



THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Eloïse BERGER

Le 23 juin 2021

**Construction des inégalités sociales de cancer du sein :
mécanismes impliqués dans l'incidence et la présentation initiale
de la maladie**

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et
Télécommunications de Toulouse**

Spécialité : **Epidémiologie**

Unité de recherche :

**CERPOP - Centre d'Epidémiologie et de Recherche en santé des POPulations
de Toulouse**

Thèse dirigée par
Cyrille DELPIERRE

Jury

Mme Virginie RINGA, Rapporteure

M. Pascal GUÉNEL, Rapporteur

M. Guy LAUNOY, Examinateur

Mme Sylvana CÔTÉ, Examinatrice

M. Cyrille DELPIERRE, Directeur de thèse

Mme Raphaële CASTAGNÉ, Co-directrice de thèse

Mme Gwenn MENVIELLE, Présidente

RESUME

Des inégalités sociales de mortalité par cancer du sein ont été mises en évidence dans la littérature. Le cancer du sein est un cancer relativement fréquent, bien documenté, pour lequel un nombre important de déterminants a été identifié. Au fil des années de plus en plus d'études se sont intéressées à l'impact de ces déterminants sur le gradient social d'incidence de cancer du sein. Cependant, ces déterminants ne suffisent pas à eux seuls à expliquer les inégalités d'incidence observées suggérant que d'autres mécanismes sont en jeu.

Nous faisons l'hypothèse que les inégalités sociales observées dans la mortalité par cancer du sein sont la conséquence d'une distribution socialement différenciée de l'incidence de la maladie mais aussi de ses caractéristiques initiales, et de l'accès au système de soin. Cette thèse propose d'étudier le lien entre la position socio-économique (PSE) et l'incidence du cancer du sein ainsi que ses caractéristiques cliniques en considérant un large ensemble de variables de médiation potentielles.

L'étude du lien entre la PSE et le fonctionnement biologique a fait l'objet de mon premier travail de thèse. Ce travail a mis en évidence un lien entre un faible niveau d'éducation et une inflammation plus élevée, indépendamment de la PSE au début de la vie ou à l'âge adulte, ainsi que des facteurs comportementaux et de l'indice de masse corporelle. J'ai ensuite étudié le lien qui pouvait exister entre PSE, inflammation et risque de cancer. J'ai approfondi cette exploration dans le champ des cancers à travers trois travaux menés au sein de la cohorte prospective française E3N (N = 98 995 femmes dont 7 877 cas de cancer du sein), dans lesquels j'ai cherché à étudier la façon dont les facteurs sociaux, biologiques, comportementaux, anthropométriques et reproductifs interagissent à différents moments de la vie pour favoriser la survenue des cancers du sein et influencer leurs caractéristiques au moment du diagnostic.

Dans un travail qui s'est intéressé à l'incidence de la maladie, mes résultats ont mis en évidence un risque de cancer du sein plus élevé chez les femmes éduquées, largement expliqué par un âge plus tardif à la première grossesse. Cependant, cette association n'était pas totalement expliquée par la prise en compte d'un nombre important de facteurs de risque, suggérant que d'autres déterminants ou mécanismes sont en jeu. Parmi ces mécanismes, nous avons étudié le lien entre le niveau d'éducation et des biomarqueurs.

Résumé

Un second travail a ainsi permis de mettre en évidence que certains biomarqueurs [la protéine C-réactive, les lipoprotéines de haute densité, l'œstradiol, la testostérone et les protéines de liaison des hormones sexuelles] étaient associés au risque de cancer du sein, indépendamment des facteurs comportementaux, anthropométriques et reproductifs, et que ces biomarqueurs étaient socialement distribués. Le lien entre l'éducation et ces biomarqueurs était en partie expliqué par les facteurs de risque considérés sauf en ce qui concerne la testostérone, suggérant que la relation entre l'éducation et cette hormone sexuelle pourrait constituer un mécanisme nouveau impliqué dans le gradient d'incidence.

Concernant les caractéristiques des cancers, un troisième travail nous a permis de montrer que les femmes moins éduquées, bien que moins à risque de développer un cancer du sein, sont plus à risque de développer des tumeurs agressives et un sous type de moins bon pronostic.

Ce travail doctoral souligne l'importance des facteurs reproductifs sur le risque de cancer du sein mais met également en évidence que d'autres mécanismes d'incorporation biologique de l'environnement social, en plus des facteurs de risque classiques, pourraient être impliqués dans les inégalités sociales d'incidence et de profils de tumeurs. D'autres travaux sont nécessaires pour mieux caractériser le rôle de l'incorporation biologique, notamment le recours à des analyses de médiation, et analyser les effets sur la survie.

Mots clés : Cancer du sein ; inégalités sociales ; incorporation biologique ; déterminants sociaux ; facteurs pronostics ; sous-types moléculaires ; approche life course

ABSTRACT

Social inequalities in breast cancer mortality have been identified in the literature. Breast cancer is a relatively common, well-documented cancer for which a large number of determinants have been identified. A growing number of studies have focused on the impact of these determinants on the social gradient of breast cancer incidence. However, these determinants are not sufficient to explain the observed inequalities of incidence, suggesting that other mechanisms are at play.

We hypothesise that the observed social inequalities in breast cancer mortality are the consequence of a socially differentiated distribution of the incidence of the disease but also of its initial characteristics, and of access to the health care system. This thesis proposes to investigate the link between socioeconomic position (SEP) and breast cancer incidence and clinical characteristics by considering a broad set of potential mediating variables.

The study of the link between SEP and biological functioning was the subject of my first thesis work. This work showed a link between low education and higher inflammation, independent of SEP in early life or in adulthood, as well as behavioural factors and body mass index. I then investigated the relationship between SEP, inflammation and cancer risk. I extended this exploration into the field of cancer through three studies conducted within the French prospective cohort E3N (N = 98,995 women, including 7,877 breast cancer cases), in which I sought to study the way in which social, biological, behavioural, anthropometric and reproductive factors interact at different times of life to promote the occurrence of breast cancers and influence their characteristics at the time of diagnosis.

In a work that looked at the incidence of the disease, my results showed a higher risk of breast cancer in educated women, largely explained by a later age at first childbirth. However, this association was not fully explained by taking into account a large number of risk factors, suggesting that other determinants or mechanisms are at play. Among these mechanisms, we studied the link between the educational attainment and some biomarkers.

A second study showed that some biomarkers [C-reactive protein, high-density lipoprotein, estradiol, testosterone and sex hormone binding globulin] were associated with breast cancer risk, independently of behavioural, anthropometric and reproductive factors, and that these

biomarkers were socially distributed. The relationship between education and these biomarkers was partly explained by the considered risk factors except for testosterone, suggesting that the relationship between education and this sex hormone may be a novel mechanism involved in the incidence gradient.

Regarding cancer characteristics, a third work allowed us to show that less educated women, although less likely to develop breast cancer, were more likely to develop aggressive tumours and a subtype with a poorer prognosis.

This doctoral work highlights the importance of reproductive factors on breast cancer risk but also suggests that other mechanisms of biological embodiment of the social environment, in addition to classical risk factors, may be involved in social inequalities in tumour incidence and profiles. Further works are needed to better characterise the role of biological embodiment, including the use of mediation analyses, and to analyse the effects on survival.

Keywords: Breast cancer; social inequalities; biological embodiment; social determinants; prognostic factors; molecular subtypes; life course approach

REMERCIEMENTS

Je tiens tout d'abord à remercier très chaleureusement mes directeurs de thèse, Cyrille Delpierre et Raphaële Castagné, pour leur accompagnement depuis mon stage de master jusqu'à la finalisation de cette thèse. Merci pour votre disponibilité, pour votre bienveillance, pour vos conseils et pour toutes les choses que vous m'avez apprises au cours de ces cinq années.

Je tiens également à remercier tout particulièrement Gianluca Severi pour avoir été présent à chaque étape de cette thèse pour échanger sur les données et les résultats.

Je remercie également Carlotta Sacerdote et Paolo Vineis pour leur implication dans les analyses de la cohorte EPIC-Italie.

Je remercie sincèrement Virginie Ringa et Pascal Guénel d'avoir accepté d'être rapporteurs de ce travail ainsi que Guy Launoy, Gwenn Menvielle et Sylvana Côté pour avoir accepté de faire partie du jury et d'évaluer ma thèse.

Un dernier grand merci à tous les membres de l'équipe Equity pour leur bienveillance, leur gentillesse et leur bonne humeur, ainsi que pour tous les moments conviviaux passés autour d'un pain au chocolat, d'une galette des rois, d'un feu de cheminée virtuel, ou d'une bière, bien trop rares en ce moment...

VALORISATION SCIENTIFIQUE

ARTICLES EN RAPPORT AVEC LA THESE

Berger E, Maitre N, Romana Mancini F, Baglietto L, Perduca V, Colineaux H, et al. The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk. *BMC Cancer*. 2020 Nov 23;20(1):1138.

Maurel M, Castagné R, **Berger E**, Bochud M, Chadeau-Hyam M, Fraga S, et al. Patterning of educational attainment across inflammatory markers: Findings from a multi-cohort study. *Brain Behav Immun*. 2020 Nov;90:303–10.

Berger E, Castagné R, Chadeau-Hyam M, Bochud M, d’Errico A, Gandini M, et al. Multi-cohort study identifies social determinants of systemic inflammation over the life course. *Nat Commun*. 2019 15;10(1):773.

Berger E, Delpierre C, Hosnijeh FS, Kelly-Irving M, Portengen L, Bergdahl IA, et al. Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts. *Sci Rep*. 2018 Jul 17;8(1):10805.

EN PRÉPARATION

Berger E, Dossus L, Perduca V, Baglietto L, Gelot A, Fournier A, Boutron-Ruault MC, Severi G, Castagné R, Delpierre C. Biological embodiment of education and future risk of breast cancer : findings from a French prospective cohort

Berger E, Boutron-Ruault MC, Severi G, Castagné R, Delpierre C. Educational attainment and breast cancer characteristics : findings from a French prospective cohort [soumission prévue d’ici juin 2021]

AUTRES PUBLICATIONS

Berger E, Delpierre C, Despas F, Bertoli S, Bérard E, Bombarde O, et al. Are social inequalities in acute myeloid leukemia survival explained by differences in treatment utilization? Results from a French longitudinal observational study among older patients. *BMC Cancer*. 2019 Sep 5;19(1):883.

Karimi M, Castagné R, Delpierre C, Albertus G, **Berger E**, Vineis P, et al. Early-life inequalities and biological ageing: a multisystem Biological Health Score approach in Understanding Society. *J Epidemiol Community Health*. 2019 Apr 3;jech-2018-212010.

COMMUNICATIONS ORALES ET AFFICHEES

COMMUNICATIONS ORALES EN RAPPORT AVEC LA THESE

Berger E, Maitre N, Mancini F, Krogh V, Panico S, Sacerdote C, Tumino R, Vineis P, Boutron-Ruault MC, Severi G, Castagné R, Delpierre C. Life course socio-economic position and individual social mobility on the future risk of breast cancer, Conférence annuelle de la Society for Longitudinal and Life course studies (SLLS) 2020 [reporté en septembre 2021 - conférence organisée en ligne].

Berger E, Krogh V, Palli D, Panico S, Sacerdote C, Tumino R, Vineis P, Severi G, Castagné R, Delpierre C. Socio-economic position across the life course and future risk of breast cancer : findings from EPIC-italy, 12ème conférence européenne de santé publique, EPHC - du 20 au 23 novembre 2019, Marseille, France.

Berger E, Delpierre C, Saberi Hosnijeh F, Kelly-Irving M, Portengen L, A. Bergdahl I, Kyrtopoulos SA, Vineis P, Chadeau-Hyam M, Vermeulen R, Castagné R. Pre-diagnostic inflammatory markers and risk of Breast cancer and Non-Hodgkin Lymphoma in two prospective cohorts; GRELL - Groupe des Registres et d'Épidémiologistes de Langue Latine; 24 au 26 mai 2017, Bruxelles, Belgique.

