. E., Di Leo, A. et al. Chemotherapy and targeted therapy for women with human epidermal growth factor receptor 2-negative (or unknown) advanced breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **32**, 3307–3329 (2014).
- André, F., Neven, P., Marinsek, N., Zhang, J., Baladi, J.-F., Degun, R. et al. Disease management patterns for postmenopausal women in Europe with hormone-receptor-positive, human epidermal growth factor receptor-2 negative advanced breast cancer. *Curr. Med. Res. Opin.* **30**, 1007–1016 (2014).
- Swallow, E., Zhang, J., Thomason, D., Tan, R.-D., Kageleiry, A. & Signorovitch, J. Real-world patterns of endocrine therapy for metastatic hormone-receptor-positive (HR+)/human epidermal growth factor receptor-2-negative (HER2-) breast cancer patients in the United States: 2002–2012. *Curr. Med. Res. Opin.* **30**, 1537–1545 (2014).
- Jerusalem, G., Neven, P., Marinsek, N., Zhang, J., Degun, R., Benelli, G. et al. Patterns of resource utilization and cost for postmenopausal women with hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer in Europe. *BMC Cancer* **15**, 787 (2015).
- Lobbezoo, D. J. A., van Kampen, R. J. W., Voogd, A. C., Dercksen, M. W., van den Berkmortel, F., Smilde, T. J. et al. In real life, one-quarter of patients with hormone receptor-positive metastatic breast cancer receive chemotherapy as initial palliative therapy: a study of the Southeast Netherlands Breast Cancer Consortium. *Ann. Oncol.* **27**, 256–262 (2016).
- Jacquet, E., Lardy-Cléaud, A., Pistilli, B., Franck, S., Cottu, P., Delaloue, S. et al. Endocrine therapy or chemotherapy as first-line therapy in hormone receptor-positive HER2-negative metastatic breast cancer patients. *Eur. J. Cancer* **95**, 93–101 (2018).
- 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).
- Gourgou-Bourgade, S., Cameron, D., Poortmans, P., Asselain, B., Azria, D., Cardoso, F. et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann. Oncol.* **26**, 873–879 (2015).
- Rosebaum, P. & Rubin, D. The central role of propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
- Hortobagyi, G. N., Stemmer, S. M., Burris, H. A., Yap, Y.-S., Sonke, G. S., Paluch-Shimon, S. et al. Ribociclib as first-line therapy for HR-positive, advanced breast cancer. *N. Engl. J. Med.* **375**, 1738–1748 (2016).
- Miller, K., Wang, M., Gralow, J., Dickler, M., Cobleigh, M., Perez, E. A. et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N. Engl. J. Med.* **357**, 2666–2676 (2007).
- Pritchard, K. I. & Sousa, B. ABC 1 (1st International consensus guidelines for advanced breast cancer): a positive step. *Breast* **21**, 225–226 (2012).

## **Annexe 2**

Article

# Increased Risk of Hospitalization for Pancreatic Cancer in the First 8 Years after a Gestational Diabetes Mellitus regardless of Subsequent Type 2 Diabetes: A Nationwide Population-Based Study

Julien Simon <sup>1,†</sup> , Karine Goueslard <sup>2,‡</sup>, Patrick Arveux <sup>3</sup> , Sonia Bechraoui-Quantin <sup>2</sup>, Jean-Michel Petit <sup>4,5,‡</sup> and Catherine Quantin <sup>1,2,6,7,\*</sup> 

- <sup>1</sup> High-Dimensional Biostatistics for Drug Safety and Genomics, Inserm U1018 Center of Research in Epidemiology and Population Health (CESP), Université Paris-Saclay, 94800 Villejuif, France; simonjulien@free.fr
  - <sup>2</sup> Biostatistics and Bioinformatics (DIM), University Hospital, 21079 Dijon, France; karine.goueslard@chu-dijon.fr (K.G.); s.bechraouiquantin@student.helmo.be (S.B.-Q.)
  - <sup>3</sup> Center for Primary Care and Public Health, Unisanté, University of Lausanne, 1003 Lausanne, Switzerland; Patrick.Arveux@unisanté.ch
  - <sup>4</sup> Department of Endocrinology, Diabetes and Metabolic Disorders, Dijon University Hospital, 21079 Dijon, France; jean-michel.petit@chu-dijon.fr
  - <sup>5</sup> INSERM Unit, LNC-UMR 1231, University of Burgundy, 21078 Dijon, France
  - <sup>6</sup> Clinical Epidemiology/Clinical Trials Unit, Clinical Investigation Center, Inserm, CIC 1432, Dijon University Hospital, 21000 Dijon, France
  - <sup>7</sup> CHU de Dijon, Service de Biostatistique et d'Informatique Médicale, BP 77908, CEDEX 21079 Dijon, France
- \* Correspondence: catherine.quantin@chu-dijon.fr; Tel.: +(33)-3-80-29-36-29; Fax: +(33)-3-80-29-39-73  
† These two authors contributed equally to this work.  
‡ These two authors contributed equally to this work.



**Citation:** Simon, J.; Goueslard, K.; Arveux, P.; Bechraoui-Quantin, S.; Petit, J.-M.; Quantin, C. Increased Risk of Hospitalization for Pancreatic Cancer in the First 8 Years after a Gestational Diabetes Mellitus regardless of Subsequent Type 2 Diabetes: A Nationwide Population-Based Study. *Cancers* **2021**, *13*, 308. <https://doi.org/10.3390/cancers13020308>

Received: 13 November 2020  
Accepted: 12 January 2021  
Published: 15 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Simple Summary:** Pancreatic cancer is one of the cancers with the highest mortality rate which is most often attributed to late diagnosis. The identification of risk factors is therefore important. While gestational diabetes also shares mechanisms of action with diabetes mellitus, several studies have provided hypotheses that could explain the pathophysiology of the link between diabetes mellitus and risk of pancreatic cancer. Accordingly, the aim of our study was to determine the risk of developing pancreatic cancer in women with a history of gestational diabetes from a quasi-exhaustive national medico-administrative database for deliveries in France. We included 1,368,755 women in our study. We showed that gestational diabetes was significantly associated with a greater risk of pancreatic cancer, regardless of subsequent type 2 diabetes. Our results suggest a better follow-up of patients after a gestational diabetes in order to identify high-risk profiles of developing more serious conditions, such as pancreatic cancer.

**Abstract:** The aim of this large retrospective cohort study was to use a quasi-exhaustive national medico-administrative database of deliveries in France to determine the risk of developing pancreatic cancer (PC) in women with a history of gestational diabetes mellitus (GDM). This nationwide population-based study included women aged 14–55 who gave birth between 1st January 2008 and 31 December 2009. The women were followed-up epidemiologically for eight years. Survival analyses using Cox regression models, adjusted for age, subsequent type 2 diabetes, and tobacco consumption, were performed on the time to occurrence of hospitalization for PC. The onset of GDM, tobacco consumption and subsequent type 2 diabetes were considered as time-dependent variables. Among 1,352,560 women included, 95,314 had a history of GDM (7.05%) and 126 women were hospitalized for PC (0.01%). Over the eight years of follow-up, GDM was significantly associated with a higher risk of hospitalization with PC in the first Cox regression model adjusted for age and subsequent type 2 diabetes (HR = 1.81 95% CI [1.06–3.10]). The second Cox regression model adjusted for the same covariates, plus tobacco consumption, showed that GDM was still significantly associated with a higher risk of hospitalization for PC with nearly the same estimated risk (HR = 1.77

95% CI [1.03–3.03]). Gestational diabetes was significantly associated with a greater risk of hospital admission for pancreatic cancer within eight years, regardless of subsequent type 2 diabetes.

**Keywords:** pancreas cancer; gestational diabetes mellitus; type 2 diabetes; postpartum follow-up; medico-administrative database; national study

## 1. Background

Pancreatic cancer (PC) is one of the cancers with the highest mortality rate in high-income countries, with a net survival of 6–9% at five years [1–3]. The high mortality rate is mainly attributed to a late diagnosis, because most patients with PC remain asymptomatic until the disease reaches an advanced stage. The identification of risk factors is therefore important for the creation of appropriate public health policies.

Many studies have reported that the risk of PC increases by approximately 80% in case of type 2 diabetes [4–8]. Several recent studies have provided hypotheses that could explain the pathophysiology of the link between diabetes mellitus and risk of PC [9–12]: hyperinsulinemia, hyperglycemia and low-grade inflammation. Although women with gestational diabetes mellitus (GDM) have an increased risk of diabetes mellitus over the years following delivery [13–15], GDM also shares mechanisms of action with type 2 diabetes, including hyperglycemia due to pancreatic  $\beta$ -cell dysfunction against a backdrop of chronic insulin resistance.

Although several studies have explored the association between GDM and cancer [16–19], the relationship between gestational diabetes and PC remains controversial due to limited evidence, because it has been the subject of only three studies, as follows. Two studies performed in Israel by Perrin et al. and Stella et al. (2007 and 2011) concluded that GDM increases the risk of subsequent PC, with an estimated risk around 7 [20,21]. A third study by Peng et al. based on the Taiwan National Health Insurance data (2019) showed that a history of GDM was not associated with subsequent PC [22]. A recent meta-analysis (2020), including the three previous studies, concluded that GDM was not associated with a risk of PC [23]. However, the results of the three studies should be taken with caution, because the diagnosis of GDM was not based on recent criteria in the two earlier studies and the follow-up was relatively short for some patients in the recent study. Diabetes could be considered as a symptom of pre-existing PC when its diagnosis is too close in time to the diagnosis of diabetes [24]. For this reason, we excluded PC occurring in the two years following a diagnosis of gestational diabetes. Accordingly, the aim of our study was to determine the risk of developing PC in women with a history of GDM. We used the quasi-exhaustive national medico-administrative database for deliveries in France over a 2-year inclusion period and a follow-up period of 8 years.

## 2. Methods

### 2.1. Data Sources and Study Design

We conducted a nationwide population-based retrospective cohort study of women who gave birth in France between 1st January 2008 and 31 December 2009. To this aim, we used the *Système National des Données de Santé* (SNDS), which is the French national health data system that contains individual, exhaustive and linkable but anonymous data on health expenditures for about 99% of the French population [25]. The SNDS pools data from the following sources: the database of the National Inter-Regime Information System on Health Insurance (*Système National d'Information Inter Régimes de l'Assurance Maladie*, SNIIRAM), which contains all prescriptions for drugs fully reimbursed by the national health insurance; the database of long-term diseases (LTD), which gives access to full coverage of health expenditures by the national health insurance scheme, and the statistical database on causes of death (*Base de Causes Médicales de Décès*, BCMD), which includes the illness or disease that caused the death, as well as other contributing factors.

The SNDS also includes data from the French national hospital database (Programme de Médicalisation des Systèmes d'Informations, PMSI), which collects principal and associated diagnoses (secondary events and current comorbidities) and procedures performed during hospital stays. The diagnoses are encoded using the International Classification of Diseases, 10th revision (ICD-10) and the procedures are encoded using the French common classification system for medical procedures (Classification Commune des Actes Médicaux, CCAM). The quality of the SNDS database is evaluated through yearly monitoring and has recently undergone validation for perinatal data [26].

Women aged 14–55 who gave birth between 1st January 2008 and 31st December 2009 were identified through the ICD-10 codes Z37 (“outcome of delivery”) as associated diagnosis and/or the procedures of deliveries in discharge abstracts from CCAM. The date of inclusion into the cohort was the date of the first delivery during the two-year inclusion period.

During the selection of study population, we excluded all women with a history of pancreatic cancer (codes C25 in the principal, associated or related diagnosis in the standardized hospital discharge abstract) during pregnancy and in the year before pregnancy. We also excluded women with a history of type 1 or 2 diabetes, who were identified using: the codes E10-E14 or O24.0-O24.3 as the principal, associated or related diagnosis in the standardized discharge abstract and/or if it was notified in the LTD and/or if at least three antidiabetic drugs or insulin (class A10 of the Anatomical, Therapeutic and Chemical classification [ATC]) or two in the event of at least one large quantity were reimbursed in the year preceding pregnancy [27]. Furthermore, women who had a diagnosis of PC within two years of pregnancy were excluded from the analysis population [24]. All of the women included for analysis were followed epidemiologically for 8 years.

## 2.2. Exposure

Exposure was defined as the presence of GDM during one pregnancy within the study period (Inclusion-2017). If a woman had more than one pregnancy, the information collected in the discharge abstracts for each pregnancy was taken into account. If a gestational diabetes code (O24.4 or O24.9) was identified in at least one of these pregnancies, the woman was considered to have GDM. To take into account the fact that GDM could occur at any time during follow-up (either during the first pregnancy or subsequent pregnancies), we considered it as a time-dependent variable (i.e., the exposure was only coded as 1 from the first-time GDM was recorded). This allowed us to consider more precisely the exact time interval between the diagnosis of GDM and the onset of PC. Women classified as “without GDM” during the entire study were those for whom a GDM code was not mentioned in any of the discharge abstracts during the epidemiological follow-up. In France, national guidelines for GDM diagnosis changed during the study period. Before 2010, GDM screening was based on a two-step procedure for all women: the first test was on venous blood glucose 1 h after ingestion of 50 g of glucose and in the event of a positive result, the second screening test was performed for oral glucose tolerance. Since 2010, new national guidelines were introduced for women with some risk factors (i.e., maternal age  $\geq 35$  years, BMI  $\geq 25$ , history of diabetes in the 1st degree, personal history of GDM or macrosomia). The recommended screening for these women is fasting blood glucose at the first prenatal consultation, and if it is not performed, an oral glucose tolerance test in the second trimester.

## 2.3. Outcome

The outcome of interest was hospitalization with an ICD-10 code for PC (C25) as a principal or associated diagnosis during 8-year follow-up period. In an effort to minimize the effect of possible reverse causality on our results, we excluded women who were diagnosed with PC within two years of pregnancy [24].

#### 2.4. Variables

The following explanatory variables were considered: maternal age, subsequent type 2 diabetes and tobacco consumption. The maternal age at the index delivery was categorized as <26, 26–29, 30–33 and  $\geq 34$  years. Subsequent type 2 diabetes during follow-up was defined by: the presence of one of the ICD-10 codes E10 to E14 as principal, associated or related diagnoses and/or if subsequent diabetes was notified in LTD and/or if there was a reimbursement of at least three antidiabetic drugs or insulin (class A10 of the Anatomical, Therapeutic and Chemical classification (ATC) with the exception of benfluorex) or two reimbursements if there was at least one large packaging [27]. Tobacco consumption in the year before the pregnancy, at inclusion or during follow-up was identified using the ICD-10 codes F17, T65.2, Z58.7, Z72.0, or Z71.6 in hospitalization records and/or at least one reimbursement of varenicline (Identification Codes of pharmaceutical speciality).

The time duration (days) from the delivery (with or without GDM) to the onset of the PC was determined.

Tobacco consumption was determined in two ways, as a time-dependent variable (i.e., onset) and a fixed variable (presence). The onset of tobacco consumption took into account the time interval from the moment smoking was detected (during hospitalization or following smoking cessation) in our database. The presence of tobacco consumption was determined as any smoking from the time of inclusion, regardless of when smoking was recorded in the year before the pregnancy and during the follow-up period.

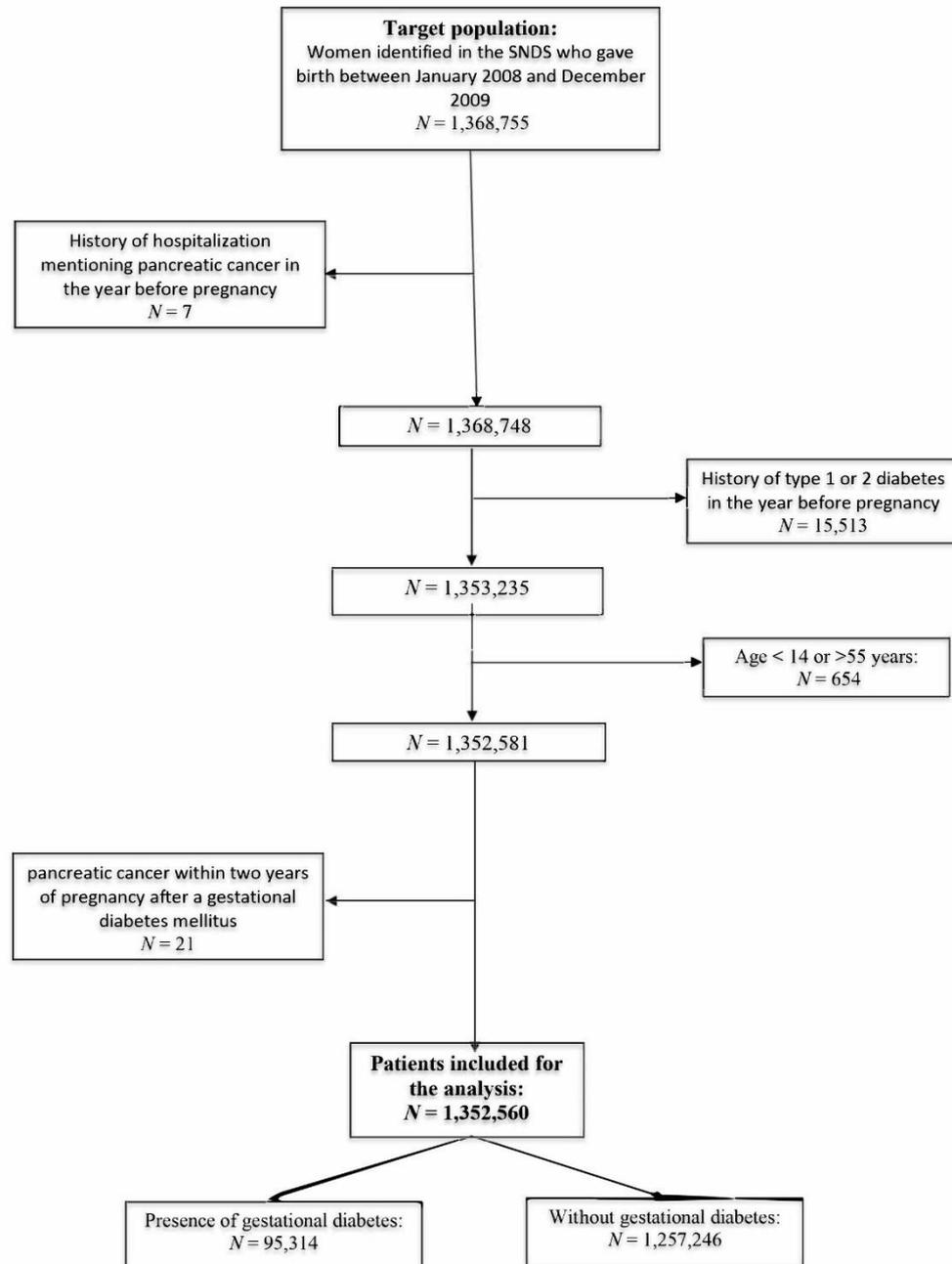
#### 2.5. Statistical Analyses

The characteristics of women that are categorical variables were described by the GDM groups (with GDM history and without GDM history) using frequency distributions. The two groups were compared using  $\chi^2$  tests. Continuous variables, notably time durations, were described using mean (with standard deviation). The risk of hospitalization with PC was investigated using Cox proportional hazards regression models. The onset of GDM, tobacco consumption and hospitalization with subsequent type 2 diabetes were considered as time-dependent variables. We developed a first model including age, GDM, and subsequent type 2 diabetes, and a second model that adjusted for the same covariates, plus onset of tobacco consumption as a time-dependent variable. This approach may be considered questionable, as a patient may have started smoking before being identified as a smoker (by being mentioned during hospitalization or by being given a smoking cessation medication). Furthermore, a sensitivity analysis was performed on the second model, where tobacco consumption onset was replaced with the presence of tobacco consumption as a fixed variable. Hazard ratios (HR) and 95% confidence intervals (CI) were calculated in each model. We also explored the interactions between age or tobacco consumption and GDM. Finally, we ran an analysis assessing the E-value [28], which is a measure of an association's robustness to potential uncontrolled confounders (e.g., obesity). Computer software SAS 9.4 was used for the descriptive and multivariate analyses.

This study was approved by the French national committee for data protection (registration number 12/11/2019-DR-2019-350). Individual written consent was not needed for this study. Data from the SNDS database was provided by the French National Health Insurance Fund (21/02/2019 n° 226847).

### 3. Results

From January 2008 to December 2009, 1,368,755 women hospitalized for delivery were identified in the SNDS database. Seven women who were hospitalized for PC before pregnancy and 21 women who were hospitalized for PC within 2 years after pregnancy were excluded (4 with gestational diabetes and 17 without) (Figure 1).



**Figure 1.** Flowchart of the study population.

Similarly, 15,531 women with type 2 diabetes and 654 women who did not meet the age criteria were also excluded, leading to 1,352,560 women who were included in the analysis population. The mean duration time from the delivery to the onset of PC over all included patients was 1910 days (sd = 604). We also found that the mean duration time

from the delivery of the 1st pregnancy with GDM (to compute this mean duration from the exposure with GDM) to the onset of PC was 1616 days (sd = 715) in the group with GDM and 1923 days (sd = 614) in the group without GDM.

Among the 1,352,560 women included in our study, 95,314 had GDM (7.05%) during follow-up and 126 women were hospitalized for PC (0.01%). A comparison of the characteristics of women with and without GDM is presented in Table 1.

**Table 1.** Comparison of characteristics of women with history of GDM or without history of GDM.

Variables	GDM+		GDM−		p
	N = 95,314 (7.05%)		N = 1,257,246 (92.95%)		
Age * (years)					<0.0001
<26	18,095	19.0%	299,873	23.9%	
26–29	26,013	27.3%	352,910	28.1%	
30–33	24,279	25.5%	313,954	25.0%	
≥34	26,927	28.2%	290,509	23.1%	
Hospitalization with Pancreatic Cancer	18	0.02%	108	0.01%	0.004
Subsequent type 2 diabetes	8184	8.6%	6840	0.5%	<0.0001
Tobacco consumption **	7818	8.2%	86,403	6.9%	<0.0001

N, number; GDM, Gestational Diabetes Mellitus; \* Age between 14 and 55 years. \*\* Presence of tobacco consumption at any time, as used in the sensitivity analysis.

The results of the first Cox model, which is adjusted for age and subsequent type 2 diabetes (as a time-dependent covariate), showed that GDM (as a time-dependent variable) was significantly associated with a higher risk of hospitalization with PC (HR = 1.81, 95% CI [1.06–3.10]). The results of the second Cox model adjusted for the same covariates, plus tobacco consumption (as a time-dependent covariate), showed that GDM was still significantly associated with a higher risk of hospitalization with PC with nearly the same estimated relative risk (HR = 1.77, 95% CI [1.03–3.03]). We found that hospitalization with subsequent type 2 diabetes or tobacco consumption (considered as a time-dependent variable or as a fixed variable in the sensitivity analysis) did not alter the relationship between GDM and hospitalization with PC. The interaction between tobacco consumption and GDM was non-significant ( $p = 0.60$ ) but the interaction between age and GDM was marginally significant ( $p = 0.05$ ). After including this interaction in Model 1, the effect of GDM was even higher. In the analysis assessing the E-value, we found that the observed HR = 1.77 could be explained away by a minimum strength of association that an unmeasured confounder must have with gestational diabetes exposure and pancreatic cancer, as low as 2.93 (Table 2).

**Table 2.** Cox models of the relation between GDM and hospitalization with pancreatic cancer (n = 126). GDM, subsequent type 2 diabetes and tobacco consumption are time-dependent covariates.

Variables	Non-Adjusted		Cox Model 1 *		Cox Model 2 **	
	HR	95% CI	HR	95% CI	HR	95% CI
GDM	2.59	1.57–4.26	1.81	1.06–3.10	1.77	1.03–3.03
Age ***						
<26	0.60	0.28–1.27	0.61	0.28–1.30	0.56	0.26–1.21
26–29	1		1		1	
30–33	1.85	1.06–3.22	1.83	1.05–3.19	1.88	1.08–3.27
≥34	3.77	2.28–6.23	3.60	2.18–5.96	3.67	2.22–6.08
Type 2 diabetes	8.06	3.76–17.32	4.85	2.13–11.06	4.74	2.08–10.79
Tobacco consumption	2.8	1.72–4.56	-	-	3.22	1.97–5.26

GDM: Gestational Diabetes Mellitus; HR: Hazard Ratio. \* Model 1 includes GDM, age and type 2 diabetes; \*\* Model 2 includes GDM, age, type 2 diabetes and tobacco consumption; \*\*\* Age between 14 and 55 years.

#### 4. Discussion

To our knowledge, this is the first study that specifically explores the relationship between GDM and PC, after adjustment for subsequent diabetes, which is the main potential confounding factor. In our population-based study, using a nationwide well documented retrospective cohort, we found that GDM was significantly associated with a greater risk of being hospitalized for PC within the first eight years, regardless of whether diabetes mellitus developed subsequently.

In the literature, the three studies on this subject report variable methodologies and divergent results. The two Israeli studies show similar significant relative risks, around 7, but with wide confidence intervals [20,21]. In addition, the Israeli studies were hospital-based, including patients from three hospitals in Perrin et al. and only one in Sella et al. They also found a limited number of cases: fifty-four in Perrin et al. and nine in Sella et al., compared to our study which identified 126 cases among the 1,352,560 included women. The larger population in our study explains the narrower confidence interval. In addition, our smaller relative risk may be explained by an adjustment for subsequent diabetes. Our inclusion period (2008–2009) is much more recent than in Perrin et al. (1964–1975), which was not based on recent diagnostic criteria for GDM. This may have impacted their results by creating a classification bias when selecting only the most severely exposed women. On the other hand, the Taiwanese study of Peng et al. is more comparable to our study, since it is also a population-based study using a national medico-administrative data [22]. They included nearly the same number of women (1,466,596 women) but observed a larger number of pancreatic cancer cases (331 cases). Though they found twice as many cases, they were not able to show a significant increased risk of PC in women with a history of GDM (HR = 1.07, 95% CI 0.65–1.75). However, they took into account all PCs occurring after pregnancy while we waited until two years after the occurrence of GDM to record a diagnosis of PC. In addition, it has been shown that Asian populations develop diabetes at a younger age, require insulin treatment more often, and see an increased risk of cancer compared to Western populations [29,30], which may explain the higher number of PC cases. Moreover, unlike our study, the follow-up in the Taiwanese study varied between women, some of whom were followed for more than twelve years while others were followed for only one year. Finally, we included GDM as a time-dependent variable, which might explain why we found a statistically significant result despite having fewer cases.

The first strength of our study is our adjustment for type 2 diabetes. To our knowledge, our study is unique in that it adjusts for this major potential confounding factor. Several studies have shown that there is a relationship between GDM and diabetes mellitus, and type 2 diabetes is recognized as directly or indirectly increasing the risk of PC through the degeneration of intraductal papillary mucinous neoplasms [31]. Therefore, it seemed important to consider from various aspects this confounding factor in the methodology of our study, as follows. Firstly, regarding our database, a validation study in a French cohort by Fuentes et al. showed that type 2 diabetes was identified in the national database (SNDS) with a positive predictive value of 97.2%, a sensitivity of 93.8%, and a specificity of 99.9% [32]. Secondly, we excluded women with a history of type 2 diabetes in order to refine the definition of GDM and remove the impact of the history of diabetes mellitus on the analysis. Finally, diabetes mellitus occurring during follow-up was included in our models as a time-dependent variable. By using Cox regression models, it was possible to introduce explanatory variables known at the time of inclusion but also information collected during the follow-up. This allowed us to take into account a change in status, for example, from non-diabetic to diabetic, but also the time of exposure to diabetes before hospitalization for PC.

Time is also an important consideration for the study of PC. Firstly, regarding the chronology of the disease itself, some studies show that hospitalization is a good marker for new cases for some cancers that are aggressive and/or requiring hospitalization [33,34]. Knowing that we took into account all types of hospitalizations (including diagnosis, radiotherapy, chemotherapy, surgery and follow-up), we were able to exclude subjects with

a history of hospitalization for PC one year before-, during- and two years after pregnancy (i.e., almost four years). It is known that PC has a net five-year survival rate of 6–9% [1–3]. By eliminating women with signs of PC in the first four years, we ensured in this study that there was a very strong probability of including only incident cases of PC. Secondly, to our knowledge, ours is the only study that has established a two-year lapse of time between pregnancy (with or without GDM) and the diagnosis of PC. It has been shown that case selection should not include newly diagnosed diabetes, because it is known that a diagnosis of diabetes can be a symptom of PC. Indeed, new-onset diabetes caused by PC may be associated with pro-inflammatory alterations, insulin resistance, and perturbations in  $\beta$ -cell functions that lead to loss of glucose homeostasis. Adipocytes and resident macrophages release many deleterious pro-inflammatory factors, and the increase in lipolysis could enhance hepatic production of acetyl coenzyme-A, causing hepatic insulin resistance and excessive glucose release by the liver [24].

Our study has other strengths. First of all, our population-based study comprised 10 years of data, allowing us to have eight years of follow-up for all included women. It relied on a quasi-exhaustive national database of deliveries (99.6%), because almost all childbirths in France take place in hospitals and all hospital stays are recorded in this database. In addition, in a national validation study [35], our team showed that GDM was identified in this database with a positive predictive value of 88.9%.

We also recognize that our study has some limitations. First, our recent validation study showed that sensitivity for GDM was only of 79%, which means that 21% of women with GDM were not identified as such in our database [26]. As a consequence, this classification bias may have led to an underestimation of the association. In addition, national guidelines for the diagnosis of GDM have changed from a two-step procedure in all patients before 2010, to a different (only one-step) procedure in patients with risk factors since 2010. Second, although we have a longer and more homogenous follow-up time compared to the three previous studies, eight years remains relatively short for the study of the onset of PC, most of which are diagnosed after 65 years, and we may thus have underestimated the full impact of GDM. Third, we did not have completed data on gestation and parity in our national database for the years under review. Fourth, although we included “tobacco consumption” as an adjustment variable, this information is incomplete in our administrative database. This is why we carried out two multivariate models, one with this variable and the other without; both of them provided similar results. Fifth, some other known risk factors could not be included in our study. Although we wanted to add “alcohol consumption” to our adjustment variables, since it is another well-known risk factor for pancreatic cancer, it was not feasible to do so because this variable is too poorly recorded in our database [36]. Moreover, the “obesity” variable, which is known to be associated with gestational diabetes and pancreatic cancer, could not be included in our model. Unfortunately, BMI before pregnancy is poorly documented in our hospital database and BMI during pregnancy may not be considered as reliable. This is why we conducted an analysis of the E-value, which showed that the observed HR (1.77) could be explained away by a minimum strength of association that an unmeasured confounder must have with gestational diabetes exposure and pancreatic cancer, as low as 2.93. The E-value is thus large and substantial unmeasured confounding would be necessary to cancel the observed association. In the literature, we found that obese women are 3.36 times more likely than non-obese women to have gestational diabetes [37] and 1.54 times more likely to have pancreatic cancer [38]. As both association levels are very different, we further calculated the bias B, or the largest factor by which the observed association could be reduced by the unmeasured confounder of these particular strengths. The B value calculated from these values should be higher than our HR of 1.77 to be explained by the unmeasured “obesity” variable, but it is equal to 1.33 and therefore lower than 1.77 [28]. This strengthens the conclusion derived from the E-value analysis. Sixth, additional studies with longer follow-up or including more confounding factors are needed

to support these findings. In the same way, further studies are needed to investigate the relationship between GDM and survival in PC patients.

Gestational diabetes is still considered by some professionals to be a disease that has little impact once pregnancy is over, resulting in poor quality short- and long-term postpartum follow-up [39–42]. However, pregnancy is a kind of “stress-test” for the body of women, which reveals some patients’ vulnerabilities. For instance, GDM seems to be an important marker for diseases that may occur later in life. Indeed, we found in a previous study that gestational diabetes exposes women to a subsequent increased risk of type 2 diabetes mellitus, metabolic syndrome and cardiovascular conditions [43]. The results of this study suggest that an increased risk of PC should be added to this list. All these results are in favor of a better adherence to the recommendations [44,45] for following-up patients with GDM in the postpartum period. In the US, like in France, screening for type 2 diabetes is recommended during the postnatal period, and then every 1–3 years, depending on the risk factors, for at least 25 years. After GDM, health professionals should encourage lifestyle changes (30 to 60 min of daily physical activity at least five days per week, balanced diet, smoking cessation . . . ) which are also useful for the prevention of PC [46,47]. Monitoring should also include regular screening and treatment of any other associated cardiovascular risk factors (arterial hypertension, dyslipidemia).