AUTRES COMMUNICATIONS ORALES

Berger E, Delpierre C, Despas F, Berard E, Bories P, Sarry A, Recher C, Lamy S. Position socio-économique et prise en charge des Leucémies Aigües Myéloïdes chez le sujet âgé, Colloque "Santé : équité ou égalité ? Définir, mesurer, agir », IFERISS : 23 au 25 mai 2018, Toulouse, France. [Résumé publié : DOI: 10.1016/j.respe.2018.12.036]

Berger E, Delpierre C, Despas F, Berard E, Bories P, Sarry A, Recher C, Lamy S, Position socio-économique et prise en charge des Leucémies Aigües Myéloïdes chez le sujet âgé. Congrès de la Société Française d'Hématologie (SFH): 28 au 30 mars 2018, Paris, France. [Communication invitée]

AUTRES COMMUNICATIONS AFFICHEES

Berger E, Delpierre C, Despas F, Berard E, Bories P, Sarry A, Recher C, Lamy S. Does patients' SEP influence therapeutic strategy in Acute Myeloid Leukaemia: Results from a longitudinal observational study, European Congress of Epidemiology: 4 au 6 juillet 2018, Lyon, France. [Résumé publié : DOI:10.1016/j.respe.2018.05.336]

Berger E, Delpierre C, Despas F, Berard E, Bories P, Sarry A, Recher C, Lamy S. La position sociale influence-t-elle le choix du traitement dans la Leucémie Aigüe Myéloïde du sujet âgé ? Une étude longitudinale en Midi-Pyrénées (France), EPICLIN 12 / 25ème Journées des Statisticiens des CLCC: 30 mai au 1 juin 2018, Nice, France

TABLE DES MATIERES

Résumé	1
Abstract	3
Remerciements	5
Valorisation scientifique	6
Articles en rapport avec la thèse.....	6
Autres publications	6
Communications orales et affichées	7
Table des matières	8
Liste des abréviations	12
Liste des figures	13
Liste des Tableaux	15
PARTIE 1 - INTRODUCTION GENERALE	17
Chapitre I. Inégalités sociales de santé : Définitions générales et concepts	18
1. Inégalités sociales de santé (ISS)	18
A. De quoi parle-t-on ?.....	18
B. Des inégalités sociales qui s’observent en terme d’espérance de vie et de mortalité	18
C. ...qui reflètent des inégalités dans l’incidence des maladies	21
D. Comment mesure-t-on les ISS ?	21
2. Des déterminants multiples qui s’enchainent et s’accumulent	26
3. L’approche biographique ou <i>life course epidemiology</i>	28
4. Incorporation biologique de l’environnement social	30
5. Mécanismes potentiels impliqués dans la construction des inégalités sociales de santé	32
A. Expositions d’origine exogène.....	33
B. Expositions d’origine endogène.....	34
C. Environnement social et modifications des processus biologiques	35
Travail 1 (box 1) : Déterminants sociaux et inflammation : étude dans six cohortes européennes... ..	36
Travail 2 (box 2): Inflammation chronique et développement du cancer du sein et des lymphomes non-Hodgkinien : résultats d’une étude cas-témoins	39
Chapitre II. Inégalités sociales observées dans la mortalité par cancers : que recouvrent-elles ?	43
1. Risque de cancer : conséquences à long terme des déterminants sociaux	43
A. Comportements de santé et facteurs anthropométriques	43
B. Expositions professionnelles, environnementales et matérielles	45
C. Stress et cancer.....	46
2. Des maladies qui n’expriment pas les mêmes caractéristiques initiales au moment du diagnostic	48
3. Une prise en charge dans les systèmes de santé différenciée	49
4. PSE et cancer : quelles spécificités ?	50
Chapitre III. Epidemiologie du cancer du sein	53
1. Le cancer du sein	53
A. Contexte général	53
B. Facteurs pronostics.....	54
C. Facteurs prédictifs et sous types de cancer du sein	56
D. Types histologiques	58

2. Incidence et facteurs de risque identifiés	59
A. Caractéristiques génétiques et familiales.....	59
B. Caractéristiques de la vie reproductive	59
C. Comportements de santé et facteurs anthropométriques	61
3. Facteurs de risque des cancers de mauvais pronostics.....	63
A. Les sous-types hormonaux dépendants	63
B. Les sous-types non dépendants aux hormones.....	63
C. Stade et grade.....	64
4. Inégalités sociales face au cancer du sein : incidence, facteurs pronostics et mortalité	64
A. Positions socio-économiques et incidence globale	65
B. Positions socio-économiques, stade, grade et sous types moléculaires	66
C. Positions socio-économiques et survie.....	69
5. Perspectives de recherche sur les inégalités sociales de cancer du sein	70
Chapitre IV. Objectifs et hypothèses de ce travail de thèse	72
1. Objectif général de ce travail de thèse	72
2. Objectifs et hypothèses des différents travaux.....	73
PARTIE 2 - MATERIEL ET METHODES.....	75
Chapitre I. La cohorte E3N	76
1. Caractéristiques générales.....	76
2. Recueil des données.....	76
3. Consentement des participants et déclaration éthique.....	77
Chapitre II. Les cancers dans E3N	78
1. Validation des pathologies	78
2. Définition des cas de cancer	78
3. Caractéristiques des cancers du sein	79
Chapitre III. Données sociales au cours de la vie et mobilité individuelle.....	81
1. Données sociales individuelles.....	81
A. L'activité professionnelle de la participante, de son père et son conjoint	81
B. L'éducation de la participante et de son conjoint	83
2. Mobilité sociale : de l'enfance à l'âge adulte	83
Chapitre IV. Marqueurs biologiques.....	85
1. Collecte des échantillons.....	85
2. Traitement des données biologiques	85
3. Définition des biomarqueurs utilisés pour les analyses	86
Chapitre V. Les covariables	87
1. Comportements de santé	89
A. Consommation de tabac.....	89
B. Consommation d'alcool et Adhérence au régime alimentaire occidental	89
C. Activité physique	90
2. Mesures anthropométriques	90
A. Taille	90
B. Poids	91
3. Facteurs reproductifs, antécédents familiaux et dépistage	91
A. Age à la première grossesse menée à terme.....	91
B. Contraceptif oral.....	91
C. Allaitement	91
D. Durée de vie reproductive.....	92
E. Statut ménopausique.....	92
F. Utilisation d'un traitement de la ménopause	92
G. Antécédents familiaux de cancers hormonaux-dépendants.....	93

G. Mammographie.....	93
Chapitre VI. Réception et gestion des données	94
Chapitre VII. Démarche d’analyses.....	96
1. Populations d’étude	96
A. Travail 3 : Gradient social de cancer du sein	96
B. Travail 4 : Incorporation biologique de l’éducation et risque de cancer du sein.....	96
C. Travail 5 : Environnement social et caractéristiques des cancers du sein	97
2. Analyses descriptives	98
3. Imputation multiple	98
4. Modèles multivariés.....	99
A. Modèles de Cox à risques proportionnels.....	99
B. Modèles de régressions logistiques multinomiales.....	101
PARTIE 3 - RESULTATS	102
Chapitre I. Impact de l’environnement social sur l’incidence du cancer du sein	103
TRAVAIL 3 : DETERMINANTS SOCIAUX DU RISQUE DE CANCER DU SEIN	103
1. Contexte.....	103
2. Méthode et schéma d’étude.....	105
3. Résultats.....	107
A. Caractéristiques de la population d’étude	107
B. Impact de chacune des PSE sur le risque de cancer du sein.....	110
C. PSE au cours de la vie et risque de cancer du sein	112
D. Mobilité sociale et risque de cancer du sein	112
E. Analyses de sensibilité et validation externe.....	115
4. Discussion.....	115
TRAVAIL 4 : INCORPORATION BIOLOGIQUE DE L’ENVIRONNEMENT SOCIAL ET RISQUE DE CANCER DU SEIN	118
1. Contexte.....	118
2. Méthode et schéma d’étude.....	119
3. Résultats.....	120
A. Caractéristiques de la population d’étude	120
B. Association entre chaque biomarqueur et le risque de cancer du sein	121
C. Association entre l’éducation et chaque biomarqueur	125
D. Analyses de sensibilité et analyses en sous-groupe	128
4. Discussion.....	128
A. Biomarqueurs et risque de cancer du sein	129
B. Education et marqueurs biologiques.....	131
Chapitre II : Déterminants sociaux et caractéristiques des cancers du sein.....	134
TRAVAIL 5 : IMPACT DE LA PSE SUR LES FACTEURS PRONOSTICS DES CANCERS DU SEIN	134
1. Contexte.....	134
2. Méthode et schéma d’étude.....	135
3. Résultats.....	137
A. Caractéristiques de la population d’étude	137
B. Association entre l’éducation et le stade du cancer du sein	138
C. Association entre l’éducation et le grade du cancer du sein	143
D. Association entre l’éducation et sous type RH du cancer du sein.....	143
E. Analyses de sensibilité et analyses exploratoires	143
4. Discussion.....	146
A. PSE, stade et grade du cancer du sein	146
B. PSE et sous types de cancer du sein	147
PARTIE 4 - DISCUSSION GENERALE.....	149

Chapitre I. Discussion des résultats	150
1. Synthèse des résultats	150
2. Forces et limites	153
A. Dans l'ensemble des travaux relatifs au cancer du sein	153
B. Spécificités des travaux 4 et 5 relatifs au cancer du sein	154
Chapitre II. PSE et cancer du sein : mécanismes possibles	157
1. Positions socio-économiques et incidence du cancer du sein	157
2. Positions socio-économiques, caractéristiques initiales et sous types de cancer du sein	161
Chapitre III. Particularités de l'incorporation biologique de l'environnement social soulevées par les différents travaux de cette thèse.....	164
1. Stress psychosocial et marqueurs biologiques : cas de la testostérone	164
2. L'éducation : déterminant majeur de l'incorporation biologique de l'environnement social et du développement des maladies	166
<i>PARTIE 5 - CONCLUSION ET PERSPECTIVES.....</i>	168
Chapitre I. Impacts potentiels des résultats	169
Chapitre II. Perspectives	171
<i>REFERENCES BIBLIOGRAPHIQUES</i>	173
<i>ANNEXES.....</i>	192

LISTE DES ABREVIATIONS

CIRC	Centre international de recherche sur le cancer
CRP	Protéine C-réactive
E3N	Étude épidémiologique auprès de femmes de la Mutuelle générale de l'Éducation nationale
EPIC	<i>European Prospective Investigation into Cancer and nutrition</i>
FC	Fréquence cardiaque
HCSP	Haut conseil de santé publique
HDL	Lipoprotéine de haute densité
HPA	<i>Hypothalamic-pituitary-adrenal axis</i>
IGF1	<i>Insulin like growth factor</i>
IL-1 β	Interleukine-1 β
IL-6	Interleukine 6
IMC	Indice de masse corporel
INCa	Institut national du cancer
ISS	Inégalités sociales de santé
LDL	Lipoprotéine de basse densité
MET	Équivalent métabolique pour une heure
NSHDS	<i>Northern Sweden Health and Disease Study</i>
OMS	Organisation mondiale de la santé
PAD	Pression artérielle diastolique
PAS	Pression artérielle systolique
PSE	Position socio-économique
SAM	Système sympathique-adrénal-médullaire
SHBG	Globuline liant les hormones sexuelles - <i>sex hormone binding globulin</i>
TNF α	Facteur de nécrose tumorale - <i>Tumor necrosis-a</i>
UICC	Union internationale contre le cancer