## 5. Conclusions

In conclusion, our study shows that gestational diabetes mellitus is a factor associated with the development of pancreatic cancer within eight years after delivery. However, the impact of gestational diabetes is overlooked by some professionals even if it exposes women to an increased risk of type 2 diabetes mellitus, metabolic syndrome and cardiovascular diseases. Our results suggest the importance of better adherence to the recommendations concerning the follow-up of patients with gestational diabetes in the postpartum period. To be effective, this follow-up would require the participation of all health professionals, including general practitioners, gynecologists, obstetricians, midwives, and endocrinologists. To render this follow-up feasible, gestational diabetes could be included among other risk factors (e.g., diabetes mellitus, obesity, metabolic syndrome, smoking or alcohol consumption) in order to identify individuals with a high risk of developing more serious conditions such as pancreatic cancer. These individuals can thus be encouraged to attend regular screening and advised about appropriate lifestyle and dietary measures.

**Author Contributions:** J.S., K.G. and C.Q. conceptualized and designed the study, interpreted data, and wrote the manuscript; J.-M.P., P.A. and S.B.-Q. participated in the interpretation of the results and reviewed and revised the manuscript drafts; C.Q. and K.G. oversaw the data analysis and interpretation and contributed substantially to writing the manuscript; J.S., K.G., J.-M.P., S.B.-Q., P.A. and C.Q. accept responsibility for the paper as published; C.Q. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the French national committee for data protection (registration number 12/11/2019-DR-2019-350).

**Informed Consent Statement:** Patient consent was waived as written consent was not required for this study because it was a retrospective study and the national data were anonymous. As a consequence, the research involves no risk to the subjects; the waiver or alteration will not adversely affect the rights and welfare of the subjects; the research could not practicably be carried out without the waiver or alteration (i.e., it is impracticable to get the consent of the subjects).

**Data Availability Statement:** Data used in this study are available for researchers who meet the criteria for access to these French data from the National Health Insurance Fund (CNAM—Caisse nationale de l'assurance maladie).

**Acknowledgments:** The authors thank Suzanne Rankin and Harbajan Chadha-Boreham for reviewing the English. Catherine Quantin who takes full responsibility for the work as a whole, including the study design, access to data, and the decision to submit and publish the manuscript.

**Conflicts of Interest:** The authors have no conflict of interest relevant to this article to disclose.

### Abbreviations

gestational diabetes mellitus (GDM); Pancreatic cancer (PC); Système National des Données de Santé (SNDS); Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIRAM); long-term diseases (LTD); Base de Causes Médicales de Décès (BCMD); Programme de Médicalisation des Systèmes d'Informations (PMSI); International Classification of Diseases, 10th revision (ICD-10); Anatomical Therapeutic and Chemical classification (ATC); Hazard Ratios (HR); Confidence Intervals (CI).

### References

1. Drouillard, A.; Manfredi, S.; Lepage, C.; Bouvier, A.-M. Epidemiology of pancreatic cancer. *Bull. Cancer JANV* **2018**, *105*, 63–69. [CrossRef] [PubMed]
2. Available online: <https://seer.cancer.gov/statfacts/html/pancreas.html> (accessed on 13 November 2020).
3. Rawla, P.; Sunkara, T.; Gaduputi, V. Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World J. Oncol.* **2019**, *10*, 10–27. [CrossRef] [PubMed]
4. Huxley, R.; Ansarymoghaddam, A.; De González, A.B.; Barzi, F.; Woodward, M.J. Type-II diabetes and pancreatic cancer: A meta-analysis of 36 studies. *Br. J. Cancer* **2005**, *92*, 2076–2083. [CrossRef] [PubMed]
5. Elena, J.W.; Steplowski, E.; Yu, K.; Hartge, P.; Tobias, G.S.; Brotzman, M.J.; Chanock, S.J.; Stolzenberg-Solomon, R.Z.; Arslan, A.A.; Bueno-De-Mesquita, H.B.; et al. Diabetes and risk of pancreatic cancer: A pooled analysis from the pancreatic cancer cohort consortium. *Cancer Causes Control* **2013**, *24*, 13–25. [CrossRef] [PubMed]
6. Chodick, G.; Heymann, A.D.; Rosenmann, L.; Green, M.S.; Flash, S.; Porath, A.; Kokia, E.; Shalev, V. Diabetes and risk of incident cancer: A large population-based cohort study in Israel. *Cancer Causes Control* **2010**, *21*, 879–887. [CrossRef]
7. Everhart, J.; Wright, D. Diabetes mellitus as a risk factor for pancreatic cancer: A meta-analysis. *JAMA* **1995**, *273*, 9. [CrossRef]
8. Zhou, X.H.; for the DECODE Study Group; Qiao, Q.; Zethelius, B.; Pyörälä, K.; Söderberg, S.; Pajak, A.; Stehouwer, C.D.A.; Heine, R.J.; Jousilahti, P.; et al. Diabetes, prediabetes and cancer mortality. *Diabetologia* **2010**, *53*, 1867–1876. [CrossRef]
9. Biadgo, B.; Abebe, M. Type 2 Diabetes Mellitus and Its Association with the Risk of Pancreatic Carcinogenesis: A Review. *Korean J. Gastroenterol.* **2016**, *67*, 168–177. [CrossRef]
10. Rahn, S.; Zimmermann, V.; Viol, F.; Knaack, H.; Stemmer, K.; Peters, L.; Lenk, L.; Ungefroren, H.; Saur, D.; Schäfer, H.; et al. Diabetes as risk factor for pancreatic cancer: Hyperglycemia promotes epithelial-mesenchymal-transition and stem cell properties in pancreatic ductal epithelial cells. *Cancer Lett.* **2018**, *415*, 129–150. [CrossRef]
11. Eibl, G.; Cruz-Monserrate, Z.; Korc, M.; Petrov, M.S.; Goodarzi, M.O.; Fisher, W.E.; Habtezion, A.; Lugea, A.P.; Pandol, S.J.; Hart, P.A.; et al. Diabetes Mellitus and Obesity as Risk Factors for Pancreatic Cancer. *J. Acad. Nutr. Diet.* **2018**, *118*, 555–567. [CrossRef]
12. Paternoster, S.; Falasca, M. The intricate relationship between diabetes, obesity and pancreatic cancer. *Biochim. Biophys. Acta (BBA) Bioenerg.* **2020**, *1873*, 188326. [CrossRef] [PubMed]
13. Bellamy, L.; Casas, J.-P.; Hingorani, A.D.; Williams, D. Type 2 diabetes mellitus after gestational diabetes: A systematic review and meta-analysis. *Lancet* **2009**, *373*, 1773–1779. [CrossRef]
14. Kaul, P.; Savu, A.; Nerenberg, K.A.; Donovan, L.E.; Chik, C.L.; Ryan, E.A.; Johnson, J.A. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: A population-level analysis. *Diabet. Med.* **2015**, *32*, 164–173. [CrossRef] [PubMed]
15. Retnakaran, R.; Shah, B.R. Role of Type 2 Diabetes in Determining Retinal, Renal, and Cardiovascular Outcomes in Women With Previous Gestational Diabetes Mellitus. *Diabetes Care* **2016**, *40*, 101–108. [CrossRef]
16. Dawson, S.I. Long-term risk of malignant neoplasm associated with gestational glucose intolerance. *Cancer* **2003**, *100*, 149–155. [CrossRef]
17. Fuchs, O.; Sheiner, E.; Meirovitz, M.; Davidson, E.; Sergienko, R.; Kessous, R. The association between a history of gestational diabetes mellitus and future risk for female malignancies. *Arch. Gynecol. Obstet.* **2017**, *295*, 731–736. [CrossRef]
18. Trabert, B.; Troisi, R.; Grotmol, T.; Ekbo, A.; Engeland, A.; Gissler, M.; Glimelius, I.; Madanat-Harjuoja, L.; Sørensen, H.T.; Tretli, S.; et al. Associations of pregnancy-related factors and birth characteristics with risk of endometrial cancer: A Nordic population-based case-control study. *Int. J. Cancer* **2020**, *146*, 1523–1531. [CrossRef]

19. Tong, G.-X.; Cheng, J.; Chai, J.; Geng, Q.-Q.; Chen, P.-L.; Shen, X.-R.; Liang, H.; Wang, D. Association Between Gestational Diabetes Mellitus and Subsequent Risk of Cancer: A Systematic Review of Epidemiological Studies. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 4265–4269. [[CrossRef](#)]
20. Perrin, M.C.; Terry, M.B.; Kleinhaus, K.R.; Deutsch, L.; Yanetz, R.; Tiram, E.; Calderon-Margalit, R.; Friedlander, Y.; Paltiel, O.; Harlap, S. Gestational diabetes as a risk factor for pancreatic cancer: A prospective cohort study. *BMC Med.* **2007**, *5*, 25. [[CrossRef](#)]
21. Sella, T.; Chodick, G.; Barchana, M.; Heymann, A.D.; Porath, A.; Kokia, E.; Shalev, V. Gestational diabetes and risk of incident primary cancer: A large historical cohort study in Israel. *Cancer Causes Control* **2011**, *22*, 1513–1520. [[CrossRef](#)]
22. Peng, Y.-S.; Lin, J.-R.; Cheng, B.-H.; Ho, C.; Lin, Y.-H.; Shen, C.-H.; Tsai, M.-H. Incidence and relative risk for developing cancers in women with gestational diabetes mellitus: A nationwide cohort study in Taiwan. *BMJ Open* **2019**, *9*, e024583. [[CrossRef](#)] [[PubMed](#)]
23. Wang, Y.; Yan, P.; Fu, T.; Yuan, J.; Yang, G.; Liu, Y.; Zhang, Z.-J. The association between gestational diabetes mellitus and cancer in women: A systematic review and meta-analysis of observational studies. *Diabetes Metab.* **2020**, *46*, 461–471. [[CrossRef](#)] [[PubMed](#)]
24. Pannala, R.; Leirness, J.B.; Bamlet, W.R.; Basu, A.; Petersen, G.M.; Chari, S.T. Prevalence and Clinical Profile of Pancreatic Cancer–Associated Diabetes Mellitus. *Gastroenterology* **2008**, *134*, 981–987. [[CrossRef](#)] [[PubMed](#)]
25. Tuppin, P.; Rudant, J.; Constantinou, P.; Gastaldi-Ménager, C.; Rachas, A.; de Roquefeuil, L.; Maura, G.; Caillol, H.; Tajahmady, A.; Coste, J.; et al. Value of a national administrative database to guide public decisions: From the système national d’information interrégimes de l’Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev. Epidemiol. Sante Publique* **2017**, *65* (Suppl. 4), S149–S167. [[CrossRef](#)]
26. Goueslard, K.; Cottenet, J.; Benzenine, E.; Tubert-Bitter, P.; Quantin, C. Validation study: Evaluation of the metrological quality of French hospital data for perinatal algorithms. *BMJ Open* **2020**, *10*, e035218. [[CrossRef](#)]
27. Fosse-Edorh, S.; Rigou, A.; Morin, S.; Fezeu, L.; Mandereau-Bruno, L.; Fagot-Campagna, A. Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes. *Rev. Epidemiol. Sante Publique* **2017**, *65* (Suppl. 4), S168–S173. [[CrossRef](#)]
28. VanderWeele, T.J.; Ding, P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann. Intern. Med.* **2017**, *167*, 268–274. [[CrossRef](#)]
29. Ma, R.C.; Chan, J.C. Type 2 diabetes in East Asians: Similarities and differences with populations in Europe and the United States. *Ann. N. Y. Acad. Sci.* **2013**, *1281*, 64–91. [[CrossRef](#)]
30. Noto, H.; Tsujimoto, T.; Noda, M. Significantly increased risk of cancer in diabetes mellitus patients: A meta-analysis of epidemiological evidence in Asians and non-Asians. *J. Diabetes Investig.* **2011**, *3*, 24–33. [[CrossRef](#)]
31. Pergolini, I.; Schorn, S.; Jäger, C.; Goess, R.; Novotny, A.; Friess, H.; Ceyhan, G.O.; Demir, I.E. Diabetes mellitus in intraductal papillary mucinous neoplasms: A systematic review and meta-analysis. *Surgery* **2020**. [[CrossRef](#)]
32. Fuentes, S.; CONSTANCES-Diab Group; Cosson, E.; Mandereau-Bruno, L.; Fagot-Campagna, A.; Bernillon, P.; Goldberg, M.; Fosse-Edorh, S. Identifying diabetes cases in health administrative databases: A validation study based on a large French cohort. *Int. J. Public Health* **2018**, *64*, 441–450. [[CrossRef](#)] [[PubMed](#)]
33. Bousquet, P.-J.; Caillet, P.; Coeuret-Pellicer, M.; Goulard, H.; Kudjawu, Y.C.; Le Bihan, C.; Lecuyer, A.; Séguret, F. 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* **2017**, *65* (Suppl. 4), S236–S242. [[CrossRef](#)] [[PubMed](#)]
34. Wu, J.W.; Azoulay, L.; Huang, A.; Paterson, M.; Wu, F.; Secrest, M.H.; Fillion, K.B. Identification of incident pancreatic cancer in Ontario administrative health data: A validation study. *Pharmacoepidemiol. Drug Saf.* **2020**, *29*, 78–85. [[CrossRef](#)] [[PubMed](#)]
35. Pierron, A.; Revert, M.; Goueslard, K.; Vuagnat, A.; Cottenet, J.; Benzenine, E.; Fresson, J.; Quantin, C. Evaluation of the metrological quality of the medico-administrative data for perinatal indicators: A pilot study in 3 university hospitals]. *Rev. Epidemiol. Sante Publique* **2015**, *63*, 237–246. [[CrossRef](#)]
36. Yadav, D.; Lowenfels, A.B. The Epidemiology of Pancreatitis and Pancreatic Cancer. *Gastroenterology* **2013**, *144*, 1252–1261. [[CrossRef](#)]
37. Torloni, M.R.; Betrán, A.P.; Horta, B.L.; Nakamura, M.U.; Atallah, A.N.; Moron, A.F.; Valente, O. Prepregnancy BMI and the risk of gestational diabetes: A systematic review of the literature with meta-analysis. *Obes. Rev.* **2009**, *10*, 194–203. [[CrossRef](#)]
38. Johansen, D.; Stocks, T.; Jonsson, H.; Lindkvist, B.; Björge, T.; Concin, H.; Almquist, M.; Häggström, C.; Engeland, A.; Ulmer, H.; et al. Metabolic Factors and the Risk of Pancreatic Cancer: A Prospective Analysis of almost 580,000 Men and Women in the Metabolic Syndrome and Cancer Project. *Cancer Epidemiol. Biomark. Prev.* **2010**, *19*, 2307–2317. [[CrossRef](#)]
39. Goueslard, K.; Cottenet, J.; Mariet, A.-S.; Sagot, P.; Petit, J.-M.; Quantin, C. Early screening for type 2 diabetes following gestational diabetes mellitus in France: Hardly any impact of the 2010 guidelines. *Acta Diabetol.* **2017**, *54*, 645–651. [[CrossRef](#)]
40. Quaresima, P.; Visconti, F.; Chiefari, E.; Puccio, L.; Foti, D.; Venturella, R.; Vero, R.; Brunetti, A.; Di Carlo, C. Barriers to Postpartum Glucose Intolerance Screening in an Italian Population. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2853. [[CrossRef](#)]
41. Boyle, D.I.; Versace, V.L.; Dunbar, J.; Scheil, W.; Janus, E.; Oats, J.J.N.; Skinner, T.C.; Shih, S.; O’Reilly, S.; Sikaris, K.; et al. Results of the first recorded evaluation of a national gestational diabetes mellitus register: Challenges in screening, registration, and follow-up for diabetes risk. *PLoS ONE* **2018**, *13*, e0200832. [[CrossRef](#)]
42. Bernstein, J.A.; Quinn, E.; Ameli, O.; Craig, M.; Heeren, T.; Iverson, R.; Jack, B.; Lee-Parritz, A.; McCloskey, L. Onset of T2DM after gestational diabetes: What the prevention paradox tells us about risk. *Prev. Med.* **2018**, *113*, 1–6. [[CrossRef](#)] [[PubMed](#)]

43. Goueslard, K.; Cottenet, J.; Mariet, A.-S.; Giroud, M.; Cottin, Y.; Petit, J.-M.; Quantin, C. Early cardiovascular events in women with a history of gestational diabetes mellitus. *Cardiovasc. Diabetol.* **2016**, *15*, 1–7. [[CrossRef](#)] [[PubMed](#)]
44. Collège National des Gynécologues et Obstétriciens Français, Société Francophone du Diabète. Gestational diabetes. *J. Gynecol. Obstet. Biol. Reprod* **2010**, *39* (Suppl. 2), S139, S338–S342.
45. American Diabetes Association. 14. Management of Diabetes in Pregnancy: Standards of Medical Care in Diabetes—2020. *Diabetes Care* **2020**, *43* (Suppl. 1), S183–S192. [[CrossRef](#)]
46. Capasso, M.; Franceschi, M.; Rodriguez-Castro, K.L.; Crafa, P.; Cambiè, G.; Miraglia, C.; Barchi, A.; Nouvenne, A.; Leandro, G.; Meschi, T.; et al. Epidemiology and risk factors of pancreatic cancer. *Acta Biomed.* **2018**, *89* (Suppl. 9), 141–146.
47. Midha, S.; Chawla, S.; Garg, P.K. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. *Cancer Lett.* **2016**, *381*, 269–277. [[CrossRef](#)]

## **Annexe 3**

Systematic Review

# Is Gestational Diabetes Mellitus a Risk Factor of Maternal Breast Cancer? A Systematic Review of the Literature

Julien Simon <sup>1</sup> , Karine Goueslard <sup>1</sup>, Sonia Bechraoui-Quantin <sup>1</sup>, Patrick Arveux <sup>2</sup> and Catherine Quantin <sup>1,3,4,\*</sup> 

<sup>1</sup> Biostatistics and Bioinformatics (DIM), University Hospital, 21000 Dijon, France; simonjulien@free.fr (J.S.); karine.goueslard@chu-dijon.fr (K.G.); s.bechraouiquantin@student.helmo.be (S.B.-Q.)

<sup>2</sup> Center for Primary Care and Public Health, Unisanté, University of Lausanne, 1010 Lausanne, Switzerland; Patrick.Arveux@unisanté.ch

<sup>3</sup> Inserm, CIC 1432, Clinical Investigation Center, Clinical Epidemiology/Clinical Trials Unit, Dijon University Hospital, 21000 Dijon, France

<sup>4</sup> High-Dimensional Biostatistics for Drug Safety and Genomics, Université Paris-Saclay, UVSQ, University Paris-Sud, Inserm, CESP, 94800 Villejuif, France

\* Correspondence: catherine.quantin@chu-dijon.fr; Tel.: +333-80-29-34-65

**Abstract:** The association between gestational diabetes mellitus (GDM) and breast cancer (BC) risk is complex. We aimed to examine this association in a systematic review of the literature. This review was done using the PubMed/Medline and Web of Science databases, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The Newcastle–Ottawa Scale was used for the assessment of bias and quality of studies. Only English-language articles published before 1 June 2021, were included. Fourteen studies were included in this systematic review. Among them, eight did not find statistically significant results. Three studies showed a statistically significant increased risk of BC after GDM, and they explained this potential increased risk by hyperinsulinemia, hyperglycemia, and low-grade inflammation. However, three studies showed a statistically significant decreased risk of BC after GDM, suggesting a possible protective effect of hormonal changes induced by GDM during pregnancy. These controversial results should be interpreted with caution due to both quantitative and qualitative methodological shortcomings. Further investigations are thus needed in order to gain a better understanding of the associations between GDM and BC, and their underlying mechanisms.

**Keywords:** systematic review; gestational diabetes mellitus; breast cancer; risk factor; qualitative analysis



**Citation:** Simon, J.; Goueslard, K.; Bechraoui-Quantin, S.; Arveux, P.; Quantin, C. Is Gestational Diabetes Mellitus a Risk Factor of Maternal Breast Cancer? A Systematic Review of the Literature. *Biomedicines* **2021**, *9*, 1174. <https://doi.org/10.3390/biomedicines9091174>

Academic Editor: Vittoria D'Esposito

Received: 21 July 2021

Accepted: 1 September 2021

Published: 7 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Breast cancer (BC) is the most common cancer, accounting for about one fourth of all cancer cases in women in Europe and worldwide [1].

Women with diabetes have a 15–20% increased risk of BC compared to women without diabetes [2–8]. Underlying mechanisms include the effects of hyperinsulinemia on sex steroid availability [9,10] and IGF-1 production [11,12]. Hormone-independent mechanisms, including chronic inflammation with elevated levels of pro-inflammatory cytokines, infiltration of fat deposits by pro-inflammatory macrophages, and associated oxidative stress, have also been suggested [13].

Women with gestational diabetes mellitus (GDM) have an increased risk of diabetes, even in the first few years following delivery [14–16]. In addition, GDM shares underlying mechanisms with type 2 diabetes, including hyperglycemia due to pancreatic  $\beta$ -cell dysfunction coupled with insulin resistance and hyperinsulinemia. It is therefore possible that GDM may be associated with an increased risk of cancer [17–25], including BC.

Several studies have explored the association between GDM and cancer. This association was quantitatively analyzed in a meta-analysis by Wang et al. in 2020 [26]. This

meta-analysis was not focused exclusively on BC and gave us few explanations for the discrepancies between the studies. Indeed, studies on this topic show divergent results. Wang et al. reported an odds ratio (OR) of 0.88 (95% CI 0.69; 1.12) for case-control studies and a relative risk (RR) of 1.08 (95% CI 0.89; 1.32) for cohort studies. They found three studies showing a statistically significant increased risk of BC after GDM, two studies showing a statistically significant protective effect, and ten studies showing non-significant results. Similarly, the most recent meta-analysis focusing exclusively on GDM and risk of BC found no significant association between GDM and BC in case-control studies (pooled OR = 0.85, 95% CI 0.65; 1.10) or in cohort studies (pooled RR = 1.00, 95% CI 0.80; 1.25) [27]. These studies demonstrate that the relationship between GDM and BC may be complex.

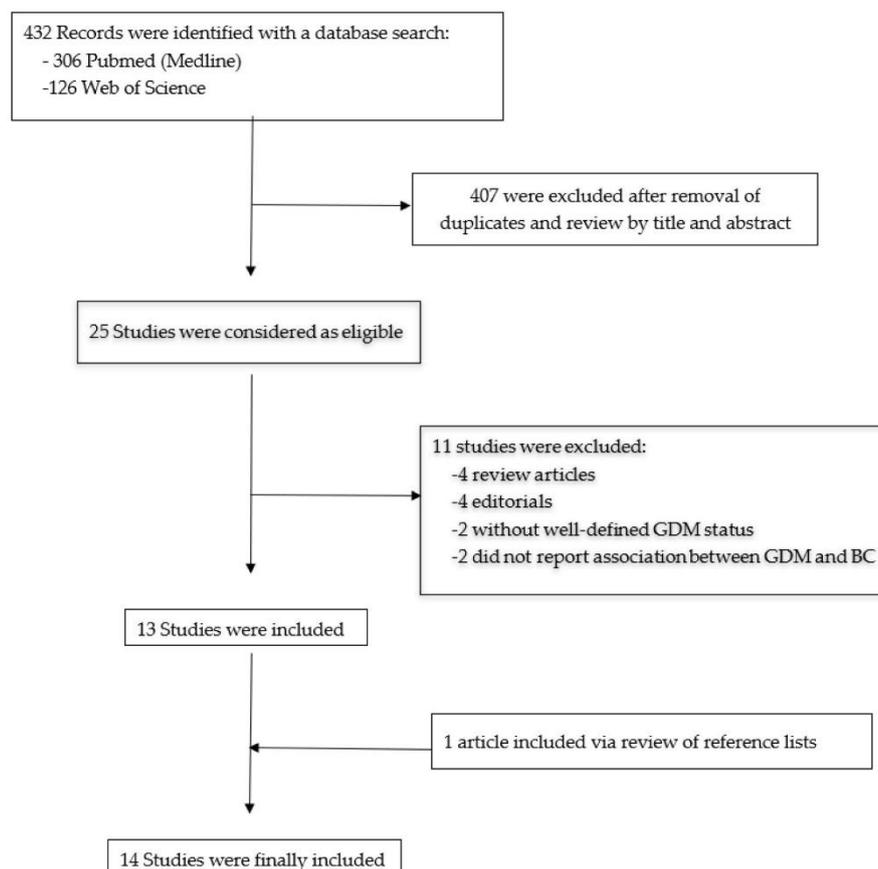
Therefore, the objective of this literature review was to compile the hypotheses studied and the methods employed in the literature that may explain this complex relationship and to encourage future studies on this subject.

## 2. Materials and Methods

We performed this review using the PubMed/Medline and Web Of Science databases using the following search terms: (gestational diabetes\* OR gestational diabetes mellitus OR GDM) AND (breast cancer OR breast tumour OR breast neoplasm\*), in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (<http://prisma-statement.org/prismastatement/Checklist.aspx> 22 July 2021). The reference lists of previous reviews of the subject and of studies included in the analysis were also searched for any further relevant studies. All data extraction was performed by one researcher.

Only English-language articles published before 1 June 2021, were included. We included all types of clinical studies reporting the association between GDM and BC with statistics available for adjusted odds ratio (OR) or relative risk (RR) and 95% confidence intervals (CI). Studies without original data, including reviews, comments, editorials, and meta-analyses were not included in the results but were used in the discussion to provide an additional perspective. Figure 1 shows the study selection process.

Data extracted included first author, publication year, country, specificities of the study population, sample size, duration of follow-up, criteria for GDM diagnosis, GDM rate, criteria for BC diagnosis, age or menopausal status at BC diagnosis, BC rate, adjusted variables, main results and sensitivity or additional analyses. These elements allowed us to qualitatively assess each study and to compare the results in view of the biases and strengths of each study. The risk of potential bias and the quality of each of the included studies was assessed using the Newcastle-Ottawa Scale (NOS), which is a standardized scale for assessing the bias and quality of non-randomized studies. ([http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp), 1 September 2021).



**Figure 1.** Flow diagram of the article selection process. GDM: gestational diabetes mellitus; BC: breast cancer.

### 3. Results

#### 3.1. Literature Search

A total of 732 articles were reviewed by title and abstract for this literature review. Among them, 25 were included for full-text review. Finally, 14 studies that met the inclusion criteria were included in the analysis—ten had a cohort design [19–24,28–31] and four had case-control designs [32–35]. One case-control study, Ardalan et al. [35], was added after a review of reference lists because it met the inclusion criteria; it investigated the relationship between gestational age and BC risk but also reported a measure of the association between GDM and BC.

#### 3.2. Characteristics of Included Studies

Table 1 shows the detailed characteristics of the ten cohort studies and Table 2 shows the detailed characteristics of the four case-control studies. The studies are listed by year of publication in ascending order.

Table 1. Characteristics of the cohort studies.

First Author	Publication Year	Country	Specificities of the Study Population	Sample Size	Duration of Follow-Up	Criteria for GDM Diagnosis	GDM Rate	Criteria for BC Diagnosis	BC Rate	Adjusted Variables	RR [CI 95%]	Sensitivity or Additional Analyses
Perrin [28]	2007	Israel	Deliveries between 1964 to 1976	37,980	Until 2015	Medical records	1%	Cancer registry	4.3%	Age, parity, social class, ethnic origin, education, immigration status	1.5 [1.0; 2.1] P = 0.03	Age at diagnosis of BC: <50 years old RR = 1 [0.5; 2.1] ≥50 years old RR = 1.7 [1.1; 2.5]
Sella [19]	2011	Israel	Aged between 15–50 years Screened between 13 March 1995 and 27 May 2009	185,315	Until June 1st 2009 (mean 5 years)	Medical records	6%	Cancer registry	0.3%	Age, socioeconomic level, smoking, BMI, gravidity, number of general practitioner visits	0.87 [0.63; 1.20]	-
Bejaimal [24]	2016	Canada	Aged between 20–50 years Screened between 1 January 1995 and July 4th 2008	149,049; 49,684 women with GDM matched on age and year of delivery in a ratio of 1:2 to pregnant women without GDM	Until 31 December 2011	Medical Records	4.6%	Cancer Registry	-	Number of physician visits	0.86 [0.75; 0.98]	After adjustment for future diabetes, the association between GDM and risk of breast cancer remained significant (HR 0.85 [0.74; 0.99])
Fuchs [22]	2017	Israel	Deliveries between 1988 to 2013	104,715	Until 2013	Medical records	9.4%	Medical records	0.5%	Univariate results only	1.8 [1.5; 2.1]	/

Table 1. Cont.

First Author	Publication Year	Country	Specificities of the Study Population	Sample Size	Duration of Follow-Up	Criteria for GDM Diagnosis	GDM Rate	Criteria for BC Diagnosis	BC Rate	Adjusted Variables	RR [CI 95%]	Sensitivity or Additional Analyses
Powe [29]	2017	US	Only nurses who declared a delivery at the beginning of the cohort in 1989 or with incident delivery through 2001	86,972	Until 2013	Self-reported, 94% confirmed by medical records	6%	Medical records	2.7%	Age, BMI at 18 years old, weight gain since age 18, height, physical activity, alcohol intake, age at menarche, birth index, breastfeeding, hormone therapy use, history of breast cancer in mother or sister, history of benign breast disease, ethnicity, mammography within the past 2 years	0.72 [0.58; 0.89]	T2D only: HR = 0.69 [0.40; 1.18] GDM + T2D: HR = 0.26 [0.10; 0.68] GDM and hormone receptor positive BC: HR 0.65 [0.50–0.84], GDM and hormone receptor negative BC: HR 0.96 [0.60–1.52]
Park [30]	2017	US	35–74 years old, all breast-cancer-free and sisters of women with BC between 2003 and 2009	39,198	Until 14 August 2015 (mean: 7.4 years)	Self-reported	4.2% (0.9% multiple GDM pregnancies)	Self-reported, 81% confirmed by medical records	5.4%	Age, ethnicity, education, age at first birth, age at menarche, weight at age 10, BMI at 30–39 years old, physical activity in their childhood and teens	1.10 [0.88; 1.36]	1 GDM pregnancy: HR = 0.94 [0.73; 1.22], more than 1 GDM pregnancies: HR = 1.68 [1.15; 2.44]
Han [21]	2018	South Korea	1st pregnancy between 2004 and 2005 and participated in a “National Health Screening examination” at least 2 years before their first delivery	102,900	Until 2015	Medico-administrative database ICD 10 codes O 24.4 and O 24.9	4.83%	Medico-administrative database ICD 10 codes	0.7%	Age, smoking, BMI before pregnancy	1.15 [0.83; 1.58]	-

Table 1. Cont.

First Author	Publication Year	Country	Specificities of the Study Population	Sample Size	Duration of Follow-Up	Criteria for GDM Diagnosis	GDM Rate	Criteria for BC Diagnosis	BC Rate	Adjusted Variables	RR [CI 95%]	Sensitivity or Additional Analyses
Peng [23]	2019	Taiwan	All deliveries between 2002 and 2012	990,572	Until 31 December 2013	Medico-administrative database ICD 9 codes 648-250	4.8%	Medico-administrative database ICD 9 codes	0.47%	Age, hypertension, dyslipidemia, liver disease, infertility and kidney disease	1.23 [1.09; 1.39]	-
Pace [20]	2020	Canada	34,294 randomly selected deliveries between April 1990 and 31 December 2007 with 2 or more diagnostic codes for GDM were matched 1:1 to 34,294 deliveries without GDM (by year of delivery, age, region)	68,588	Until 2012 (mean: 13.1 years)	2 or more GDM codes	-	Hospital discharge abstracts using ICD codes	1.1%	Gestational hypertension, preterm delivery, infant size, parity, prior comorbidity, material deprivation index, ethnicity	0.93 [0.80; 1.09]	-
Bertrand [31]	2020	US	African American parous women between March 1995 and March 2017	41,767	Until March 2017	Self-reported	4.9%	Self-reported and confirmed in cancer registry records or medical records	4%	Age, BMI at age 18, recent BMI, parity, menarche, age at first birth, oral contraceptive duration, and family history of BC	0.98 [0.77; 1.25]	No significant results after adjustment on ER status

GDM: gestational diabetes mellitus; BC: breast cancer; RR: relative risk; HR: hazard ratio; BMI: body mass index; T2D: type 2 diabetes; ICD: International Classification of Disease.

Table 2. Characteristics of the case-control studies.