LISTE DES FIGURES

Figure 1. Espérance de vie à 35 ans par sexe, année et diplôme. Source : INSEE Premiere 1584 -----	19
Figure 2. Inégalités relatives de mortalité selon le niveau d'éducation, Source : Mackenbach JP & al, Health & place, 2017-----	20
Figure 3 : Cadre global présentant les déterminants de la santé. SOURCE : Anctil & al, 2012, Direction des communications du ministère de la Santé et des Services sociaux.-----	28
Figure 4 : Mécanismes par lesquels la position socio-économique peut influencer sur l'état de santé. Source : adaptée de adler & al, Ann NY Acad Sci, 2010.-----	33
Figure 5 : Évolution du taux d'incidence des cancers du sein par âge chez les femmes, entre 1990 et 2018 en France. SOURCE : Réseau français des registres du cancer, 2019-----	54
Figure 6 : Aperçu des données sélectionnées dans les deux cohortes et proportion des données disponibles pour chacune des populations étudiées. Le pourcentage représente la proportion de femmes disposant de données disponibles pour chaque covariable dans la population avant imputation.-----	106
Figure 7 : Sélection de la population avec une mesure des biomarqueurs et variables utilisées pour cette étude -----	122
Figure 8 - Résultats présentant les biomarqueurs associés de manière indépendante à la fois au cancer du sein et à l'éducation. (A) graphique représentant les résultats des analyses de régressions des modèles de cox à risques proportionnels (HR [intervalle de confiance à 95 %]) pour l'association entre les biomarqueurs et les cancers du sein en post-ménopause et (B) graphique représentant les résultats des analyses de régressions multinomiales (OR [intervalle de confiance à 95 %]) pour l'association entre l'éducation et les marqueurs biologiques. Les traits pleins représentent les résultats des modèles M1 et les traits pointillés représentent les modèles totalement ajustés.-----	127
Figure 9 : Ajustements en fonction de la caractéristique du cancer du sein considérée -----	136
Figure 10 : Sélection de la population -----	137
Figure 11 : Schéma récapitulatif des résultats obtenus dans le cadre du troisième travail. Les flèches en pointillé indiquent un lien indirect entre l'éducation et le risque de cancer du sein. -----	157
Figure 12 : Mécanismes impliqués dans le lien entre l'éducation et le taux d'œstradiol ----	158

Figure 13 : Mécanismes impliqués dans le lien entre l'éducation et les taux de CRP, d'HDL et de SHBG -----	159
Figure 14 : Mécanismes impliqués dans le lien entre l'éducation et le taux de testostérone -----	159
Figure 15 : Schéma récapitulatif des résultats obtenus dans le cadre de mON travail. Les fleches en pointillé indiquent un lien indirect entre l'éducation et le risque de cancer du sein. -----	160
Figure 16 : Effets délétère et protecteur du niveau d'éducation sur le risque de cancer du sein mis en évidence dans les travaux 3 et 4. -----	161
Figure 17 : Mécanismes impliqués dans le lien entre l'éducation et les sous types moléculaires des cancers du sein -----	162
Figure 18 : L'éducation est liée à quatre dimensions. -----	167

LISTE DES TABLEAUX

Tableau 1 : Classification CIM-10 des tumeurs malignes du sein	79
Tableau 2 : Proportion des différentes caractéristiques de cancer du sein dans la population de E3N (N=7877).....	80
Tableau 3 : Catégorisation de l'activité professionnelle en trois classes.....	82
Tableau 4 : Données collectées lors de l'envoi des différents questionnaires dans la cohorte E3N (adaptée de Clavel-Chapelon F.& al, IJE, 2014). La zone grisée correspond aux données utilisées pour ce travail de thèse.	88
Tableau 5 : Caractéristiques des femmes avec les données disponibles dans E3N selon le statut vis-à-vis du cancer du sein.	108
Tableau 6 : Résultats des modèles de Cox à risques proportionnels sur le risque de cancer du sein à partir des données imputées de E3N (N = 83 436).....	111
Tableau 7 : Résultats des régressions après ajustement chronologique des PSE à partir des données imputées de E3N (N = 83 436)	113
Tableau 8 : Association entre la mobilité sociale et le risque de cancer du sein à partir des données imputées de E3N (N = 83 436)	114
Tableau 9 : Caractéristiques biologiques des femmes avec des données disponibles dans la cohorte E3N (N = 2603) par niveau d'éducation et statut vis-à-vis du cancer du sein	123
Tableau 10 : Association entre chaque biomarqueur avec le risque de cancer du sein diagnostiqué en post-ménopause à partir de la population de E3N avec données imputées (N = 2 603)	124
Tableau 11 : Analyses de régressions logistiques multinomiales de l'association entre l'éducation et chaque biomarqueur précédemment identifié comme étant associé au risque de cancer du sein à partir de la population avec données imputées dans E3N (N = 2 603).	126
Tableau 12 : Caractéristiques des femmes avec les données disponibles dans E3N selon la caractéristique du cancer du sein. ND = Non définissable.....	139
Tableau 13 : Résultats des modèles des régressions logistiques multivariées de l'association entre l'éducation et le stade du cancer du sein à partir des données imputées de E3N (N = 6 701).....	142

Tableau 14 : Résultats des modèles des régressions logistiques multivariées de l'association entre l'éducation et le grade du cancer du sein à partir des données imputées de E3N (N = 6 701)	144
Tableau 15 : Résultats des modèles des régressions logistiques multivariées de l'association entre l'éducation et le sous type du cancer du sein à partir des données imputées de E3N (N = 6 701)	145

PARTIE 1 - INTRODUCTION GENERALE

CHAPITRE I. INEGALITES SOCIALES DE SANTE : DEFINITIONS GENERALES ET CONCEPTS

1. INEGALITES SOCIALES DE SANTE (ISS)

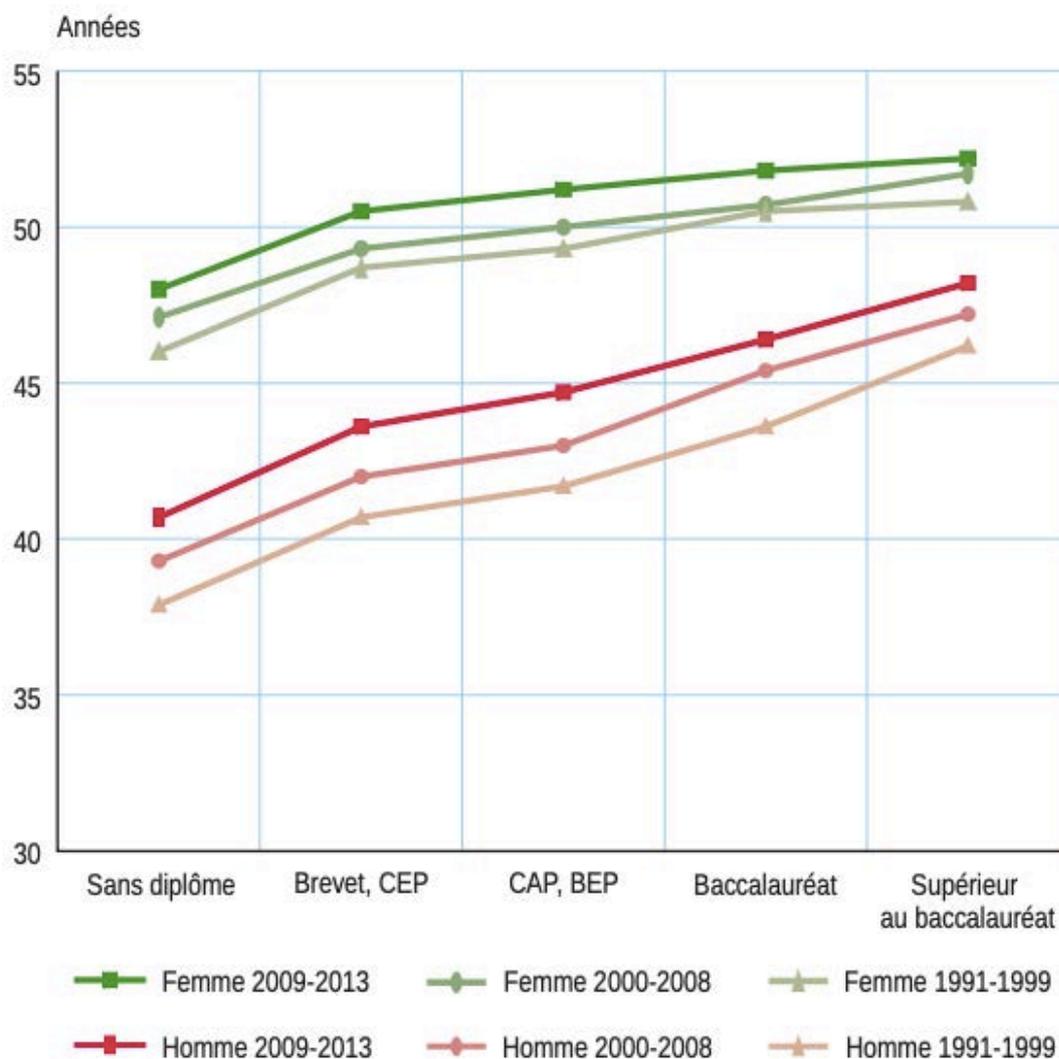
A. DE QUOI PARLE-T-ON ?

Les inégalités sociales de santé représentent des différences injustes et évitables en matière de santé au sein de la population et entre différents groupes de la société¹. Elles ont été définies par Margaret Whitehead et Göran Dahlgren² comme étant systématiques, c'est-à-dire que ces différences présentent un schéma constant dans l'ensemble de la population ainsi que dans la majorité des pays et pour la plupart des maladies. De plus, elles sont socialement construites : ce sont ici les processus sociaux qui sont à l'origine des différences dans l'état de santé. En ce sens, les différences observées devraient pouvoir être modifiées par un effort concerté du pays. Enfin, elles ne relèvent pas de la responsabilité des personnes, ce qui les rend injustes². Ces écarts de santé s'observent tout au long de la hiérarchie sociale sous forme d'un gradient. Il ne s'agit pas seulement d'opposer les catégories les plus et les moins favorisées ou éduquées : cela concerne l'ensemble de la population.