First Author	Publication Year	Country	Specificities of the Study Population	Sample Size	Criteria for GDM Diagnosis	GDM Rate	Criteria for BC Diagnosis	Adjusted Variables	OR [CI 95%]	Sensitivity or Additional Analyses
Troisi [32]	1998	US	Age between 20 and 44 years	3174	Self-report	7.5%	Medical records	Age at menarche, mammography, alcohol, BMI, site, race, combination variables representing parity and age at first birth	1.1 [0.7; 1.4]	-
Rollison [33]	2008	US	Non-Hispanic white, Hispanic, and American Indian women (mean age: 56 years)	2523	Self-report	3.2%	Cancer registry	Age, study site, menopausal status, BMI, BMI at 15 years old, parity, age at menarche, physical activity, family history of BC, breastfeeding history	0.71 [0.52; 0.98]	Stratification on age at onset of GDM: 0.56 [0.38; 0.82] ≥ 35 years old 1.34 [0.72; 2.52] ≥ 35 years old Stratification on age at onset of GDM and ER/PR status: <35 years old and ER + 0.52 [0.31; 0.85] <35 years old and PR + 0.53 [0.32; 0.89] ≥35 years old and ER – 2.52 [1.07; 5.53] ≥35 years old and PR – 3.08 [1.45; 6.54]
Brasky [34]	2013	US	Age between 35 and 79 years	2812	Self-report	3.3%	Medical records	Age, education, history of benign breast disease, family history of BC, age at first pregnancy, number of pregnancies, menopausal status, age at menopause	0.79 [0.48; 1.30]	-

Table 2. Cont.

First Author	Publication Year	Country	Specificities of the Study Population	Sample Size	Criteria for GDM Diagnosis	GDM Rate	Criteria for BC Diagnosis	Adjusted Variables	OR [CI 95%]	Sensitivity or Additional Analyses
Ardalan [35]	2016	US	Women diagnosed with premenopausal BC (age 21–49) within the 5 years after delivering a live baby between 1 January 1994 and 31 December 2003	630	Recorded birth certificates	1.4%	Cancer registry	Mother's age at delivery, race/ethnicity, level of education, birth weight, parity, gestational age, weight gain in pregnancy, smoking habits, drinking habits, induction of labor, gestational hypertension	1.62 [0.30; 8.68]	-

GDM: gestational diabetes mellitus; BC: breast cancer; OR: odds ratio; BMI: body mass index; ER, estrogen receptor; PR, progesterone receptor.

### 3.2.1. Cohort Studies

Of the ten cohort studies, five were conducted in North America, three in Israel, and two in Asia. The first was published in 2008, and at least one study was published each year between 2015 and 2020. The cohorts were rather large, from 37,980 subjects for Perrin et al. (2008) [28] to 990,572 subjects for Peng et al. (2019) [23]. However, the vast majority of these studies recruited their populations before the most recent diagnostic criteria for GDM was published in 2010 [36]. The oldest cohorts were mostly prospective cohorts based on medical records and cancer registry data, while the most recent cohorts were mostly retrospective and based on self-reported or medico-administrative data.

It should also be noted that some studies had specific inclusion criteria, for example, Powe et al. [29] included only nurses, Park et al. [30] included only sisters of women with diagnosed BC, Han et al. [21] included only patients with a national screening examination within two years of pregnancy, and Bertrand et al. [31] included only African-American women. Follow-up was long in some studies, up to several decades in Perrin et al. [28] (median follow-up of 34 years), but in other studies, follow-up lasted only a few weeks for some of the cases (e.g., Bertrand et al. [31] and Fuchs et al. [22]). The rates of GDM and BC differed between these four studies, ranging from 1% to 9.4% and 0.3% to 5.4%, respectively. Only three cohort studies clearly stated if the BC was in situ or invasive [19,29,31], and two cohort studies differentiated between hormone-dependent and non-hormone-dependent cancers [29,31].

The results of these studies also likely differed due to the choice of adjustment variables. We found three studies showing a statistically significant increased risk of BC after GDM, two studies showing a statistically significant protective effect, and five studies showing non-significant results. Concerning the adjustment variables, it should be noted that Fuchs et al. [22] published only univariate results and found a statistically significant protective effect [22]. However, Bejeimal et al. only took into account the number of visits to the doctor and found a statistically significant increased risk [24]. The vast majority of the other cohort studies adjusted for age, parity, ethnicity, or BMI. Four studies [19,20,28,30] included socio-economic variables (e.g., education, social class) and five studies [19,21,29–31] took into account lifestyle habits (smoking, alcohol, physical activity, contraception).

### 3.2.2. Case-Control Studies

The four case-control studies were conducted in the US. They were published between 1998 and 2016 and included between 630 and 3174 subjects. These studies were based on self-reported data for the diagnosis of GDM, except for Ardalan et al. [35], which was based on recorded birth certificates. The diagnosis of BC was based on medical records, except for Rollison et al. [33], which was based on a cancer registry. Only one case-control study clearly presented whether BC was in situ or invasive [33] and two case-control studies specified hormone-dependent and non-hormone-dependent cancers [33,34].

In Troisi et al. [32] and Ardalan et al. [35], the subjects were younger than in the other two studies and the GDM rate ranged from 1.4% to 7.5%. It should also be noted that Rollison et al. [33] only included non-Hispanic whites, Hispanics, and American Indian women.

Among the case-control studies, we found only one study showing a statistically significant protective effect of GDM and three studies showing non-significant results. Regarding adjustment variables, all case-control studies included age and parity, three studies considered menopausal status and ethnicity, two took into account BMI, two took into account family history of BC, education, and alcohol consumption. Finally, one adjusted for physical activity and another for smoking habits.

### 3.3. Risk of Bias in the Included Studies

Tables 3 and 4 show, respectively, the risk of bias for cohort studies and case-control studies, assessed via star rating according to the NOS.

Table 3. Risk of bias and quality assessment of cohort studies.

First Author	Publication Year	Selection			Comparability			Outcome		Total	
		Representativeness of the Exposed Cohort	Selection of the Non-Exposed Cohort	Ascertainment of Exposure	Demonstration that Outcome of Interest Was Not Present at Start of Study	Controls for the Most Important Factor	Controls for any Additional Factor	Assessment of Outcome	Was Follow up Long Enough for Outcome to Occur		Adequacy of Follow up of Cohorts
Perrin [28]	2008	*	*	*	*	*	*	*	*	*	9
Sella [19]	2011	*	*	*	*	*	*	*	*	*	8
Bejaimal [24]	2015	*	*	*	*	*	*	*	*	*	9
Fuchs [22]	2016	*	*	*	*	*	*	*	*	*	7
Powe [29]	2017		*	*	*	*	*	*	*	*	6
Park [30]	2017	*	*	*	*	*	*	*	*	*	6
Han [21]	2018		*	*	*	*	*	*	*	*	8
Peng [23]	2019	*	*	*	*	*	*	*	*	*	9
Pace [20]	2020	*	*	*	*	*	*	*	*	*	8
Bertrand [31]	2020	*	*			*	*	*			5

Note: Assessments are based on Newcastle-Ottawa Scale. 'high' quality choices are identified with a '\*'.

Table 4. Risk of bias and quality assessment of case-control studies.

First Author	Publication Year	Selection			Comparability			Exposure		Total	
		Is the Case Definition Adequate?	Representativeness of the Cases	Selection of Controls	Definition of Controls	Controls for the Most Important Factor	Controls for Any Additional Factor	Ascertainment of Exposure	Same method of Ascertainment for Cases and Controls		Non-Response Rate
Troisi [32]	1998	*	*	*		*	*		*		6
Rollison [33]	2008		*	*		*	*		*	*	6
Brasky [34]	2013	*	*	*		*	*		*	*	7
Ardalan [35]	2016		*	*	*			*	*	*	6

Note: Assessments are based on Newcastle-Ottawa Scale. 'high' quality choices are identified with a '\*'.

Among the ten cohort studies, three studies had nine stars, three studies had eight stars, one study had seven stars, three studies had six stars, and another had five stars. Two studies selected the group of users (nurses, participants in a screening) [21,29]. In two studies, the diagnosis of GDM was self-reported [30,31]. In three studies, the outcome of interest was not present at the start [20,29,31]. In two studies, the follow-up was not long enough for outcomes to occur [22,31]. Five studies had a follow-up rate below 94% and/or no description of patients who were lost to follow-up [19,22,29–31]. In the study by Park et al. [30], the diagnosis of BC was self-reported and confirmed by medical records in only 81% of the cases [30].

Among the four case-control studies, three studies had six stars and one study had seven stars. Two studies did not have an adequate case definition, with record linkage using registers [33,35]. Three studies did not describe their source for the definition of controls and interviews were not blinded to case/control status [32–34]. In the study by Troisi et al. [32], the response rate to the interview was lower for controls than for cases [32]. In Ardalan et al. [35], there were no controls for GDM, given that the main objective of this study was to investigate the relationship between gestational age and BC [35].

#### 4. Discussion

In this literature review, we found 14 studies investigating the relationship between GDM and BC risk. Among the 10 cohorts, three studies showed a statistically significant increased risk of BC after GDM, two studies showed a statistically significant protective effect, and five studies reported non-significant results. Among the four case-control studies, one study showed a statistically significant protective effect and three studies reported non-significant results. The 14 studies were of varying quality and used different methodologies. Using the NOS assessments, only 3 out of the 14 obtained all stars, which could explain the diversity of the results.

The case-control studies were all conducted in the US. The cohort studies showing either non-significant results or a statistically significant protective effect were also mostly carried out in North America. All the studies showing a statistically significant increased risk were carried out on the Asian continent (Israel, China). Regarding ethnicity and BC risk, Perrin et al. [28] showed that women of West Asian and North African origin have a significantly lower risk of developing BC. Therefore, a first hypothesis would be that ethnicity, culture, and/or country of residence has an influence on the relationship between GDM and BC.

A second hypothesis would be that the heterogeneity of the results may be partly explained by the use of different definitions of exposure (i.e., the diagnosis of GDM), which resulted in different rates of GDM. In the case-control studies, the history of GDM was self-reported, except for the study Ardalan et al. [35], which used birth certificate records, resulting in a GDM rate of 1.4%, but the exhaustiveness of the certificates is questionable. In the study by Troisi et al. [32], the women were young (between 20 and 44 years old), so, it can be assumed that the time between the diagnosis of GDM and the subjects' responses was rather short, and they found a higher GDM rate of 7.5%. In contrast, the other two case-control studies had older populations, which may have led to recall bias, and the GDM rate was lower at around 3.3%. It should also be taken into account that the diagnostic criteria for GDM have evolved over the last 50 years, and earlier criteria detected only the most severe forms of GDM [36]. For example, the study by Perrin et al. [28] on pregnancies between 1964 and 1976 showed a GDM rate of 1%, whereas the more recent cohort studies report a rate of between 4.5% and 6%.

There is also the question of the treatment and monitoring of GDM. If the pathophysiological mechanism of an increased risk of BC is a combination of hyperglycemia coupled with insulin resistance, and hyperinsulinemia is presumed, then GDM balanced by regular follow-up and diet should not generate an excess risk compared with poorly followed or unbalanced GDM. Indeed, in 2004, Dawson et al. showed that "after adjusting for age, BMI, and smoking behavior, the risk of BC was found to be 1.2 to 5.1 times higher, respectively,

in the highest quartile compared with the lowest quartile depending on the glucose indices used" [37]. One hypothesis would be that insulin-treated GDM may increase the risk of BC but not GDM balanced with diet.

We could go further by assuming that the protective effect of GDM found in several studies could be explained by the improvement of lifestyle habits (physical activity, diet) over time. However, in our literature review the one study that explored these elements (Powe et al. [29]) reported that GDM had a protective effect, but there was no significant differences in diet and physical activity between women with and without a history of GDM. However, it should also be noted that Powe et al. [29] focused on nurses, who are more likely to be conscious about their health, but who work at night, which is known as a BC risk factor [38].

According to the literature, some pathophysiological mechanisms could explain the protective effect, such as the altered hormonal milieu of a pregnancy complicated by GDM conferring protection to breast tissue. Pregnancy is known to induce proliferation and functional differentiation of breast lobules and ducts, which may explain the association between pregnancy history and BC risk [39–41]. Differences in circulating growth factors or hormones in GDM pregnancies may also have an impact on these processes and future BC risk.

Similarly, the criteria for diagnosing BC differ between studies and may lead to various classification biases. For instance, we can take the study by Park et al. [30] where the diagnosis of BC was self-reported and was confirmed by medical records in only 81% of cases. Cancer registries can also lead to a classification bias if they are not complete or up to date. The same applies to the completeness of medico-administrative databases based on hospital admissions, and the question arises for cancers that do not always require hospitalization, such as BC. Furthermore, only four studies clearly presented whether the cancer was in situ or invasive [19,29,31,33] and only four studies differentiated between hormone-dependent and non-hormone-dependent cancers [29,31,33,34], even though these cancers do not have the same pathophysiology and the same risk factors. Additionally, it can be assumed that patients in the control group may have had undetected BC, thus contributing to another classification bias.

The included studies also had very heterogeneous adjustment variables. Among the six cohort studies that did not focus solely on BC, most did not adjust for several known BC risk factors (e.g., oral contraceptive use, breastfeeding, alcohol consumption, or family history of BC). Only Powe et al. [29] and Bertrand et al. [31] adjusted for these risk factors, excluding alcohol consumption. Among the case-control studies, Troisi et al. [32] and Ardalan et al. [35] adjusted for alcohol consumption, while Rollison et al. [33] and Brasky et al. [34] adjusted for family history of BC; no case-control study adjusted for oral contraceptive use. There are also two risk factors for BC that are also known risk factors for GDM—obesity and physical inactivity. Therefore, it can be assumed that these factors play a confounding role in the relationship between BC and GDM [42], as well as type 2 diabetes. Thus, several studies have shown that GDM increased the risk of type 2 diabetes [14,16], and type 2 diabetes is recognized as increasing the risk of BC [43]. Although the vast majority of studies adjusted for BMI, only three studies adjusted for physical activity and none adjusted for type 2 diabetes in the main analyses. However, in their secondary analysis, Bejaimal et al. [24] showed that after adjustment for type 2 diabetes, the association between GDM and risk of BC remained significant (HR = 0.85, 95% CI [0.74; 0.99]). After a stratification on type 2 diabetes, Powe et al. [29] showed that GDM had a statistically significant protective effect on the risk of BC. In an additional analysis in their population of nurses, they showed that those with a history of GDM and type 2 diabetes had an even lower risk of BC (HR = 0.26, 95% CI [0.10; 0.68]).

Several studies have also shown that age is an important factor to take into account. Firstly, regarding age at onset of GDM, Rollison et al. [33], who showed a statistically significant protective effect of GDM on BC risk in their population, showed an even more protective effect in women who were younger than 35 years at onset of GDM

(OR = 0.56, 95% CI [0.38; 0.82]). Secondly, the age at diagnosis of BC or the menopausal status of patients would influence the relationship between GDM and BC risk. For example, Perrin et al. [28] showed that there is a significantly increased risk with a history of GDM for BC diagnosed after age 50 (RR = 1.7 CI 95% [1.1; 2.5]).

Regarding BC subtypes, three studies adjusted or stratified on estrogen and/or progesterone receptor (ER/PR) status. No study worked on HER2 status. Bertrand et al. [31] did not show a statistically significant association between GDM and BC by estrogen receptor status. However, Rollison et al. [33] and Powe et al. [29], who had already found a statistically significant protective effect of GDM on BC risk, found an even more protective relationship for hormone-receptor-positive BC. However, this statistically significant relationship in Rollison et al. [33] was only found in women with GDM diagnosed before age 35. It would be interesting to further investigate this relationship between GDM and the different subtypes of BC. Only one study considered the effect of a history of multiple occurrences of GDM on the risk of BC. This study by Park et al. [30] was conducted in a particular population of sisters of women who had both BC and history of more than one occurrence of GDM. They studied the cumulative effect of the number of GDM occurrences and showed a statistically significant increased risk of BC.