B. DES INEGALITES SOCIALES QUI S'OBSERVENT EN TERME D'ESPERANCE DE VIE ET DE MORTALITE ...

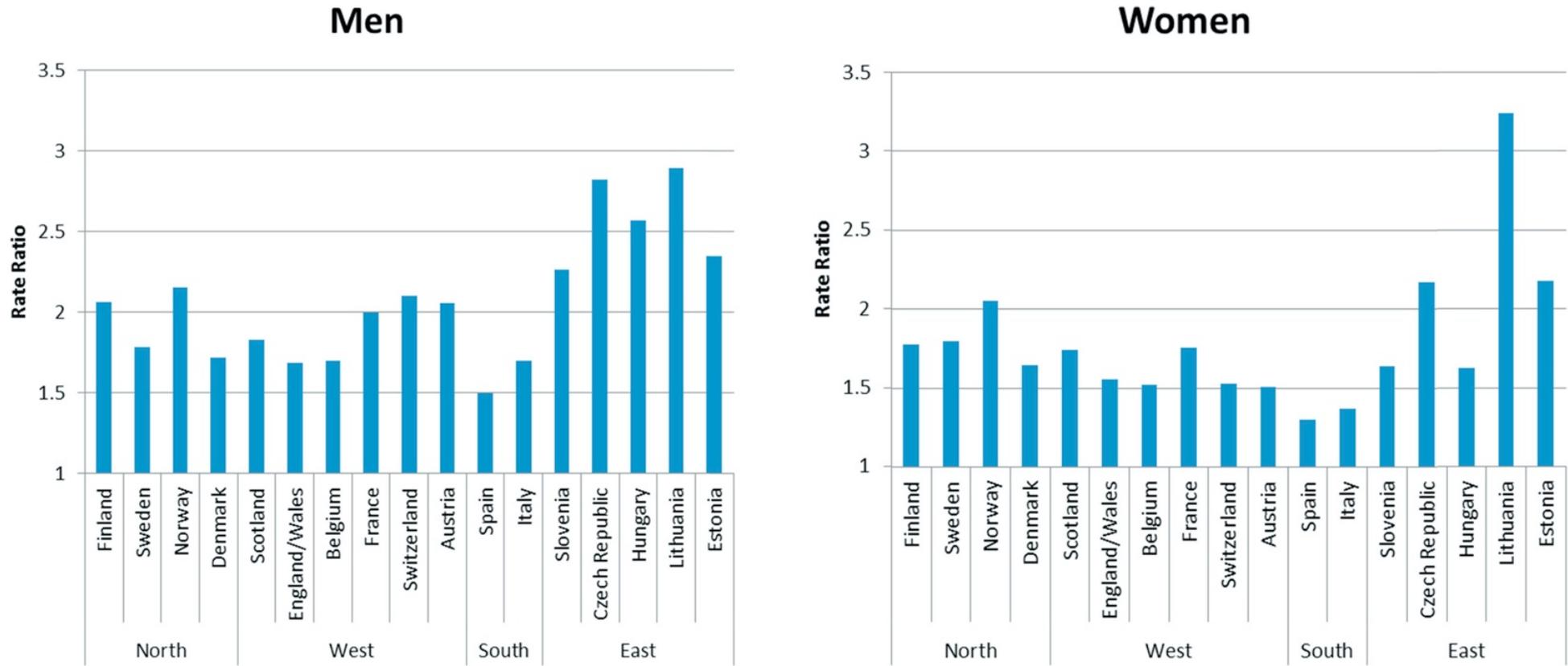
Un des principaux estimateurs utilisés pour mettre en lumière ces inégalités est l'espérance de vie. En France sur la période 2009-2013, à 35 ans, un homme titulaire d'un diplôme supérieur au baccalauréat a une espérance de vie de 48,2 ans alors que pour un homme sans diplôme, elle est de 40,7 ans, soit une différence de 7,5 ans. La différence chez les femmes est également présente quoique moins marquée (4,2 ans). Malgré une augmentation de l'espérance de vie au cours de ces trente dernières années, qui a touché l'ensemble des groupes sociaux, les inégalités semblent persister à la fois chez les hommes et les femmes³ (*Figure 1*).

FIGURE 1. ESPERANCE DE VIE A 35 ANS PAR SEXE, ANNEE ET DIPLOME.
SOURCE : INSEE PREMIERE 1584



Ces inégalités s'observent également en terme de mortalité. La figure 2 montre les inégalités relatives de mortalité toutes causes confondues dans 17 pays européens en fonction du niveau d'éducation (référence : haut niveau d'éducation). Nous voyons ici que la mortalité est plus élevée chez les personnes moins éduquées, que ce soit chez les hommes ou les femmes, avec une taille d'effet qui varie entre les pays (allant du plus élevé pour les pays d'Europe de l'Est, au plus faible dans les pays du Sud de l'Europe)

FIGURE 2. INEGALITES RELATIVES DE MORTALITE SELON LE NIVEAU D'EDUCATION, SOURCE : MACKENBACH JP & AL, HEALTH & PLACE, 2017



C. ...QUI REFLETENT DES INEGALITES DANS L'INCIDENCE DES MALADIES

Les inégalités socio-économiques se retrouvent également pour un large éventail de maladies chroniques. Il a notamment été mis en évidence que les personnes moins favorisées ont un risque plus élevé de problèmes cardiaques, d'hypertension artérielle, de problèmes respiratoires, de douleurs articulaires, de maux de tête, de diabète, d'obésité, de dépression et de cancer^{4,5}. Une étude s'est intéressée à la contribution des grandes causes de décès dans la différence de mortalité totale des hommes âgés de 45-59 ans dans les classes manuelles et non manuelles. Il a été suggéré qu'en France, les cancers en sont les principaux contributeurs comptant pour près de 38% de ces inégalités de mortalité⁶.

D. COMMENT MESURE-T-ON LES ISS ?

L'obtention des données sur la santé de la population en fonction des conditions socio-économiques est primordiale pour surveiller et comprendre les inégalités sociales de santé, mais il apparaît également nécessaire de clarifier les paramètres socio-économiques que nous mesurons et pourquoi⁷. C'est le constat établi par Nancy Krieger dans un des textes fondateurs sur la mesure de la classe sociale et des autres aspects de la position socio-économique (PSE).

Avant d'aborder plus en détail les indicateurs utilisés pour mesurer les inégalités sociales de santé, il convient de revenir sur la définition de trois notions majeures quand nous parlons d'inégalités sociales de santé : *la classe sociale*, *le statut social* et *la PSE*. Ces notions approchent différentes dimensions des inégalités qui peuvent ne pas entretenir les mêmes relations avec différents types de problème de santé. Nancy Krieger⁷ et Mel Bartley⁸ ont défini de manière plus approfondie et nuancée ces différentes notions dans leurs travaux respectifs.

La classe sociale se définit au regard de la division d'une profession en fonction des conditions d'emplois et des relations de travail. La classe sociale définit la situation structurelle d'une personne au sein de l'économie, d'une part par le fait qu'elle possède ou non de capitaux, terres, biens, entreprises, etc, et d'autre part par son statut d'employeur ou d'employé, d'indépendant ou de chômeur^{7,8}. Cette notion repose sur des théories de la structure sociale dont sont à l'origine les sociologues Karl Marx et Max Weber. Selon Karl Marx, il s'agit d'un ensemble d'hommes caractérisé par la place qu'il tient au sein du système productif. D'après

Max Weber, la société est hiérarchiquement stratifiée selon de nombreuses dimensions, créant des groupes dont les membres partagent une position commune sur le marché et donc des "chances de vie" communes.

Le *statut social* apporte une autre dimension, faisant plus souvent référence aux caractéristiques d'une personne, définies à partir de sa famille, de sa religion, de son appartenance ethnique, etc⁸. Un des exemples utilisés pour expliciter cette notion est le système des castes dans la culture hindoue. Ce système représente une hiérarchie allant du statut le plus faible, les *intouchables*, au statut le plus prestigieux, les *prêtres* et les castes intermédiaires sont chacune associées à une profession. Cependant, cela n'implique pas que toutes les personnes d'une même caste exercent cette même profession : leur statut et position dans le système de castes sont déterminés par le statut qu'avaient leurs ancêtres ainsi que d'autres dimensions (culturelles, alimentaires, etc)⁸.

Le terme de *PSE* est un concept global qui comprend à la fois des mesures basées sur les ressources (les revenus, la richesse, les diplômes) et sur le prestige (au rang ou au statut d'un individu dans une hiérarchie sociale, généralement évalué par rapport à l'accès et à la consommation de biens, de services et de connaissances, en lien avec le prestige professionnel, les revenus et le niveau d'éducation)⁷. La PSE est complexe à définir et relève de plusieurs caractéristiques comme cela a été souligné dans un rapport du haut conseil de la santé publique (HCSP)⁹ : premièrement c'est une notion multifactorielle qui renvoie à plusieurs dimensions et peut difficilement être résumée par une seule variable. De ce fait, les indicateurs permettant de mesurer la PSE partagent une dimension commune mais ne peuvent pas se substituer complètement les uns aux autres. C'est également une notion dynamique qui peut évoluer au cours du temps.

Les notions évoquées précédemment peuvent être approchées à travers des indicateurs socio-économiques qui reflètent non seulement différents aspects des expériences de vie mais également les perceptions et/ou les croyances sur la santé. Différents travaux, notamment ceux de Nancy Krieger⁷ et de Bruna Galobardes¹⁰, ont apporté des définitions détaillées de ces différents indicateurs. Dans cette section, nous allons donner un aperçu des indicateurs de PSE les plus couramment utilisés dans la recherche en épidémiologie sociale en portant une attention particulière sur les indicateurs utilisés dans ce travail de thèse.

Éducation

C'est un indicateur qui tente de saisir les ressources liées aux connaissances d'une personne. De façon directe, il influence les styles de vie et les comportements de santé, grâce aux connaissances acquises, à la capacité à en acquérir d'autres. Il permet d'améliorer "l'empowerment", le contrôle sur sa vie, l'estime de soi, la littératie en santé et encourage la participation dans la communauté. Indirectement, le niveau d'études oriente vers une profession et donc une catégorie sociale et un niveau de revenu. Comme l'éducation formelle est normalement achevée au début de l'âge adulte et qu'elle est fortement déterminée par les caractéristiques parentales, elle peut être conceptualisée dans le cadre d'un parcours de vie comme un indicateur qui mesure en partie la PSE au début de la vie¹⁰. Le niveau d'éducation est souvent utilisé car il présente plusieurs avantages : contrairement à la plupart des indicateurs il est fixe au cours du temps et tend à ne plus évoluer. De plus c'est un indicateur qui permet une comparaison entre pays⁹. Nous reviendrons plus en détails sur cet indicateur dans la partie discussion de ce travail de thèse.

Profession

C'est un indicateur qui est à la fois basé sur le niveau d'éducation et qui est un important déterminant des revenus. La profession reflète d'une part le prestige social mais également les conditions de travail (expositions physiques et chimiques néfastes par exemple) et les relations professionnelles (un environnement stressant, l'autonomie, le contrôle, les sollicitations, etc)¹¹. Il existe différentes manières de définir un indicateur de profession, parmi lesquelles nous pouvons citer la distinction entre les métiers manuels (ouvriers non qualifiés, qualifiés, agriculteurs, etc) et non manuels (commerçants, employés, indépendants, etc) ou alors diverses classifications comme notamment la classification socio-économique européenne dont nous reparlerons plus en détails dans la partie méthode de cette thèse. Par ailleurs, il peut y avoir des phénomènes de causalité inverse, la santé pouvant également jouer sur l'accessibilité à l'emploi¹¹.

Revenus

Le revenu est l'indicateur de PSE qui mesure le plus directement la composante des ressources matérielles (alimentation, abri, culture) et des services (soins de santé, activités de loisir et

éducation)¹¹. Les revenus vont pouvoir avoir un impact sur la santé à travers la capacité d'acheter des produits sains et d'accéder à des services bénéfiques pour la santé¹⁰. Par ailleurs, comme pour la profession, la santé peut également avoir un impact sur les revenus. C'est un indicateur qui peut rapidement changer dans le temps¹⁰.

Situation financière perçue

Les indicateurs de PSE précédemment évoqués font référence à des mesures objectives. D'autres mesures, comme la situation financière perçue, peuvent être utilisées, relevant d'une dimension plus subjective, permettant d'approcher la perception par l'individu de sa place dans la structure socio-économique¹². La situation financière perçue permet d'avoir une idée des ressources économiques d'un individu en tenant compte du contexte de vie actuel et passé¹³. A l'inverse des questions sur les revenus, qui peuvent être considérées comme sensibles et regrouper un nombre important de non répondants¹⁰, la situation financière perçue est moins sujette aux données manquantes.

B. INDICATEURS A L'ECHELLE DU FOYER

Profession et revenus du ménage

Si le revenu individuel permet de saisir les caractéristiques matérielles individuelles, le revenu du ménage peut être un indicateur utile, en particulier pour les femmes, qui ne sont pas nécessairement les principaux soutiens de famille. Pour que le revenu soit comparable d'un ménage à l'autre, il faut obtenir des informations supplémentaires sur la taille de la famille ou le nombre de personnes dépendant du revenu déclaré¹⁰. La profession la plus élevée du ménage peut également être un bon indicateur dans cette situation.

Caractéristiques du logement

Les caractéristiques du logement mesurent les aspects matériels des circonstances socio-économiques. La caractéristique la plus couramment utilisée est le mode d'occupation du logement - qu'il s'agisse d'un logement occupé par son propriétaire ou d'un logement loué auprès d'un propriétaire privé ou social¹⁰. Un certain nombre d'autres indicateurs peuvent être utilisés, comme la présence d'équipements ménagers (l'accès à l'eau chaude et froide

dans la maison, la présence d'un chauffage central, le fait d'avoir un réfrigérateur, une machine à laver ou un téléphone, etc) ou encore les conditions de vie des ménages (comme la présence d'humidité et de condensation, le surpeuplement du logement, etc)¹⁰.

C. INDICATEURS INDIRECTS A L'ECHELLE D'UNE ZONE

Les indicateurs écologiques, ou au niveau de la zone, sont également utilisés comme mesures de la PSE. Le plus souvent, ils sont agrégés à partir de données individuelles, généralement issues de recensements ou d'autres bases de données administratives. Ils peuvent être utilisés pour caractériser une zone sur un continuum allant de la défavorisation à la prospérité ainsi qu'une approximation de la PSE des personnes vivant dans ces zones¹⁴. En ce sens ils combinent des effets de contexte et des effets dits de composition liés aux caractéristiques des personnes vivant dans ces zones.

Pour ce travail de thèse nous avons utilisé *la profession du père, le niveau d'éducation et la profession de la participante* comme indicateurs de PSE au cours de la vie. La prise en compte de mesures relatives à la période de l'enfance nécessite de se replacer dans un contexte qui peut être différent de celui dans lequel nous vivons actuellement. La répartition travail - tâches domestiques était probablement bien différente dans les années 1950 que de nos jours. Pour étudier la PSE dans laquelle les individus évoluent à cette période, il nous a semblé plus pertinent de se concentrer sur la profession du père. Ces mesures de la PSE basées sur l'individu présentent plusieurs avantages et pourraient notamment saisir des facteurs individuels (par exemple, des facteurs matériels, comportementaux ou psychosociaux), fournir des informations sur l'accès des individus aux ressources sociales et économiques, et être liées à des caractéristiques macro-environnementales (par exemple, la situation géographique).

2. DES DETERMINANTS MULTIPLES QUI S'ENCHAINENT ET S'ACCUMULENT

D'après l'OMS, les déterminants sociaux de la santé peuvent être définis comme : « les circonstances dans lesquelles les individus naissent, grandissent, vivent, travaillent et vieillissent ainsi que les systèmes mis en place pour faire face à la maladie »¹⁵. La santé peut être impactée par un ensemble des facteurs qui vont intervenir à des niveaux plus ou moins proche des individus. Ce phénomène est décrit sur la figure 3 qui représente la santé autour de laquelle se trouve quatre différentes couches d'influence allant des phénomènes les plus proches des individus aux causes les plus contextuelles¹⁶. En effet, la santé de la population peut être influencée par :

- les caractéristiques individuelles : cette première couche regroupe les éléments qui s'approchent au plus près de la santé des individus. Nous retrouvons ici les *caractéristiques biologiques et génétiques*, comme le sexe, l'origine ethnique ou certaines prédispositions génétiques. Cette première couche tient également compte des ressources dont disposent les individus pour faire face aux événements de la vie à travers *les compétences personnelles et sociales*, ainsi que leurs *habitudes et leurs comportements* comme l'alimentation, l'activité physique, la consommation de tabac mais aussi *les caractéristiques socio-économiques* comme les revenus, la profession et l'éducation.
- les milieux de vie : cette deuxième couche regroupe les différents milieux dans lesquels nous évoluons chaque jour, au travail, à la maison, à l'école, dans le quartier, etc. Ces milieux peuvent avoir un impact sur la santé à différents niveaux notamment les conditions matérielles qui font que les familles, les enfants, les travailleurs se trouvent dans des locaux sains, mais aussi à travers les relations sociales qui se développent dans ces différents milieux. Le développement des quartiers, la présence de commerces, de parcs, de pistes cyclables jouent également un rôle important sur l'état de santé, en influençant notamment les comportements individuels.
- les systèmes : cette troisième couche regroupe les différents systèmes mis en place par l'État et les collectivités afin de permettre aux citoyens d'avoir accès à l'éducation, au système de santé, à l'emploi, etc.

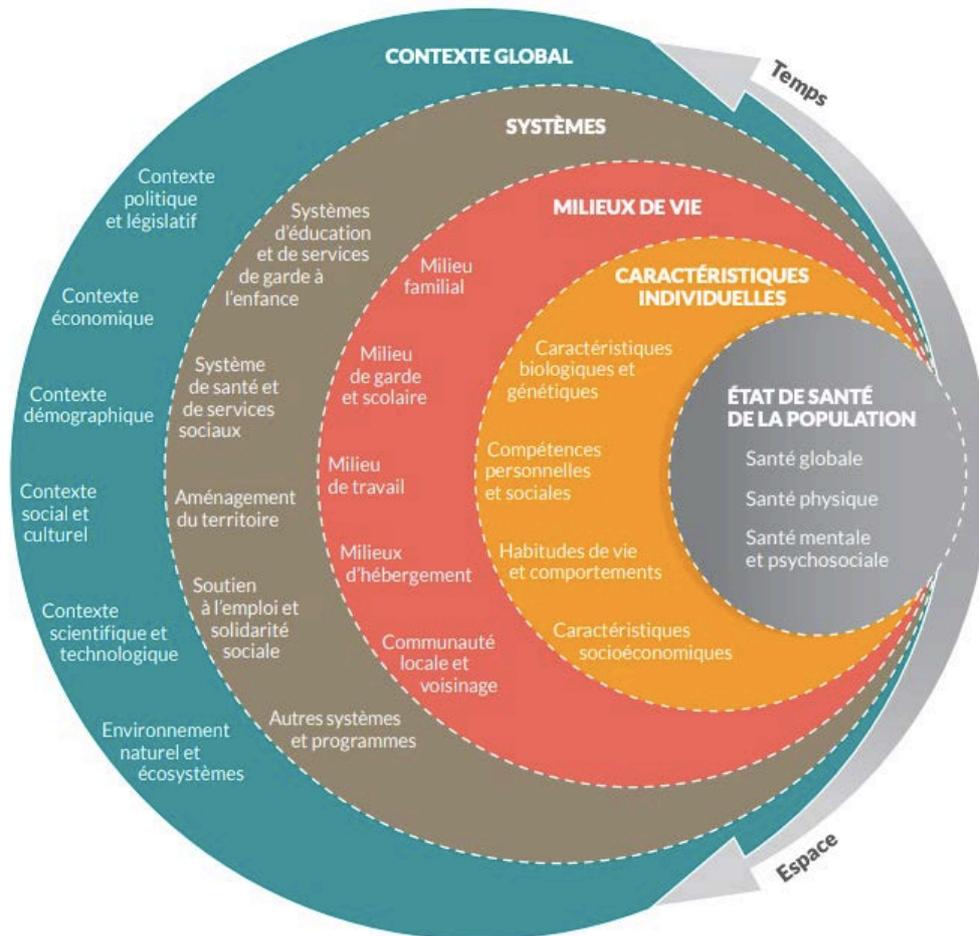
- le contexte global : cette quatrième et dernière couche est la plus éloignée de la santé de la population mais va avoir un effet majeur sur la vie en société, comme le contexte politique, économique, social. Elle va avoir un effet important sur les déterminants des autres couches et va pouvoir avoir un effet positif sur les individus et leur santé mais peut également avoir un effet négatif notamment en accroissant les inégalités.

Ces différents niveaux interagissent entre eux. Cette interaction entre les différents niveaux n'est pas fixe et universelle, et va pouvoir évoluer au cours du temps et des lieux, pays dans lesquels nous nous situons. Par ailleurs l'état de santé des populations peut également influencer sur le contexte global et les décisions prises au niveau des politiques sociales et publiques, comme nous pouvons le constater en cette période de pandémie.

Il est important de noter que chacun des éléments d'une même couche peuvent également interagir les uns avec les autres et qu'il peut être difficile de distinguer clairement une couche des trois autres. En fonction de leurs caractéristiques socio-économiques, les individus vont être confrontés à différents types de conditions matérielles, comportementales et psychosociales ce qui va à terme impacter leur état de santé : les personnes ayant une profession manuelle ou qui gagnent de faibles revenus ou encore qui ont un faible niveau d'éducation ne vont pas vivre dans les mêmes conditions ou dans les mêmes quartiers, avoir les mêmes comportements et le même soutien social que les personnes avec de meilleurs revenus ou un niveau d'éducation plus élevé.

Ainsi, d'après Mel Bartley⁸, l'exposition à des conditions sociales spécifiques, à des événements stressants ou encore des expositions dangereuses ne peut plus être pensée comme un événement unique se produisant simplement et ayant un effet sur la santé. Ces événements sont vécus par des personnes qui ont chacune une histoire, qui a débuté avant même leur naissance. L'apparition d'une maladie va être la conséquence d'une accumulation d'expositions au fil du temps et d'interactions complexes entre de multiples facteurs causaux à travers un processus dynamique¹⁷. Il apparaît donc important d'adopter une approche longitudinale, prenant en compte l'environnement de vie dès le plus jeune âge, quand on s'intéresse à la construction des inégalités sociales de santé.

FIGURE 3 : CADRE GLOBAL PRESENTANT LES DETERMINANTS DE LA SANTE. SOURCE : ANCTIL & AL, 2012, DIRECTION DES COMMUNICATIONS DU MINISTRE DE LA SANTE ET DES SERVICES SOCIAUX.



3. L'APPROCHE BIOGRAPHIQUE OU LIFE COURSE EPIDEMIOLOGY

C'est sur ce concept que repose l'approche biographique, ou « vie entière », ou « trajectoire de vie » ou *life course epidemiology*^{18,19} qui se définit comme l'étude des effets à long terme sur la santé ou le risque de maladies d'expositions physiques, chimiques, comportementales et sociales pendant la gestation, l'enfance, l'adolescence, le jeune âge adulte et la vie adulte ultérieure²⁰. Son objectif est de « mieux comprendre les processus biologiques, comportementaux et psychosociaux qui interviennent tout au long de la vie d'une personne et qui influent sur le risque de développer une maladie ».

Les modèles théoriques proposés dans le cadre de l'approche lifecourse soulignent l'effet sur la santé à l'âge adulte des événements précoces, vécus au début de la vie ainsi que les événements qui se déroulent au cours de la vie. Ces modèles reflètent des processus dynamiques, qui nécessitent de tenir compte du moment de survenue mais aussi de l'enchaînement des expositions et des relations qui peuvent exister entre elles au cours du temps.

Trois modèles théoriques ont été proposés²⁰ :

- Période sensible ou modèle latent - *critical / sensitive period*

Ce modèle suggère que les êtres humains sont plus sensibles aux expositions vécues durant une phase de développement rapide, notamment in utero et pendant la petite enfance, ce qui pourrait entraîner un effet à long terme sur le fonctionnement biologique et le développement de maladies. Les expériences menées par Hubel et Wiesel dans les années 1960 ont montré que les chatons privés d'expériences visuelles dans un œil pendant six semaines après leur naissance restaient définitivement aveugles de l'œil privé²¹. Une période critique a donc été identifiée pour le système visuel des chatons : le cerveau "s'attend" à être exposé à une entrée visuelle pendant une fenêtre de développement, et lorsque cette expérience n'a pas lieu, le système est endommagé de manière irréversible.

- Accumulation des risques - *accumulation of risk*

Ce modèle suggère que les expositions vécues vont graduellement s'accumuler pour causer des dommages à long terme, aboutir au développement d'une maladie ou altérer le bien-être des personnes.