Genomic characteristics were not discussed by any of the studies included in our literature review, even though genetic mutations could influence BC and GDM. We found several studies showing an increased frequency of certain mutations in BC patients. The most well-known germline mutations are BRCA1 and BRCA2, but others, such as PALB2 or TP53, may also be found [44–47]. Similarly, several hypotheses have been formulated regarding the link between genome mutations, such as GCK, HLA antigens, INSR, IGF2, HNF4A, INS-VNTR, HNF4a, CAPN10, MBL2, KCNJ11, ABCC8, ND1, TCF7L2, ADIPOQ, and PAI-1, and the development of GDM [48–50]. Chen et al. also suggested that mitochondrial DNA mutations could promote the development of GDM in some patients [51]. In addition, there are genetic mutations that could have an impact on the development of both GDM and BC, such as mutations in the ACE I/D gene [52,53]. The literature provides several hypotheses about the link between GDM, genetic mutations, and cancer risk. For example, Witczak et al. explained that: “gestational diabetes mellitus (GDM) is defined as carbohydrate intolerance resulting in hyperglycemia first diagnosed during pregnancy. It is associated with increased levels of oxidative stress due to overproduction of reactive oxygen species (ROS). The overproduction of ROS induces protein oxidation, lipid peroxidation and various types of DNA damage” [54]. However, it can be assumed that the currently established follow-up and treatments for GDM reduce the consequences on DNA. It would, therefore, be interesting in future studies to clarify the link between genotypes and the influence that they may have on the risk of GDM and/or BC in order to improve prevention and treatment. Regarding the length of follow-up required to study the relationship between GDM and BC risk, the existing literature has not clearly established the ideal length. If the follow-up is less than 15 years, as in the vast majority of cohorts in this review, there is a risk of studying predominantly pre-menopausal BC. A follow-up of more than 15 years would make it possible to evaluate the effects of GDM in the longer term, particularly on menopausal BC. While studies suggest that type 2 diabetes and obesity are not associated with premenopausal BC or with a lower risk [19,28,55], high BMI is positively associated with risk of postmenopausal BC [56]. A large sample size is also required in this type of study since the BC rate is usually less than 1%.

Our literature review has several limitations. Firstly, it was conducted by a single reviewer. However, we were able to rely on two recent meta-analyses [26,27] to ensure that our review was exhaustive. Secondly, it was difficult to make a straightforward comparison between studies. Indeed, the study populations were either too small, with respect to the number of cases, to have sufficient power, or too unrepresentative (majority of studies done in the US, with a particular ethnicity, with cohorts of nurses, based on participation in a national health screening examination, or with a cohort of sisters of women with BC). There were also potential selection biases, including incomplete follow-up in cohort

studies and low response rates in case-control studies. Exposure misclassification is another particular methodological issue and several studies included in this systematic review used self-reported estimates instead of objective clinical diagnosis based on biochemical testing [57]. The adjustment variables are also very heterogeneous and sometimes do not correspond to known risk factors for GDM or BC.

## 5. Conclusions

To conclude, the results of studies addressing the link between GDM and BC seem to be inconsistent and they should be interpreted with caution seeing as many present quantitative or qualitative methodological issues. The link between GDM and BC, therefore, warrants further investigation and a better understanding of the associations and the underlying mechanisms is needed. It would be advisable for studies to be conducted not only in North America and Asia, but also in the rest of the world, and for them to be representative of their populations. They should have a clear and up-to-date definition of GDM, as well as the different types of BC (menopausal or not, hormone-dependent or not, invasive or not). It would be useful to know what measures or treatments were put in place after the diagnosis of GDM and to study the cumulative effect of GDM. It would also be interesting to adjust for known risk factors for BC and GDM, especially obesity and/or physical inactivity, as well as type 2 diabetes, which could be considered as potential confounders in this relationship.

**Author Contributions:** Conceptualization, J.S., K.G., and C.Q.; methodology, C.Q.; software, K.G.; validation, J.S., K.G., and C.Q.; formal analysis, K.G.; data curation, C.Q.; writing—original draft preparation, J.S., K.G., and C.Q.; writing—review and editing, J.S., K.G., S.B.-Q., P.A., and C.Q.; visualization, C.Q.; supervision, C.Q.; project administration, C.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** We performed this review using the PubMed/Medline and Web Of Science databases using the following search terms: (gestational diabetes\* OR gestational diabetes mellitus OR GDM) AND (breast cancer OR breast tumour OR breast neoplasm\*), in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

**Acknowledgments:** The authors thank Harbajan Chadha-Boreham and Suzanne Rankin for reviewing the English and Gwenaëlle Periard for her help with the layout and management of this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
2. Michels, K.B.; Solomon, C.G.; Hu, F.B.; Rosner, B.A.; Hankinson, S.E.; Colditz, G.; Manson, J.E. Type 2 Diabetes and Subsequent Incidence of Breast Cancer in the Nurses' Health Study. *Diabetes Care* **2003**, *26*, 1752–1758. [[CrossRef](#)]
3. Xue, F.; Michels, K.B. Diabetes, metabolic syndrome, and breast cancer: A review of the current evidence. *Am. J. Clin. Nutr.* **2007**, *86*, 823S–835S. [[CrossRef](#)] [[PubMed](#)]
4. Larsson, S.C.; Mantzoros, C.S.; Wolk, A. Diabetes mellitus and risk of breast cancer: A meta-analysis. *Int. J. Cancer* **2007**, *121*, 856–862. [[CrossRef](#)] [[PubMed](#)]
5. Liao, S.; Li, J.; Wei, W.; Wang, L.; Zhang, Y.; Li, J.; Wang, C.; Sun, S. Association between diabetes mellitus and breast cancer risk: A meta-analysis of the literature. *Asian Pac. J. Cancer Prev.* **2011**, *12*, 5.
6. Boyle, P.; Boniol, M.; Koehlin, A.; Robertson, C.; Valentini, F.; Coppens, K.; Fairley, L.-L.; Zheng, T.; Zhang, Y.; Pasterk, M.; et al. Diabetes and breast cancer risk: A meta-analysis. *Br. J. Cancer* **2012**, *107*, 1608–1617. [[CrossRef](#)]
7. Linde, J.S.; Karlstad, O.; Eriksen, S.A.; Vestergaard, P.; Bronsveld, H.K.; de Vries, F.; Andersen, M.; Auvinen, A.; Haukka, J.; Hjellvik, V.; et al. CARING (CAncer Risk and INSulin analogues): The Association of Diabetes Mellitus and Cancer Risk with Focus on Possible Determinants—A Systematic Review and a Meta-Analysis. *Curr. Drug Saf.* **2013**, *8*, 296–332. [[CrossRef](#)]

8. Dankner, R.; Boffetta, P.; Balicer, R.D.; Boker, L.K.; Sadeh, M.; Berlin, A.; Olmer, L.; Goldfracht, M.; Freedman, L.S. Time-Dependent Risk of Cancer After a Diabetes Diagnosis in a Cohort of 2.3 Million Adults. *Am. J. Epidemiol.* **2016**, *183*, 1098–1106. [[CrossRef](#)] [[PubMed](#)]
9. Nyholm, H.; Djursing, H.; Hagen, C.; Agner, T.; Bennett, P.; Svenstrup, B. Androgens and Estrogens in Postmenopausal Insulin-Treated Diabetic Women. *J. Clin. Endocrinol. Metab.* **1989**, *69*, 946–949. [[CrossRef](#)] [[PubMed](#)]
10. Lipworth, L.; Adami, H.-O.; Trichopoulos, D.; Cartström, K.; Mantzoros, C. Serum Steroid Hormone Levels, Sex Hormone-Binding Globulin, and Body Mass Index in the Etiology of Postmenopausal Breast Cancer. *Epidemiology* **1996**, *7*, 96–100. [[CrossRef](#)]
11. Lawlor, D.A.; Smith, G.D.; Ebrahim, S. Hyperinsulinaemia and increased risk of breast cancer: Findings from the British Women's Heart and Health Study. *Cancer Causes Control.* **2004**, *15*, 267–275. [[CrossRef](#)]
12. Chaudhuri, P.K.; Chaudhuri, B.; Patel, N. Modulation of estrogen receptor by insulin and its biologic significance. *Arch. Surg.* **1986**, *121*, 1322–1325. [[CrossRef](#)] [[PubMed](#)]
13. Ferroni, P.; Riondino, S.; Buonomo, O.; Palmirotta, R.; Guadagni, F.; Roselli, M. Type 2 Diabetes and Breast Cancer: The Interplay between Impaired Glucose Metabolism and Oxidant Stress. *Oxidative Med. Cell. Longev.* **2015**, *2015*, 1–10. [[CrossRef](#)] [[PubMed](#)]
14. Bellamy, L.; Casas, J.-P.; Hingorani, A.; Williams, D. Type 2 diabetes mellitus after gestational diabetes: A systematic review and meta-analysis. *Lancet* **2009**, *373*, 1773–1779. [[CrossRef](#)]
15. Kaul, P.; Savu, A.; Nerenberg, K.A.; Donovan, L.E.; Chik, C.L.; Ryan, E.A.; Johnson, J. Impact of gestational diabetes mellitus and high maternal weight on the development of diabetes, hypertension and cardiovascular disease: A population-level analysis. *Diabet. Med.* **2015**, *32*, 164–173. [[CrossRef](#)]
16. Retnakaran, R.; Shah, B.R. Role of Type 2 Diabetes in Determining Retinal, Renal, and Cardiovascular Outcomes in Women With Previous Gestational Diabetes Mellitus. *Diabetes Care* **2016**, *40*, 101–108. [[CrossRef](#)] [[PubMed](#)]
17. Wartko, P.D.; Beck, T.L.; Reed, S.; Mueller, B.A.; Hawes, S.E. Association of endometrial hyperplasia and cancer with a history of gestational diabetes. *Cancer Causes Control* **2017**, *28*, 819–828. [[CrossRef](#)]
18. Perrin, M.C.; Terry, M.B.; Kleinhaus, K.; Deutsch, L.; Yanetz, R.; Tiram, E.; Calderon, R.; Friedlander, Y.; Paltiel, O.; Harlap, S. Gestational diabetes as a risk factor for pancreatic cancer: A prospective cohort study. *BMC Med.* **2007**, *5*, 25. [[CrossRef](#)]
19. Sella, T.; Chodick, G.; Barchana, M.; Heymann, A.D.; Porath, A.; Kokia, E.; Shalev, V. Gestational diabetes and risk of incident primary cancer: A large historical cohort study in Israel. *Cancer Causes Control* **2011**, *22*, 1513–1520. [[CrossRef](#)]
20. Pace, R.; Rahme, E.; Dasgupta, K. Gestational diabetes mellitus and risk of incident primary cancer: A population-based retrospective cohort study. *J. Diabetes* **2020**, *12*, 87–90. [[CrossRef](#)]
21. Han, K.-T.; Cho, G.J.; Kim, E.H. Evaluation of the Association between Gestational Diabetes Mellitus at First Pregnancy and Cancer within 10 Years Postpartum Using National Health Insurance Data in South Korea. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2646. [[CrossRef](#)] [[PubMed](#)]
22. Fuchs, O.; Sheiner, E.; Meirovitz, M.; Davidson, E.; Sergienko, R.; Kessous, R. The association between a history of gestational diabetes mellitus and future risk for female malignancies. *Arch. Gynecol. Obstet.* **2017**, *295*, 731–736. [[CrossRef](#)]
23. Peng, Y.-S.; Lin, J.-R.; Cheng, B.-H.; Ho, C.; Lin, Y.-H.; Shen, C.-H.; Tsai, M.-H. Incidence and relative risk for developing cancers in women with gestational diabetes mellitus: A nationwide cohort study in Taiwan. *BMJ Open* **2019**, *9*, e024583. [[CrossRef](#)]
24. Bejaimal, S.A.D.; Wu, C.F.; Lowe, J.; Feig, D.S.; Shah, B.R.; Lipscombe, L.L. Short-term risk of cancer among women with previous gestational diabetes: A population-based study. *Diabet. Med.* **2016**, *33*, 39–46. [[CrossRef](#)] [[PubMed](#)]
25. Simon, J.; Goueslard, K.; Arveux, P.; Bechraoui-Quantin, S.; Petit, J.-M.; Quantin, C. Increased Risk of Hospitalization for Pancreatic Cancer in the First 8 Years after a Gestational Diabetes Mellitus regardless of Subsequent Type 2 Diabetes: A Nationwide Population-Based Study. *Cancers* **2021**, *13*, 308. [[CrossRef](#)] [[PubMed](#)]
26. Wang, Y.; Yan, P.; Fu, T.; Yuan, J.; Yang, G.; Liu, Y.; Zhang, Z.-J. The association between gestational diabetes mellitus and cancer in women: A systematic review and meta-analysis of observational studies. *Diabetes Metab.* **2020**, *46*, 461–471. [[CrossRef](#)] [[PubMed](#)]
27. Xie, C.; Wang, W.; Li, X.; Shao, N.; Li, W. Gestational diabetes mellitus and maternal breast cancer risk: A meta-analysis of the literature. *J. Matern. -Fetal Neonatal Med.* **2017**, *32*, 1022–1032. [[CrossRef](#)] [[PubMed](#)]
28. Perrin, M.C.; Terry, M.B.; Kleinhaus, K.; Deutsch, L.; Yanetz, R.; Tiram, E.; Calderon-Margalit, R.; Friedlander, Y.; Paltiel, O.; Harlap, S. Gestational diabetes and the risk of breast cancer among women in the Jerusalem Perinatal Study. *Breast Cancer Res. Treat.* **2007**, *108*, 129–135. [[CrossRef](#)]
29. Powe, C.E.; Tobias, D.K.; Michels, K.B.; Chen, W.Y.; Eliassen, A.H.; Manson, J.E.; Rosner, B.; Willett, W.C.; Hu, F.B.; Zhang, C.; et al. History of Gestational Diabetes Mellitus and Risk of Incident Invasive Breast Cancer among Parous Women in the Nurses' Health Study II Prospective Cohort. *Cancer Epidemiol. Prev. Biomark.* **2017**, *26*, 321–327. [[CrossRef](#)]
30. Park, Y.-M.; O'Brien, K.M.; Zhao, S.; Weinberg, C.R.; Baird, D.D.; Sandler, D.P. Gestational diabetes mellitus may be associated with increased risk of breast cancer. *Br. J. Cancer* **2017**, *116*, 960–963. [[CrossRef](#)]
31. Bertrand, K.A.; Castro-Webb, N.; Cozier, Y.C.; Li, S.; O'Brien, K.M.; Rosenberg, L.; Palmer, J.R. Gestational Diabetes and Risk of Breast Cancer in African American Women. *Cancer Epidemiol. Prev. Biomark.* **2020**, *29*, 1509–1511. [[CrossRef](#)]
32. Troisi, R.; Weiss, H.A.; Hoover, R.N.; Potischman, N.; Swanson, C.A.; Brogan, D.R.; Coates, R.J.; Gammon, M.D.; Malone, K.E.; Daling, J.R.; et al. Pregnancy Characteristics and Maternal Risk of Breast Cancer. *Epidemiology* **1998**, *9*, 641–647. [[CrossRef](#)]
33. Rollison, D.E.; Giuliano, A.R.; Sellers, T.A.; Laronga, C.; Sweeney, C.; Risendal, B.; Baumgartner, K.B.; Byers, T.; Slattery, M.L. Population-based Case-Control Study of Diabetes and Breast Cancer Risk in Hispanic and Non-Hispanic White Women Living in US Southwestern States. *Am. J. Epidemiol.* **2007**, *167*, 447–456. [[CrossRef](#)]