- Chaine de causalité ou modèle de cheminement - *chains of risk / pathway model*

Ce modèle suggère qu'une première exposition va augmenter la probabilité d'être confrontée à un deuxième type d'exposition et ainsi de suite jusqu'à ce que le dernier maillon de la chaîne aboutisse à l'évènement final. Dans ce modèle, la dernière exposition seule peut avoir un impact sur l'évènement mais il est également possible que les différentes expositions qui composent la chaîne aient leur propre effet sur l'évènement, en plus de leur effet conduisant à la dernière exposition.

Il apparaît difficile d'isoler complètement ces différents mécanismes liant les expositions aux événements de santé, notamment en ce qui concerne les modèles de périodes sensibles et ceux des chaînes de causalité.

4. INCORPORATION BIOLOGIQUE DE L'ENVIRONNEMENT SOCIAL

L'approche biographique repose sur l'idée que les expositions psychosociales et comportementales sont susceptibles d'agir sur les processus biologiques, et ainsi influencer sur le risque de développer une maladie. Une littérature de plus en plus abondante soutient que l'environnement, notamment social, dans lequel nous vivons peut-être littéralement incorporé, en pénétrant dans le corps. D'après Clyde Hertzman & Tom Boyce dans leurs travaux de 2010, l'existence du gradient social ne fait plus aucun doute, mais nous commençons seulement à caractériser les voies sociales et biologiques qui l'expliquent à partir des données observationnelles permettant de réinterpréter la biologie humaine dans son contexte social. Plusieurs mécanismes, reposant sur un ensemble de théories proposées indépendamment par Nancy Krieger et Clyde Hertzman, ont été suggérés pour expliquer les relations entre les expositions sociales, les réponses biologiques et la survenue des maladies.

En 1994, Nancy Krieger a proposé la *théorie éco-sociale de la distribution des maladies*^{22,23} qui vise à comprendre les inégalités sociales de santé en rendant compte de la complexité des relations entre l'environnement social et la santé. Cette théorie tient compte d'un ensemble de facteurs opérant à différentes échelles notamment le contexte social, écologique, historique mais aussi l'environnement de vie et de travail ou encore des facteurs individuels concernant les comportements et les histoires de vie. Cette théorie se distingue des précédentes en apportant une vision globale des déterminants de la santé. Ce modèle repose principalement sur le principe de l'incorporation biologique (ou *embodiment* en anglais), autrement dit : « la façon dont nous incorporons littéralement, biologiquement, notre expérience vécue dans un contexte sociétal et écologique, créant ainsi des modèles de santé et de maladies dans la population »²⁴.

Cette théorie repose sur différentes bases fondamentales²³:

- l'incorporation : qui correspond à un processus dynamique qui tient compte de la manière dont les évènements passés ont pu impacter notre corps et la façon dont il répond aux évènements présents. L'incorporation peut être reflétée par une modification de l'état de santé général mais peut également impliquer une modification biologique et moléculaire plus profonde.
- les mécanismes de l'incorporation : différents mécanismes peuvent être à l'œuvre, que ce soit de manière concurrente ou en interaction, incluant entre autres des expositions environnementales (physiques, chimiques ou biologiques), ou des expositions sociales (des traumatismes ou des évènements défavorables) qui vont se produire différemment en fonction du contexte et de la hiérarchie sociale. Nous reviendrons sur les différents mécanismes possibles dans la section suivante.
- les interactions cumulées entre l'exposition, la sensibilité et la résistance : qui tiennent compte de l'importance du moment et de l'accumulation des expositions qui vont être incorporées et des réponses que cela va provoquer de la part de l'organisme.
- la reconnaissance de l'incorporation et de ses mécanismes dans les inégalités sociales de santé et dans les recherches pour expliquer ces inégalités.

D'après Clyde Hertzman, les expositions et stimulations vécues notamment pendant l'enfance vont affecter le fonctionnement et la structure du système nerveux central, ce qui va à terme nuire au développement cognitif, social et comportemental. Le système nerveux central, qui joue un rôle important dans l'interprétation de l'environnement, interagit étroitement avec le système immunitaire et hormonal. De ce fait, une exposition différenciée à certains évènements et le stress des expériences et des circonstances auquel un individu doit répondre vont avoir un impact sur ses processus physiologiques ce qui peut à terme impacter l'état de santé et participer à la construction des ISS si ces expositions sont socialement différenciées²⁵. Clyde Hertzman a alors introduit le terme *d'incorporation biologique* afin de parler des processus par lesquels l'expérience humaine modifie les processus biologiques de manière stable et durable qui influencent la santé tout au long de la vie¹⁷.

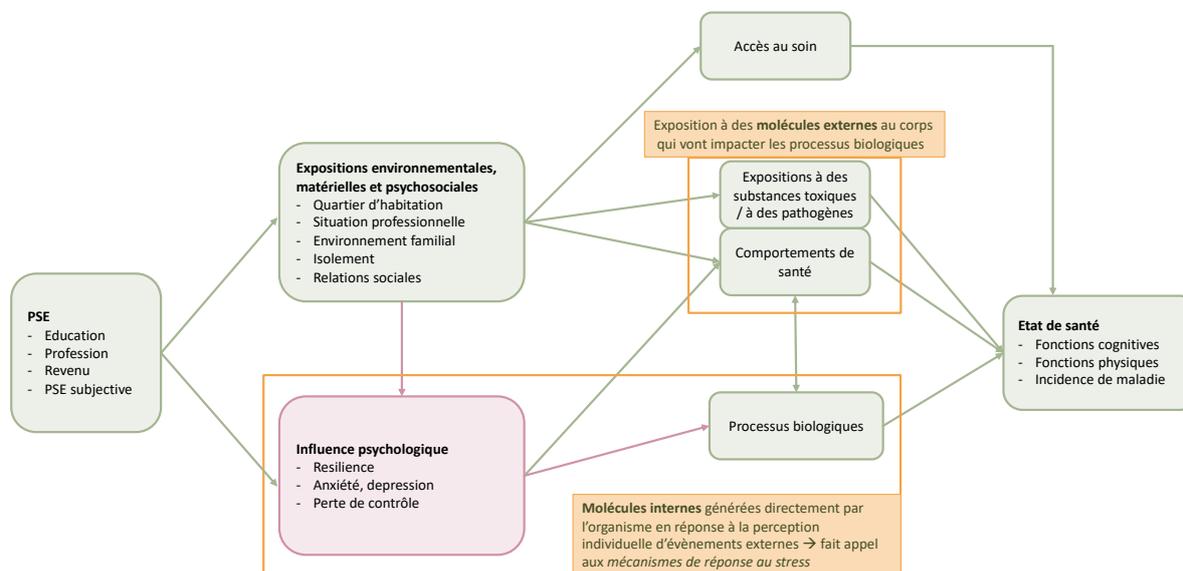
5. MECANISMES POTENTIELS IMPLIQUES DANS LA CONSTRUCTION DES INEGALITES SOCIALES DE SANTE

A travers la figure 4 adaptée des travaux de Nancy Adler et Judith Stewart en 2010²⁶, nous pouvons voir trois mécanismes par lesquels la position sociale peut avoir un impact sur l'état de santé des individus

- Le premier mécanisme sous-tend que la position sociale va impacter l'accessibilité au système de santé, ce qui va avoir un effet sur le moment de la prise en charge et le type de traitement mis en place, et donc sur la santé.
- Le second mécanisme sous-tend que la position sociale va pouvoir avoir un impact indirect sur la santé à travers les comportements de santé ainsi que les conditions de vie.
- Le troisième mécanisme sous-tend que la position sociale va pouvoir avoir un impact direct sur la santé à travers les expositions psychosociales.

Les deux derniers mécanismes vont faire appel à différents types d'expositions, que notre équipe a distingué en deux types, exogènes et endogènes²⁷, qui vont entraîner la modification des processus biologiques ce qui va à terme jouer sur l'état de santé. La séparation des mécanismes biologiques en types exogènes et endogènes correspond principalement à une construction qui pourrait faciliter notre compréhension de la dynamique d'incorporation. Dans de nombreux processus pathologiques, les deux sont probablement en jeu²⁸ : la position sociale peut jouer sur l'un ou l'autre ou les deux types d'expositions et les processus biologiques associés peuvent se cumuler et interagir entre eux²⁹.

FIGURE 4 : MECANISMES PAR LESQUELS LA POSITION SOCIO-ECONOMIQUE PEUT INFLUER SUR L'ETAT DE SANTE.
SOURCE : ADAPTEE DE ADLER & AL, ANN NY ACAD SCI, 2010.



A. EXPOSITIONS D'ORIGINE EXOGENE

La position sociale va pouvoir avoir un impact indirect sur la santé en jouant sur la nature de l'environnement externe dans lequel nous vivons (qualité de l'air, de l'eau, etc) et évoluons au quotidien (qualité du logement, des locaux où nous travaillons et le type de travail exercé, etc) ainsi que par nos comportements (tabagisme, consommation d'alcool, alimentation) qui sont socialement construits. Nos systèmes biologiques sont alors sensibles à des éléments extérieurs et peuvent être modifiés par l'introduction de ces éléments à l'intérieur de notre organisme²⁷.

Ces éléments ont précédemment été qualifiés comme pouvant être des « matériaux vivants (bactéries, virus) et inertes (fibres d'amiante, acide folique) qui ont un impact sur la structure de l'organisme et le fonctionnement biologique. L'impact peut être bénéfique (flore intestinale essentielle, développement du tube neural embryonnaire dépendant de l'acide folique), nocif mais contenu (anticorps, tissu cicatriciel) ou pathologique (tuberculose respiratoire, mésothéliome) »²⁹.

B. EXPOSITIONS D'ORIGINE ENDOGENE

La position sociale peut également avoir un impact sur les expositions psychosociales. Le modèle d'expositions psychosociales suggère que la perception psychologique des événements vécus en lien avec la position sociale va avoir des conséquences sur la santé, tout comme la réponse émotionnelle et comportementale qu'elle va engendrer⁸. Ces expositions peuvent se produire dans différents contextes : dans la vie personnelle et familiale à travers le manque de soutien social, au travail avec notamment les exigences et la pression exercée par les supérieurs, le manque d'autonomie professionnelle ou le déséquilibre effort-récompense ou encore au sein de la communauté à travers des discriminations (racisme) ou l'isolement social par exemple.

Cette perception psychologique peut être vécue comme un stress. Le stress a été précédemment défini par Selye comme « une réponse non spécifique de l'organisme à toute demande de changement »³⁰. Les mécanismes de réponses au stress sont majoritairement gérés par l'axe hypothalamo-hypophyso-surrénalien (*hypothalamic-pituitary-adrenal axis* - HPA). L'adaptation ponctuelle à certaines demandes aboutit à une activation de courte durée de l'axe HPA. Cela va entraîner une cascade de médiateurs hormonaux et de neurotransmetteurs qui vont amener à la libération de cortisol, l'hormone du stress, qui va à son tour déclencher une réponse physiologique afin de répondre à la demande. C'est un phénomène naturel et commun pour l'organisme qui, face à un environnement hostile, va mettre en place une série de réponses coordonnées pour favoriser la survie. Il va notamment permettre d'augmenter la pression sanguine, la fréquence cardiaque et libérer l'énergie nécessaire pour faire face à une demande de type « combat ou fuite » (*fight-or-flight*). Ce concept a été proposé dès 1932 par Cannon pour décrire les réactions physiologiques et comportementales d'un organisme face à des menaces³¹. Ce phénomène se termine en général par un retour à l'équilibre et à une stabilité homéostatique. Cependant, l'exposition chronique à un stress peut être délétère pour l'organisme et son fonctionnement. L'activation répétée et chronique des systèmes de réponses au stress peut ainsi conduire à une usure physiologique, une charge allostatique comme proposé par Teresa Seeman et Bruce McEwen³². Cette charge allostatique va se caractériser par une modification du système neuroendocrinien qui va lui-même pouvoir conduire, via par exemple une production au long cours d'un niveau élevé de cortisol, à des modifications en cascade du système inflammatoire,

immunitaire, métabolique et cardiovasculaire. Ces effets des expositions à des stress chroniques peuvent être particulièrement marqués en fonction du moment de leurs survenus dans la vie d'une personne. Ainsi de précédents travaux menés dans notre équipe ont mis en évidence une association entre les adversités vécues pendant l'enfance avec une usure physiologique à l'âge adulte, mesurée par la charge allostatique³³.

C. ENVIRONNEMENT SOCIAL ET MODIFICATIONS DES PROCESSUS BIOLOGIQUES

En pratique l'identification de ces mécanismes est complexe car peu d'études contiennent à la fois des données biologiques et d'expositions (exogènes et surtout endogènes). Néanmoins quelques études soutiennent l'hypothèse que les indicateurs de PSE ont un effet sur différents processus physiologiques indépendamment des facteurs de risque, suggérant un autre rôle de l'environnement social dans le développement des maladies, via des mécanismes endogènes. Bien que la littérature soit encore peu développée dans ce domaine, de plus en plus de travaux mettent en évidence le rôle de la PSE sur les biomarqueurs cardiovasculaires³⁴, métaboliques³⁵⁻³⁷, inflammatoires³⁷⁻³⁹, le vieillissement biologique et l'usure physiologique³³, indépendamment des facteurs de risque « classiques » (notamment le tabac, l'alcool, l'Indice de Masse Corporelle (IMC) et l'activité physique).

Néanmoins certaines de ces études ne contiennent pas l'ensemble des variables d'ajustement nécessaires pour expliquer les mécanismes impliqués, ou ne sont pas longitudinales. Afin d'approfondir cette question, et en continuité avec des travaux antérieurs de mon équipe, mon premier travail de thèse a porté sur l'étude du lien entre l'environnement social et le système inflammatoire du fait que plusieurs travaux suggèrent que ce système biologique pourrait être sensible à la position social et de son implication dans le développement de maladies comme nous l'aborderons à la fin de ce chapitre. Pour cela j'ai utilisé des données de cohortes disposant d'informations variées au niveau social et comportemental. Dans la littérature disponible sur la relation entre PSE et inflammation, l'étude du moment des expositions sociales au cours de la vie comme l'influence des contextes restent à être mieux élucidées. Le premier travail que j'ai conduit dans le cadre de ma thèse a donc porté sur

l'étude de l'association entre la PSE à trois moments de la vie et le niveau de CRP⁴⁰ (Box 1).
L'article correspondant est disponible en annexe 1.

Travail 1 (box 1) : Déterminants sociaux et inflammation : étude dans six cohortes européennes

Au cours d'un travail réalisé dans le cadre du projet Européen Lifepath qui visait à étudier les voies biologiques qui sous-tendent les différences sociales dans le vieillissement en bonne santé, j'ai étudié l'association entre la PSE à trois moments de la vie (durant l'enfance, en tant que jeune adulte et à l'âge adulte) et le niveau de CRP à l'âge adulte. Cette étude a porté sur une population de 23 008 participants issus de six cohortes différentes : deux cohortes suisses (Skipogh et CoLaus), trois cohortes britanniques (NCDS, Whitehall et ELSA) et une cohorte italienne (EPIC-Italie). Ces cohortes ont été sélectionnées parmi celles disponibles dans le projet car elles possédaient à la fois des données sur la CRP, la PSE à au moins deux moments différents de la vie, et les comportements/IMC. Les données de toutes les cohortes du projet Lifepath ont été harmonisées dans le but de pouvoir comparer les cohortes dont les données ont été collectées à différentes périodes (dimension temporelle) et dans différents pays (effet du contexte) et/ou les analyser ensemble.

Cette étude montre qu'une PSE moins favorisée, quelle que soit la période étudiée, était associée à une augmentation de l'inflammation. J'ai ensuite évalué le rôle déterminant joué par certains comportements et mesures anthropométriques comme le tabac, l'alcool, la sédentarité ou l'IMC sur la relation entre la PSE et le risque d'avoir un taux moyen élevé de CRP et mis en évidence que l'IMC expliquait en partie cette relation.

Pour tenir compte des différentes expériences vécues au cours de la vie dans un même modèle, j'ai par la suite fait des analyses en ajustant chronologiquement les modèles sur les différentes périodes de vie. Ces analyses ont mis en évidence que les personnes qui avaient un faible niveau d'éducation avaient un niveau de CRP plus élevé, indépendamment de la PSE au début de la vie ou à l'âge adulte, ainsi que des facteurs comportementaux et de l'IMC.

La persistance de l'association suggère que le niveau d'éducation peut être un facteur de risque du niveau d'inflammation, indépendamment des facteurs de risque classiques (tabac, alcool, sédentarité, IMC, niveau socioéconomique à l'âge adulte). Ce résultat indique, qu'en plus du mécanisme indirect qui passe par l'IMC, des mécanismes alternatifs pourraient jouer un rôle

dans le lien entre niveau d'éducation et inflammation, comme par exemple l'exposition à des substances chimiques, à des pathogènes ou encore à des stress psychosociaux.



ARTICLE

<https://doi.org/10.1038/s41467-019-08732-x>

OPEN

Multi-cohort study identifies social determinants of systemic inflammation over the life course

Eloïse Berger¹, Raphaële Castagné¹, Marc Chadeau-Hyam², Murielle Bochud³, Angelo d'Errico⁴, Martina Gandini⁴, Maryam Karimi², Mika Kivimäki^{5,6}, Vittorio Krogh⁷, Michael Marmot⁵, Salvatore Panico⁸, Martin Preisig³, Fulvio Ricceri⁴, Carlotta Sacerdote⁹, Andrew Steptoe⁵, Silvia Stringhini¹⁰, Rosario Tumino¹¹, Paolo Vineis^{2,12}, Cyrille Delpierre¹ & Michelle Kelly-Irving¹

Chronic inflammation has been proposed as having a prominent role in the construction of social inequalities in health. Disentangling the effects of early life and adulthood social disadvantage on inflammation is key in elucidating biological mechanisms underlying socio-economic disparities. Here we explore the relationship between socioeconomic position (SEP) across the life course and inflammation (as measured by CRP levels) in up to 23,008 participants from six European cohort studies from three countries conducted between 1958 and 2013. We find a consistent inverse association between SEP and CRP across cohorts, where participants with a less advantaged SEP have higher levels of inflammation. Educational attainment is most strongly related to inflammation, after adjusting for health behaviours, body mass index and later-in-life SEP. These findings suggest socioeconomic disadvantage in young adulthood is independently associated with later life inflammation calling for further studies of the pathways operating through educational processes.

¹ LEASP, UMR 1027, Inserm-Université Toulouse III Paul Sabatier, Toulouse 31000, France. ² MRC-PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. ³ Psychiatric Epidemiology and Psychopathology Center, Department of Psychiatry, Lausanne University Hospital, Lausanne 1004, Switzerland. ⁴ Epidemiology Unit, ASL TO3 Piedmont Region, Grugliasco 10095, Italy. ⁵ Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK. ⁶ Clincium, Faculty of Medicine, University of Helsinki, P. O. Box 20 Helsinki FI-00014, Finland. ⁷ Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan 20133, Italy. ⁸ Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples 80131, Italy. ⁹ Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin 10126, Italy. ¹⁰ Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne 1010, Switzerland. ¹¹ Cancer Registry and Histopathology Department, 'Civic - M. P. Arezzo' Hospital, ASP Ragusa, Ragusa 97100, Italy. ¹² Molecular and Genetic Epidemiology Unit, Italian Institute for Genomic Medicine (IIGM), Torino 10126, Italy. These authors contributed equally: Eloïse Berger, Raphaële Castagné. These authors jointly supervised: Paolo Vineis, Cyrille Delpierre, Michelle Kelly-Irving. Correspondence and requests for materials should be addressed to M. K.-I. (email: michelle.kelly@inserm.fr)

Il n'y a pas de consensus dans la littérature sur les marqueurs d'inflammation les plus pertinents en terme de modèle social ni sur les biomarqueurs standards qui seraient pertinents pour indiquer la présence d'une inflammation chronique délétère à la santé. La plupart des études qui se sont intéressées au lien entre PSE et inflammation se sont principalement focalisées sur la CRP et dans une moindre mesure sur l'interleukine 6 [IL-6]. La difficulté d'interpréter le rôle du système inflammatoire dans la dynamique d'incorporation est due à l'hétérogénéité des études qui utilisent un nombre limité mais varié de marqueurs inflammatoires. Suite au premier travail mené sur la CRP, j'ai participé à une étude visant à décrire en détails la distribution sociale de plusieurs marqueurs inflammatoires différents (CRP, fibrinogène, IL-6, interleukine-1 β [IL-1 β] et facteur de nécrose tumorale- α [TNF- α]) et l'impact de l'ajustement sur les comportements de santé sur ces différentes associations³⁸. L'article correspondant se trouve en annexe 2. Ce travail, mené dans le cadre du projet Lifepath sur les données harmonisées de 17 470 participants issus de six cohortes suisses (Skipogh et CoLaus), britanniques (Whitehall, NCDS et ELSA) et finlandaises (Young Finns)³⁸, a mis en évidence que l'ampleur de la relation entre le niveau d'éducation et ces cinq biomarqueurs inflammatoires était variable, et que le biomarqueur le plus marqué socialement était la CRP, suivie du fibrinogène et, dans une moindre mesure, de l'IL-6, pour lesquels un faible niveau d'éducation était associé à un niveau de biomarqueur plus élevé. L'association entre le niveau d'éducation et la CRP ou le fibrinogène à l'âge adulte n'a été que partiellement atténuée après la prise en compte des facteurs comportementaux (tabagisme, consommation d'alcool et sédentarité) et de l'IMC, ce qui suggère d'une part que le niveau d'éducation peut être une variable indépendante importante et constante qui précède une inflammation élevée et que, d'autre part, le niveau d'éducation peut également fonctionner par des voies intermédiaires autres que les comportements de santé qui méritent une attention particulière dans les analyses futures, notamment les voies psychosociales. Ces résultats suggèrent que la manière dont la PSE est incorporée par l'inflammation peut impliquer des processus complexes qui méritent d'être approfondis.

Les deux travaux décrits précédemment soulignent la relation particulière qui semble exister entre l'éducation et l'inflammation. L'inflammation étant liée à un certain nombre de maladies⁴¹, ce lien entre éducation et inflammation pourrait expliquer une partie du lien entre éducation et santé. Néanmoins les travaux analysant le lien entre PSE, fonctionnement

biologique et état de santé restent rares, manque que ma thèse se propose de partiellement combler. Les cancers représentent les pathologies contribuant le plus aux inégalités sociales de mortalité comme nous en avons discuté au début de ce chapitre. De plus ils constituent un ensemble intéressant de pathologies qui ont à la fois une racine commune dans le système immunitaire et inflammatoire, mais qui se composent également d'un certain nombre de processus étiologiques et de mécanismes biologiques différents. Dans le cadre de mon deuxième travail de thèse, je me suis intéressée au lien entre l'inflammation et le risque de deux types de cancer : les cancers du sein et les lymphomes non Hodgkinien (LNH). Cette étude est décrite dans la box 2 et l'article correspondant se trouve en annexe 3.

Travail 2 (box 2): Inflammation chronique et développement du cancer du sein et des lymphomes non-Hodgkinien : résultats d'une étude cas-témoins

Au cours d'un travail réalisé dans le cadre du projet Envirogenomarkers (EGM) qui visait à identifier de nouveaux biomarqueurs associés à des maladies chroniques pour lesquelles l'environnement pourrait jouer un rôle important, j'ai mené une étude dont l'objectif était de tester le lien entre un score protéique inflammatoire et le risque de cancer du sein et de LNH⁴² ainsi que d'étudier l'impact de différents comportements de santé et du niveau d'éducation sur ces associations. Le score inflammatoire utilisé dans ce travail a été construit à partir de 28 cytokines, chemokines et facteurs de croissance dosés à partir d'un échantillon sanguin fourni par les participants.