34. Brasky, T.M.; Li, Y.; Jaworowicz, D.J.; Potischman, N.; Ambrosone, C.B.; Hutson, A.D.; Nie, J.; Shields, P.G.; Trevisan, M.; Rudra, C.B.; et al. Pregnancy-related characteristics and breast cancer risk. *Cancer Causes Control* **2013**, *24*, 1675–1685. [[CrossRef](#)]
35. Ardalan, A.; Bungum, T. Gestational Age and the Risk of Maternal Breast Cancer: A Population-Based Case-Control Study. *Breast J.* **2016**, *22*, 657–661. [[CrossRef](#)]
36. Karagiannis, T.; Bekiari, E.; Manolopoulos, K.; Paletas, K.; Tsapas, A. Gestational diabetes mellitus: Why screen and how to diagnose. *Hippokratia* **2010**, *14*, 151–154. [[PubMed](#)]
37. Dawson, S.I. Long-term risk of malignant neoplasm associated with gestational glucose intolerance. *Cancer* **2003**, *100*, 149–155. [[CrossRef](#)] [[PubMed](#)]
38. Cordina-Duverger, E.; Menegaux, F.; Popa, A.; Rabstein, S.; Harth, V.; Pesch, B.; Brüning, T.; Fritschi, L.; Glass, D.; Heyworth, J.; et al. Night shift work and breast cancer: A pooled analysis of population-based case-control studies with complete work history. *Eur. J. Epidemiol.* **2018**, *33*, 369–379. [[CrossRef](#)]
39. Schedin, P. Pregnancy-associated breast cancer and metastasis. *Nat. Rev. Cancer* **2006**, *6*, 281–291. [[CrossRef](#)] [[PubMed](#)]
40. Russo, I.H.; Russo, J. Pregnancy-Induced Changes in Breast Cancer Risk. *J. Mammary Gland. Biol. Neoplasia* **2011**, *16*, 221–233. [[CrossRef](#)] [[PubMed](#)]
41. Meier-Abt, F.; Bentires-Alj, M.; Rochlitz, C. Breast Cancer Prevention: Lessons to be Learned from Mechanisms of Early Pregnancy-Mediated Breast Cancer Protection. *Cancer Res.* **2015**, *75*, 803–807. [[CrossRef](#)]
42. Tankó, L.B.; Bagger, Y.Z.; Christiansen, C. Long-term risk of malignant neoplasm associated with gestational glucose intolerance. *Cancer* **2004**, *100*, 2680–2681. [[CrossRef](#)] [[PubMed](#)]
43. Chodick, G.; Zucker, I. Diabetes, Gestational Diabetes and the Risk of Cancer in Women: Epidemiologic Evidence and Possible Biologic Mechanisms. *Women's Health* **2011**, *7*, 227–237. [[CrossRef](#)]
44. Tariq, H.; Gul, A.; Zubair, M.; Jaffer, S.R.; Zafar, N.; Sadaf, G. Targeted Next-generation Sequencing for Reliable Detection of Genetic Status in Breast Cancer. *J. Coll. Physicians Surg. -Pak. JCPSP* **2021**, *30*, 837–840. [[PubMed](#)]
45. Guglielmi, C.; Scarpitta, R.; Gambino, G.; Conti, E.; Bellè, F.; Tancredi, M.; Cervelli, T.; Falaschi, E.; Cosini, C.; Aretini, P.; et al. Detection of Germline Variants in 450 Breast/Ovarian Cancer Families with a Multi-Gene Panel Including Coding and Regulatory Regions. *Int. J. Mol. Sci.* **2021**, *22*, 7693. [[CrossRef](#)]
46. Ren, M.; Orozco, A.; Shao, K.; Albanez, A.; Ortiz, J.; Cao, B.; Wang, L.; Barreda, L.; Alvarez, C.S.; Garland, L.; et al. Germline variants in hereditary breast cancer genes are associated with early age at diagnosis and family history in Guatemalan breast cancer. *Breast Cancer Res. Treat.* **2021**, *189*, 1–7. [[CrossRef](#)] [[PubMed](#)]
47. Rogoża-Janiszewska, E.; Malińska, K.; Górski, B.; Scott, R.J.; Cybulski, C.; Kluźniak, W.; Lener, M.; Jakubowska, A.; Gronwald, J.; Huzarski, T.; et al. Prevalence of germline TP53 variants among early-onset breast cancer patients from Polish population. *Breast Cancer* **2021**, *28*, 226–235. [[CrossRef](#)]
48. Wise, A.; Nguyen, T.; Herring, A.; North, K.E.; Siega-Riz, A.M.; Stuebe, A.M. Maternal Genotype and Gestational Diabetes. *Am. J. Perinatol.* **2013**, *31*, 069–076. [[CrossRef](#)] [[PubMed](#)]
49. Lambrinouadaki, I.; Vlachou, S.A.; Creatas, G. Genetics in gestational diabetes mellitus: Association with incidence, severity, pregnancy outcome and response to treatment. *Curr. Diabetes Rev.* **2010**, *6*, 393–399. [[CrossRef](#)]
50. Robitaille, J.; Grant, A.M. The genetics of gestational diabetes mellitus: Evidence for relationship with type 2 diabetes mellitus. *Genet. Med.* **2008**, *10*, 240–250. [[CrossRef](#)]
51. Chen, Y.; Liao, W.; Roy, A.; Loganath, A.; Ng, S. Mitochondrial gene mutations in gestational diabetes mellitus. *Diabetes Res. Clin. Pract.* **2000**, *48*, 29–35. [[CrossRef](#)]
52. Khan, I.A.; Jahan, P.; Hasan, Q.; Rao, P. Angiotensin-converting enzyme gene insertion/deletion polymorphism studies in Asian Indian pregnant women biochemically identifies gestational diabetes mellitus. *J. Renin-Angiotensin-Aldosterone Syst.* **2014**, *15*, 566–571. [[CrossRef](#)]
53. Singh, A.; Srivastava, N.; Amit, S.; Prasad, S.; Misra, M.; Ateeq, B. Association of AGTR1 (A1166C) and ACE (I/D) Polymorphisms with Breast Cancer Risk in North Indian Population. *Transl. Oncol.* **2018**, *11*, 233–242. [[CrossRef](#)]
54. Witczak, M.; Wilczyński, J.; Gulczyńska, E.; Talar, T.; Mordalska, A.; Łopaczyńska, D.; Ferenc, T. What is the impact of gestational diabetes mellitus on frequency of structural chromosome aberrations in pregnant women and their offspring? *Mutat. Res. Genet. Toxicol. Environ. Mutagenesis* **2017**, *818*, 27–30. [[CrossRef](#)] [[PubMed](#)]
55. Cnattingius, S.; Torráng, A.; Ekblom, A.; Granath, F.; Petersson, G.; Lambe, M. Pregnancy characteristics and maternal risk of breast cancer. *Jama* **2005**, *294*, 2474–2480. [[CrossRef](#)] [[PubMed](#)]
56. Van den Brandt, P.A.; Spiegelman, D.; Yaun, S.-S.; Adami, H.-O.; Beeson, L.; Folsom, A.R.; Fraser, G.R.; Goldbohm, A.; Graham, S.; Kushi, L.; et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am. J. Epidemiol.* **2000**, *152*, 514–527. [[CrossRef](#)] [[PubMed](#)]
57. Xie, S.; Yang, S.-J.; Chen, J.; Cheng, J.-Q. Gestational Diabetes Mellitus and Subsequent Cancer Risk: Shared Risk Factors, Causality or Confounding? *Asian Pac. J. Cancer Prev.* **2015**, *15*, 11001. [[CrossRef](#)] [[PubMed](#)]

## **Annexe 4**



→ Nbre Antécédents familiaux de cancer du sein 1 et 2 : 1<sup>er</sup> degré \_\_\_/\_\_\_ 2<sup>ème</sup> degré \_\_\_/\_\_\_  
0 sans atcd ; 1<sup>er</sup> degré (mère, sœur, fille) ; 2<sup>ème</sup> degré (grand-mère ; tante et cousine maternelles, nièce, petite-fille ...) ; 88 enquête effectuée/non interrogeable ; 87 sans information DCD ; 99 inconnu

→ Age atcd familiaux de cancer du sein 1 et 2 : \_\_\_/\_\_\_ \_\_\_/\_\_\_

→ Résultat des facteurs de risque génétiques (si enquête effectuée) :

- BRCA1 :  1-positif  2-négatif  9- en attente
- BRCA2 :  1-positif  2-négatif  9- en attente
- Autres :  1-positif  2-négatif  9- en attente - si oui, en clair : \_\_\_\_\_

→ Antécédents personnels néoplasiques 1 et 2 : \_\_\_/\_\_\_ \_\_\_/\_\_\_

00 aucun atcd; 01 sein controlatéral; 02 ovaire; 03 poumon; 04 endomètre; 05 thyroïde; 06 CIN3 CIS ; 08 col utérin INF; 16 colon rectum; 32 autre INF; 33 autre CIS ; 64 CIS homolatéral ; 88 autre (codage GC); 99 inc

Années atcd personnels néoplasiques 1 et 2 : \_\_\_/\_\_\_/\_\_\_/\_\_\_ \_\_\_/\_\_\_/\_\_\_/\_\_\_

→ Pathologies associées : \_\_\_/\_\_\_/\_\_\_/\_\_\_ \_\_\_/\_\_\_/\_\_\_/\_\_\_ \_\_\_/\_\_\_/\_\_\_/\_\_\_

00 aucune pathologie associée; 01 thyroïde; 02 traitement anxiolytique; 04 traitement hormonal prolongé; 08 obésité; 16 tuberculose; 32 HTA; 64 autre; 128 anomalie génétique; 256 maladie familiale; 512 psychiatrique; 666 maladie neurologique (Alzheimer, Parkinson, épilepsie...); 888 diabète; 444 vieillissement (majeur); 555 atcd modifiant les traitements loco-régionaux possibles; 777 atcd modifiant les traitements généraux possibles

Tabagisme :  1-fumeur  2-non fumeur  3-ancien fumeur  9-non mentionné

Nombre de paquets-années : \_\_\_/\_\_\_ 99 renseignement non connu 98 pour > 100 paquets année

CIM10 : \_\_\_/\_\_\_/\_\_\_/\_\_\_ Histologie : \_\_\_\_\_ ADICAP : \_\_\_/\_\_\_/\_\_\_/\_\_\_

Stade pTNM : pT \_\_\_ pN \_\_\_ pM \_\_\_ pTNM' : pT \_\_\_ pN' \_\_\_ pM \_\_\_

Ttt Néo-adjuvant :  1-oui  2-non  9-inc Stade ypTNM : ypT \_\_\_ ypN \_\_\_ ypM \_\_\_

Siège de la lésion :  0-mamelon/aréole  1-région centrale/rétromamelonnaire  2-Q supéro-interne/union des Q internes/union des Q sup  3-Q inféro-interne/union des Q inf  4-Q supéro-externe/union des Q externes  5-Q inféro-externe  6-prolongement axillaire  8-région profonde du sein/plusieurs quadrants atteints/sillon sous-mammaire  9-sein SAI

Nbre total de ganglions prélevés : \_\_\_/\_\_\_ Nbre total de ganglions envahis : \_\_\_/\_\_\_

Technique ganglion sentinelle :  1-oui  2-non  8-échec  9-non précisé

→ repérage:  1-bleu  2-sonde  3-bleu + sonde  9-non précisé

→ site injection:  1-périaréolaire  2-péritumorale  9-non précisé

→ lymphoscintigraphie :  1-oui  2-non  8-échec  9-non précisé

Si oui,  1-foyer axillaire  Nbre de foyer : \_\_\_/\_\_\_

2-foyer mammaire interne

3-foyer axillaire + chaîne mammaire interne  9-non précisé

→ nbre de ganglions prélevés : \_\_\_/\_\_\_

→ examen extemporané :  1-oui  2-non  9-non précisé

si oui, envahissement ganglion sentinelle :  1-oui  2-non  9-non précisé

→ examen définitif, envahissement ggl sentinelle :  1-oui  2-non. nbre envahi : \_\_\_/\_\_\_

si oui, type :  1- amas cellulaire (<= 0,2 mm)  2- micrométastase (0,2 - 2 mm)

3- métastase (> 2 mm)  4- oui SAI

si oui, atteinte multifocale dans le ganglion :  1-oui  2-non  9-non précisé

Rupture capsulaire :  1-oui  2-non  9-non précisé

Emboles vasculaires sanguin / lymphatique :  1-oui  2-non  9-non précisé

Taille macroscopique de la tumeur (mm) : \_\_\_/\_\_\_/\_\_\_

Emboles carcinomateux:  1-oui  2-non  9-non précisé

Limite d'exérèse :  1-berges saines  2-berges atteintes  3-berges limites <1mm  9-inconnu

Nombre de foyers :  1-unifocale  2-plurifocale micro  3-plurifocale macro  9-inconnu

➤ **Cancer infiltrant** :  1-oui  2-non  9-non précisé

Si oui, type de cancer infiltrant :  1-canaulaire  2-lobulaire  3-autre  9-inconnu

autre, type : \_\_\_\_\_

Grade histopronostique (SBR, EE ou grade de Nottingham) :  
 1-grade 1  2-grade 2  3-grade 3  8-non applicable  9-inconnu

- Différenciation tissulaire : \_\_\_/ - Pléiomorphisme tissulaire : \_\_\_/ - Mitoses \_\_\_/ - Index mitotique : \_\_. \_\_

➤ **Contingent in situ associé** :  1-oui  2-non  9-non précisé

Si oui, type canalaire :  1-oui  2-non  9-non précisé **grade nucléaire** : \_\_\_/

type lobulaire :  1-oui  2-non  9-non précisé

**Extension loco-régionale** :  1-oui  2-non  9-non précisé

**Extension type** :  04 mamelon/peau micro  08 peau/mamelon macro  32 muscle

➤ **Contingent in situ pur** :  1-oui  2-non  9-non précisé

Si oui, type canalaire :  1-oui  2-non  9-non précisé **grade nucléaire** : \_\_\_/

type lobulaire :  1-oui  2-non  9-non précisé

**Autres lésions associées** :  1-oui  2-non  9-non précisé

Si oui, type : \_\_\_\_\_

**DIAGNOSTIC** Date du diagnostic : \_\_\_/\_\_\_/\_\_\_\_\_/ **Date imagerie** : \_\_\_/\_\_\_/\_\_\_\_\_/

**Type imagerie** :  1mammo +-écho  2écho seule  3 IRM  4 autre  9 inconnu **Inclusion DO** :  oui  non  inc

**Circonstances du diagnostic** :  1-clinique  2-imagerie systématique  3-dépistage organisé  
 4-dépistage individuel  5-chir. prophylactique  6-diag. sur métastase  7-autre  9-inconnu

**Imagerie** :  1-désorganisation architecturale  2-nodule  3-micro-calcifications isolées  4-autre

Autre en clair : \_\_\_\_\_

**Base du diagnostic** :  0-certificat de décès uniquement  1-clinique  2-investigation clinique  
 3-chirurgie exploratrice / autopsie  4-marqueurs tumoraux spécifiques / biologie  5-cytologie  
 6-histologie des métastases  7-histologie d'une tumeur primaire  8-autopsie avec histologie  9-inconnu

**Cs oncogériatrie** :  1-oui  2-non  9-inc. **Statut OMS** : \_\_\_/ **Poids** : \_\_\_/\_\_\_/ kg **Taille** : \_\_\_/\_\_\_/ cm

**CLINIQUE** Côté :  1 droit  2 gauche  3 bilat. Drt (NI)  4 bilat. Gche (NI)  9-inc. **Bonnet** : \_\_\_/\_\_\_/\_\_\_/

**Taille clinique (mm)** : \_\_\_/\_\_\_/\_\_\_/ **Stade TNM** : T \_\_\_/ N \_\_\_/ M \_\_\_/

**Type de métastase 1 et 2** :

<input type="checkbox"/> 1-os	<input type="checkbox"/> 4-cerveau	<input type="checkbox"/> 7-plèvre	<input type="checkbox"/> 10-péritoine	<input type="checkbox"/> 99-inconnu
<input type="checkbox"/> 2-poumons	<input type="checkbox"/> 5-peau	<input type="checkbox"/> 8-multiples	<input type="checkbox"/> 11-œil/orbite	<input type="checkbox"/> 98-à préciser
<input type="checkbox"/> 3-foie	<input type="checkbox"/> 6-ganglions	<input type="checkbox"/> 9-moelle	<input type="checkbox"/> 12-autres	

➔ **Récepteurs hormonaux** :

- RO :  1-**P**  2-**N**  8-non déterminés  9-inconnu % : \_\_\_/\_\_\_/ **Intensité** : \_\_\_/

- RP :  1-**P**  2-**N**  8-non déterminés  9-inconnu % : \_\_\_/\_\_\_/ **Intensité** : \_\_\_/

➔ **Méthode** :  1-immunohistochimique  2-biochimique  8-non déterminés  9-inconnu

➔ **Expression de HER2** :  0-0  1-1+  2-2+  3-3+  8-non fait

**Méthode FISH** :  1 amplifié  2 non amplifié  3 non assuré  8 non fait  9 inconnu

**Méthode TOPO2A** :  1 amplifié  2 non amplifié  3 non assuré  8 non fait  9 inconnu

➔ **% Ki67** : \_\_\_/\_\_\_/ **Autres marqueurs hormonaux recherchés** :  1-oui  2-non  9-non fait

si oui, le(s)quel(s) : \_\_\_\_\_

## CHIRURGIE

Chirurgie :  1-oui  2-non  9-inconnu

Si non, raison :  1-refus  2-contre-indication  3-non indiquée  4-autre  9-non spécifié

Essai clinique :  1-oui  2-non  9-inconnu Si oui, nom de l'essai : \_\_\_\_\_

Type chirurgie :	Date :	Nom ébt / dptmt :	Type :	Nom CHIR :
1/ Micro-biopsie	__/__/____/		__/	
2/ Macro-biopsie	__/__/____/		__/	
3/ Tumorectomie	__/__/____/		__/	
4/ Mastectomie radicale	__/__/____/		__/	
5/ Mastectomie sous-cutanée	__/__/____/		__/	
6/ Halsted	__/__/____/		__/	
7/ Chirurgie non précisée	__/__/____/		__/	
8/ Reconstruction mammaire	__/__/____/		__/	
9/ Curage ganglionnaire non précisé	__/__/____/		__/	
10/ Curage axillaire	__/__/____/		__/	
11/ Curage axillaire sus-claviculaire	__/__/____/		__/	
12/ Curage chaîne mammaire interne	__/__/____/		__/	
13/ Chirurgie de(s) métastase(s)	__/__/____/		__/	
14/ Reprise chirurgicale	__/__/____/		__/	
15/ Ganglion sentinelle	__/__/____/		__/	

## CHIMIOThERAPIE

Poids : \_\_/\_\_/\_\_/ kg

Taille : \_\_/\_\_/\_\_/ cm

Facteurs de croissance :  1-oui  2-non  9-non précisé

Type d'établissement :  01-CHU  02-CHG  03-CLCC  04-clinique privée  99-inconnu

Oncologue : \_\_\_\_\_

Nom de l'établissement : \_\_\_\_\_ département : \_\_/\_\_/

NEO-ADJUVANTE :  1-oui  2-non  9-non précisé Nbre de cycles : \_\_/\_\_/

Si non, raison :  1-refus  2-contre-indication  3-non indiquée  4-autre  9-non spécifié

Date du 1er cycle : \_\_/\_\_/\_\_\_\_/ Date du dernier cycle : \_\_/\_\_/\_\_\_\_/

Type de chimio (nom du protocole) : \_\_\_\_\_

Essai clinique :  1-oui  2-non  9-inconnu Si oui, nom de l'essai : \_\_\_\_\_

Drogue	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8

**ADJUVANTE** :  1-oui  2-non  9-non précisé Nbre de cycles : \_\_/\_\_/

Si non, raison :  1-refus  2-contre-indication  3-non indiquée  4-autre  9-non spécifié

Date du 1er cycle : \_\_/\_\_/\_\_\_\_\_/ Date du dernier cycle : \_\_/\_\_/\_\_\_\_\_/

Type de chimio (nom du protocole) : \_\_\_\_\_

Essai clinique : 1-oui 2-non 9-inconnu Si oui, nom de l'essai : \_\_\_\_\_

Drogue	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8

### **RADIOTHERAPIE**

Radiothérapie :  1-oui  2-non  9-non précisé

Si non, raison :  1-refus  2-contre-indication  3-non indiquée  4-autre  9-non spécifié

Essai clinique : 1-oui 2-non 9-inconnu Si oui, nom de l'essai : \_\_\_\_\_

Dose totale (en GY)	Cible	Cible en clair	Date de début	Date de fin
__/__/	__/__/		__/__/_____/	__/__/_____/
__/__/	__/__/		__/__/_____/	__/__/_____/

Type d'établissement :  01-CHU  02-CHG  03-CLCC  04-clinique privée  99-inconnu

Radiothérapeute : \_\_\_\_\_

Nom de l'établissement : \_\_\_\_\_ département : \_\_/\_\_/

### **HORMONOTHERAPIE**

Hormonothérapie :  1-oui  2-non  9-non précisé

Si non, raison :  1-refus  2-contre-indication  3-non indiquée  4-autre  9-non spécifié

Essai clinique : 1-oui 2-non 9-inconnu Si oui, nom de l'essai : \_\_\_\_\_

Type	Date début	Date fin
__/__/	__/__/_____/	__/__/_____/
__/__/	__/__/_____/	__/__/_____/
__/__/	__/__/_____/	__/__/_____/

Type d'établissement :  01-CHU  02-CHG  03-CLCC  04-clinique privée  99-inconnu

Nom de l'établissement : \_\_\_\_\_ département : \_\_/\_\_/

Type d'hormonothérapie :

80 Castration chirurgicale, 81 Castration radioT, 82 Anti-œstrogène (tamoxifène) : Nolvadex®, 83 Permastril, 84 Nolvadex + Permastril, 85 Progestatif : Norfor®, Prodasone® ou Dépo-prodasone®, Farlutal®, 86 Œstrogène (Distilbène®...), 87 Castration ChimioT, 88 Nolvadex + progestérone, 90 Anti-œstrogène (torémifène): Fareston®, 91 Antiaromatase (anastrozole): Arimidex®, 92 Antiaromatase type I : Aromasine®, 93 Antiaromatase (létrozole): Femara®, 94 Agoniste LH-RH : Zoladex®, 95 antioestrogènes : Faslodex (Fulvestrant), 96 Agoniste LH-RH (Decapeptyl), 97 Analogue GnRH : Enantone (Leuproréline), 99 inc

Clé fin traitement : \_\_/\_\_/

Date absence évolution : \_\_/\_\_/\_\_\_\_\_/

Date recueil : \_\_/\_\_/\_\_\_\_\_/

### SOURCES

CRAP :  1-oui  2-non  9-non précisé

-Labo CRAP Biopsie  1-CHU  2-Chalon  3-CLCC  4-CP  5- autre  6- mixte  8-non applicable  9-inc

-Labo CRAP CHIR  1-CHU  2-Chalon  3-CLCC  4-CP  5- autre  6- mixte  8-non applicable  9-inc

RCP  1-oui  2-non  9-non précisé

Autre source de signalement (CD, PMSI, ALD...) :  1-oui  2-non  9-non précisé

Participation Etude : CANTO  1-oui  2-non Autre : \_\_\_\_\_

Refus étude REGISTRE :  1-oui  2-non étude/date : \_\_\_\_\_

### BIOLOGIE RECHUTE

→ Récepteurs hormonaux :

- RO :  1-  2-  8-non déterminés  9-inconnu % : \_\_\_ / Intensité : \_ /

- RP :  1-  2-  8-non déterminés  9-inconnu % : \_\_\_ / Intensité : \_ /

→ HER2 :  0-0  1-1+  2-2+  3-3+  8-non fait FISH :  1 AMP  2 NON AMP  3 non assuré  8 non fait

### RECHUTES ET TRAITEMENT DES RECHUTES

→ Rechute 1 : \_\_\_/\_\_\_/\_\_\_\_\_/

Type de rechute :  1-Récidive locale/régionale  2-Métastases à distance  3-Récidive mixte

Type de métastase 1 et 2 :

<input type="checkbox"/> 1-os	<input type="checkbox"/> 4-cerveau	<input type="checkbox"/> 7-plèvre	<input type="checkbox"/> 10-péritoine	<input type="checkbox"/> 98- à préciser
<input type="checkbox"/> 2-poumons	<input type="checkbox"/> 5-peau	<input type="checkbox"/> 8-multiples	<input type="checkbox"/> 11-œil/orbite	<input type="checkbox"/> 99- inconnu
<input type="checkbox"/> 3-foie	<input type="checkbox"/> 6-ganglions	<input type="checkbox"/> 9-moelle	<input type="checkbox"/> 12-autres	

→ Rechute 2 : \_\_\_/\_\_\_/\_\_\_\_\_/

Type de rechute :  1-Récidive locale/régionale  2-Métastases à distance  3-Récidive mixte

Type de métastase 1 et 2 :

<input type="checkbox"/> 1-os	<input type="checkbox"/> 4-cerveau	<input type="checkbox"/> 7-plèvre	<input type="checkbox"/> 10-péritoine	<input type="checkbox"/> 98- à préciser
<input type="checkbox"/> 2-poumons	<input type="checkbox"/> 5-peau	<input type="checkbox"/> 8-multiples	<input type="checkbox"/> 11-œil/orbite	<input type="checkbox"/> 99- inconnu
<input type="checkbox"/> 3-foie	<input type="checkbox"/> 6-ganglions	<input type="checkbox"/> 9-moelle	<input type="checkbox"/> 12-autres	

→ Rechute 3 : \_\_\_/\_\_\_/\_\_\_\_\_/

Type de rechute :  1-Récidive locale/régionale  2-Métastases à distance  3-Récidive mixte

Type de métastase 1 et 2 :

<input type="checkbox"/> 1-os	<input type="checkbox"/> 4-cerveau	<input type="checkbox"/> 7-plèvre	<input type="checkbox"/> 10-péritoine	<input type="checkbox"/> 98- à préciser
<input type="checkbox"/> 2-poumons	<input type="checkbox"/> 5-peau	<input type="checkbox"/> 8-multiples	<input type="checkbox"/> 11-œil/orbite	<input type="checkbox"/> 99- inconnu
<input type="checkbox"/> 3-foie	<input type="checkbox"/> 6-ganglions	<input type="checkbox"/> 9-moelle	<input type="checkbox"/> 12-autres	

→ Rechute 4 : \_\_\_/\_\_\_/\_\_\_\_\_/

Type de rechute :  1-Récidive locale/régionale  2-Métastases à distance  3-Récidive mixte

Type de métastase 1 et 2 :

<input type="checkbox"/> 1-os	<input type="checkbox"/> 4-cerveau	<input type="checkbox"/> 7-plèvre	<input type="checkbox"/> 10-péritoine	<input type="checkbox"/> 98- à préciser
<input type="checkbox"/> 2-poumons	<input type="checkbox"/> 5-peau	<input type="checkbox"/> 8-multiples	<input type="checkbox"/> 11-œil/orbite	<input type="checkbox"/> 99- inconnu
<input type="checkbox"/> 3-foie	<input type="checkbox"/> 6-ganglions	<input type="checkbox"/> 9-moelle	<input type="checkbox"/> 12-autres	

# **Annexe 5**

Section and Topic	Item #	Checklist item	Location where item is reported
<b>TITLE</b>			
Title	1	Identify the report as a systematic review.	
<b>ABSTRACT</b>			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
<b>METHODS</b>			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	

Section and Topic	Item #	Checklist item	Location where item is reported
Synthesis methods	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	
<b>RESULTS</b>			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	

Section and Topic	Item #	Checklist item	Location where item is reported
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
<b>DISCUSSION</b>			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	
	23b	Discuss any limitations of the evidence included in the review.	
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	
<b>OTHER INFORMATION</b>			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71  
For more information, visit: <http://www.prisma-statement.org/>

# **Annexe 6**

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE  
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

- 1) Is the case definition adequate?
  - a) yes, with independent validation \*
  - b) yes, eg record linkage or based on self reports
  - c) no description
- 2) Representativeness of the cases
  - a) consecutive or obviously representative series of cases \*
  - b) potential for selection biases or not stated
- 3) Selection of Controls
  - a) community controls \*
  - b) hospital controls
  - c) no description
- 4) Definition of Controls
  - a) no history of disease (endpoint) \*
  - b) no description of source

Comparability

- 1) Comparability of cases and controls on the basis of the design or analysis
  - a) study controls for \_\_\_\_\_ (Select the most important factor.) \*
  - b) study controls for any additional factor \* (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

- 1) Ascertainment of exposure
  - a) secure record (eg surgical records) \*
  - b) structured interview where blind to case/control status \*
  - c) interview not blinded to case/control status
  - d) written self report or medical record only
  - e) no description
- 2) Same method of ascertainment for cases and controls
  - a) yes \*
  - b) no

- 3) Non-Response rate
  - a) same rate for both groups \*
  - b) non respondents described
  - c) rate different and no designation

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE  
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

- 1) Representativeness of the exposed cohort
  - a) truly representative of the average \_\_\_\_\_ (describe) in the community \*
  - b) somewhat representative of the average \_\_\_\_\_ in the community \*
  - c) selected group of users eg nurses, volunteers
  - d) no description of the derivation of the cohort
- 2) Selection of the non exposed cohort
  - a) drawn from the same community as the exposed cohort \*
  - b) drawn from a different source
  - c) no description of the derivation of the non exposed cohort
- 3) Ascertainment of exposure
  - a) secure record (eg surgical records) \*
  - b) structured interview \*
  - c) written self report
  - d) no description
- 4) Demonstration that outcome of interest was not present at start of study
  - a) yes \*
  - b) no

Comparability

- 1) Comparability of cohorts on the basis of the design or analysis
  - a) study controls for \_\_\_\_\_ (select the most important factor) \*
  - b) study controls for any additional factor \* (This criteria could be modified to indicate specific \_\_\_\_\_ control for a second important factor.)

## Outcome

### 1) Assessment of outcome

- a) independent blind assessment \*
- b) record linkage \*
- c) self-report
- d) no description

### 2) Was follow-up long enough for outcomes to occur

- a) yes (select an adequate follow up period for outcome of interest) \*
- b) no

### 3) Adequacy of follow up of cohorts

- a) complete follow up - all subjects accounted for \*
- b) subjects lost to follow up unlikely to introduce bias - small number lost - > \_\_\_\_\_ % (select an adequate %) follow up, or description provided of those lost) \*
- c) follow up rate < \_\_\_\_\_ % (select an adequate %) and no description of those lost
- d) no statement



**Titre :** Abord de l'épidémiologie des cancers par des données de registre et des données médico-administratives : complémentarité des deux approches.

**Mots clés :** épidémiologie, cancer, registre, base médico-administrative, avantages, limites

**Résumé :** À partir des données des registres, qui couvrent environ 20% de la population, les tendances des estimations d'incidence nationale et régionales sont publiées régulièrement. Les études de survie et les publications de l'incidence selon les facteurs pronostiques contribuent à l'évaluation de l'impact du dépistage ou de nouveaux traitements des cancers.

Les données médico-administratives sont de plus en plus utilisées, à la fois à des fins de recherche et de surveillance sanitaire. L'intérêt potentiel de ces bases de données dans une optique épidémiologique apparaît clairement dans la mesure où elles fournissent des données individuelles médicalisées, structurées et codées de manière standardisée. Toutefois, ayant été construites pour répondre à des objectifs médico-économiques, leur utilisation à des fins épidémiologiques nécessite un travail important de contrôle et de validation des données, et de nombreux algorithmes et études de validation ont été publiés ces dernières années.

Dans une première partie analysant les apports et limites des registres de cancer, nous avons effectué une première étude à partir des données du Registre des cancers du sein et autres cancers gynécologiques de Côte d'Or.

Dans une seconde partie analysant les apports et limites des bases médico-administratives, nous avons effectué une seconde étude à partir des données du Système National des Données de Santé.

Dans une dernière partie, nous avons effectué une revue de la littérature composée d'études utilisant des données issues de registres ou de bases médico-administratives. Cette revue de la littérature et la description de plusieurs exemples de chaînage entre des bases médico-administratives et des registres de cancer nous ont permis d'analyser les différences et la complémentarité qu'il peut exister entre ces deux approches.

**Title:** Epidemiological studies of cancer using registry data and medico-administrative data: complementarity of the two approaches.

**Keywords:** epidemiology, cancer, registry, medico-administrative database, advantages, limits.

**Abstract:** Based on registry data, which cover approximately 20% of the population, trends in national and regional incidence estimates are published regularly. Survival studies and reports of incidence by prognostic factors can contribute to the evaluation of the impact of cancer screening and new treatments.

Medico-administrative data are increasingly used for both research and health surveillance purposes. The potential value of these databases from an epidemiological point of view is clear insofar as they provide individualized medical data that is structured and coded in a standardized way. However, having been built to meet medico-economic objectives, their use for epidemiological purposes requires significant data control and validation work, and many algorithms and validation studies have been published in recent years.

First, we conducted a study using data from the Côte d'Or Breast and other gynaecological cancer registry in order to analyse the contributions and limitations of cancer registries.

We then conducted a second study using data from the "Système National des Données de Santé" (SNDS), which is the French national health data system that contains individual, exhaustive and linkable but anonymous data on health expenditures for about ninety-nine percent of the French population, in order to analyse the contributions and limitations of medico-administrative databases.

The last part of the investigation was a literature review of studies using data from registries or medico-administrative databases. We used this review of the literature and several examples of chaining between medico-administrative databases and cancer registries to analyse the differences between these two approaches and their potential complementarity.