Ce travail a été mené sur les données de deux études cas-témoins nichées dans les cohortes EPIC-Italie (N = 261) et NSHDS (N = 402). Les résultats ont mis en évidence une moindre inflammation chez les cas de LNH et dans une moindre mesure, chez les cas de cancer du sein, au moment du prélèvement. Ces associations n'ont pas été affectées par les différents ajustements sur les facteurs de confusion potentiels (comportementaux ou la PSE), ce qui suggère qu'elles peuvent être indépendantes des signatures inflammatoires potentielles de ces facteurs.

Les résultats obtenus dans ce travail dans le cas du cancer du sein sont contraires avec les récentes données de la littérature, à savoir un lien entre un niveau d'inflammation plus élevé et un risque accru de cancer du sein, ce qui peut être dû non seulement à l'approche que nous avons utilisée pour mesurer l'inflammation mais également à la taille de la population

d'étude. En effet, nous avons combiné différentes mesures de protéines liées à l'inflammation pour mieux saisir le statut inflammatoire, alors que d'autres études se sont principalement intéressées à un seul marqueur inflammatoire, notamment la CRP. Deux méta-analyses récentes ont mis en évidence une augmentation du risque de cancer du sein chez les femmes qui ont un niveau élevé de CRP^{43,44}. Ces différences peuvent également être dues à la sélection de la population pour le projet EGM ainsi que du moment du prélèvement de l'échantillon. Les études précédentes sur le lien entre inflammation et LNH retrouvent également une association dont la direction de l'effet varie en fonction des biomarqueurs considérés, très peu ayant considéré un score global. Quoiqu'il en soit, ces résultats suggèrent un lien entre inflammation et cancer du sein et LNH, faisant du système inflammatoire et de ces cancers des candidats pertinents pour étudier le lien entre PSE, fonctionnement biologique et risque de cancer à plus grande échelle dans d'autres contextes. J'ai donc choisi de m'intéresser aux cancers comme marqueur d'état de santé en proposant un approche sur le parcours de vie pour étudier le lien entre PSE, fonctionnement biologique et cancer.

SCIENTIFIC REPORTS

OPEN

Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts

Received: 24 January 2018
Accepted: 4 July 2018
Published online: 17 July 2018

Eloise Berger¹, Cyrille Delpierre¹, Fatemeh Saberi Hosnijeh^{2,3}, Michelle Kelly-Irving¹, Lutzen Portengen², Ingvar A. Bergdahl⁴, Ann-Sofie Johansson⁵, Vittorio Krogh⁶, Domenico Palli⁷, Salvatore Panico⁸, Carlotta Sacerdote⁹, Rosario Tumino¹⁰, Soterios A. Kyrtopoulos¹¹, Paolo Vineis^{12,13,14}, Marc Chadeau-Hyam^{2,13,14}, Roel Vermeulen^{2,13}, Raphaële Castagné^{1,13,14} & EnviroGenoMarkers*

Chronic inflammation may be involved in cancer development and progression. Using 28 inflammatory-related proteins collected from prospective blood samples from two case-control studies nested in the Italian component of the European Prospective Investigation into Cancer and nutrition (n = 261) and in the Northern Sweden Health and Disease Study (n = 402), we tested the hypothesis that an inflammatory score is associated with breast cancer (BC) and B-cell Non-Hodgkin Lymphoma (B-cell NHL, including 68 multiple myeloma cases) onset. We modelled the relationship between this inflammatory score and the two cancers studied: (BC and B-cell NHL) using generalised linear models, and assessed, through adjustments the role of behaviours and lifestyle factors. Analyses were performed by cancer types pooling both populations, and stratified by cohorts, and time to diagnosis. Our results suggested a lower inflammatory score in B-cell NHL cases ($\beta = -1.28$, $p = 0.012$), and, to lesser, extent with BC ($\beta = -0.96$, $p = 0.33$) compared to controls, mainly driven by cancer cases diagnosed less than 6 years after enrolment. These associations were not affected by subsequent adjustments for potential intermediate confounders, notably behaviours. Sensitivity analyses indicated that our findings were not affected by the way the inflammatory score was calculated. These observations call for further studies involving larger populations, larger variety of cancer types and repeated measures of larger panel of inflammatory markers.

¹LEASP, UMR 1027, Inserm-Université Toulouse III Paul Sabatier, Toulouse, France. ²Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands. ³Immunology Department, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁴Department of Biobank Research, Umeå University, Umeå, Sweden. ⁵Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden. ⁶Fondazione IRCCS- Istituto Nazionale dei Tumori, Milan, Italy. ⁷Istituto per lo Studio e la Prevenzione Oncologica (ISPO Toscana), Florence, Italy. ⁸Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy. ⁹Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy. ¹⁰Cancer registry and Histopathology Unit, Azienda Ospedaliera 'Civile -M.P. Arezzo', Ragusa, Italy. ¹¹National Hellenic Research Foundation, Institute of Biology, Medicinal Chemistry and Biotechnology, Athens, Greece. ¹²HuGeF, Human Genetics Foundation, Torino, Italy. ¹³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ¹⁴MRC-PHE Centre for Environment and Health, Imperial College, London, London, UK. Roel Vermeulen and Raphaële Castagné jointly supervised this work. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to R.C. (email: raphaele.castagne@inserm.fr)

Il existe des inégalités sociales de santé en terme de morbidité et de mortalité. Ces inégalités sont systématiques, socialement construites, injustes et apparaissent sous la forme d'un gradient. Elles sont présentes à travers différents pays.

Le développement des maladies chroniques n'est pas le reflet d'une exposition unique mais peut être la conséquence d'une accumulation et d'interactions complexes entre différentes expositions et ce tout au long de la vie.

Différents modèles théoriques suggèrent que nous pouvons incorporer biologiquement l'environnement social dans lequel nous vivons et que notre santé dépend des interactions entre les caractéristiques sociales et biologiques.

Deux types d'expositions, d'origine exogène et endogène, peuvent impacter le fonctionnement biologique et à terme la santé. Les mécanismes exogènes suggèrent que des molécules externes à l'organisme (dû à nos comportements et nos conditions de vie) peuvent pénétrer l'organisme et altérer le fonctionnement biologique alors que les mécanismes endogènes suggèrent que ce sont des molécules directement générées par l'organisme, en réaction à un stress (dû à des expositions psychosociales notamment) qui vont intervenir.

Il existe une relation particulière entre le niveau d'éducation et l'inflammation qui semble par ailleurs liée au risque de cancer. Il apparait alors que le cancer peut être un modèle de maladie particulièrement pertinent pour étudier comment le social devient biologique, via une approche fondée sur le parcours de vie.

CHAPITRE II. INEGALITES SOCIALES OBSERVEES DANS LA MORTALITE PAR CANCERS : QUE RECOUVRENT-ELLES ?

D'un point de vue général, le cancer est l'une des principales causes de décès dans le monde et il joue un rôle majeur dans les inégalités sociales en matière de mortalité⁶. Les inégalités sociales sont présentes à chaque étape du continuum du cancer, depuis le développement de la maladie, en passant par ses caractéristiques biologiques ainsi que l'accès au système de soin et la prise en charge, qui vont tous les trois pouvoir avoir un impact sur la survie des patients. Ces inégalités peuvent s'expliquer en partie par différents mécanismes qui vont être détaillés au cours de ce chapitre.

1. RISQUE DE CANCER : CONSEQUENCES A LONG TERME DES DETERMINANTS SOCIAUX

A. COMPORTEMENTS DE SANTE ET FACTEURS ANTHROPOMETRIQUES

D'après des recherches menées au Royaume-Uni, 40% des cancers seraient évitables si nous changions nos modes de vie et certains de nos comportements⁴⁵, notamment face au tabac et à l'alcool. Ces comportements apparaissent socialement distribués dans la population, en étant plus prononcés chez les personnes présentant une PSE moins favorable. Cependant, ces différences doivent être vues comme la conséquence de déterminants qui vont agir en amont pour influencer l'adoption de ces comportements plutôt que la conséquence de choix individuels. Les comportements de santé peuvent être en effet vus comme des actes socialisés, dont les caractéristiques proviennent des groupes d'appartenance des individus⁴⁶. Parmi les comportements de santé et les facteurs anthropométriques, la consommation de tabac et d'alcool, une « mauvaise » alimentation et l'obésité apparaissent comme les facteurs prédominants face au risque de cancer⁴⁷.

Il a été estimé que le tabagisme était à l'origine d'environ 71 % de tous les décès par cancer du poumon⁴⁷. Le tabagisme était auparavant un comportement répandu chez les personnes avec une PSE élevée. Depuis les années 1950, avec la mise en évidence des effets néfastes du tabac sur la santé, les tendances se sont inversées et la prévalence du tabagisme est devenue plus importante chez les personnes avec une PSE moins favorable. Cet effet a été exacerbé

par un accès limité aux services de traitement de la dépendance ainsi que par la publicité issue de l'industrie du tabac ciblant particulièrement ces populations²⁶. D'après les chiffres de la fondation ARC pour la recherche sur le cancer, 17 cancers sont associés au tabagisme. Le tabac contient près de 70 substances cancérigènes qui vont pouvoir passer dans le sang via les poumons et affecter l'organisme entier. Ces substances vont agir sur l'ADN et augmenter le risque de mutations génétiques qui vont pouvoir par la suite être à l'origine du développement d'un cancer⁴⁸.

La consommation d'alcool est observée chez tous les individus, quelle que soit leur PSE. Le gradient n'est pas aussi marqué que celui observé pour le tabac. Il est même inversé chez les femmes⁴⁹. Cependant, il semblerait que les personnes avec une faible PSE soit plus à risque de développer des cancers liés à la consommation d'alcool. Cela peut s'expliquer par un comportement différent dans les habitudes de consommation, notamment face au phénomène d'alcoolisation massive⁵⁰. Plusieurs types de cancers ont été associés à la consommation d'alcool, notamment les cancers des voies aérodigestives supérieures, de l'œsophage, de foie, du colon-rectum et du sein chez la femme. L'alcool peut agir sur le risque de cancer à travers différents mécanismes⁵¹ : certains métabolites de l'éthanol qui le compose ont des propriétés cancérigènes capables d'altérer l'ADN. De plus, l'alcool va augmenter la concentration de certaines hormones sexuelles dans le sang, ce qui va favoriser le développement de certains cancers, notamment du sein. Il a également été mis en évidence que l'alcool peut favoriser la pénétration d'autres substances cancérigènes dans les cellules, ce qui provoque un effet synergique lorsque la consommation d'alcool s'accompagne de tabac.

L'obésité est un problème majeur de santé publique : en 2015, 17% de la population française était obèse (IMC \geq 30)⁵². La proportion de personnes obèses a largement augmenté ces dernières années en particulier chez les personnes avec une plus faible PSE²⁶, ce qui peut s'expliquer entre autres par des modes de vie de plus en plus sédentaires ainsi que par la multiplication d'aliments riches et peu coûteux sur le marché, facilement accessibles pour une population à faibles revenus. Il a précédemment été suggéré que le surpoids et l'obésité favoriseraient le développement du cancer de l'œsophage, de l'endomètre, du rein, du colon, du pancréas et du sein⁵³. L'obésité peut agir le risque de cancer en favorisant un état

d'inflammation chronique ainsi qu'en modulant la concentration d'insuline, de leptine et des hormones sexuelles⁵³.

B. EXPOSITIONS PROFESSIONNELLES, ENVIRONNEMENTALES ET MATERIELLES

L'environnement social dans lequel nous vivons et travaillons peut nous amener à être confronté à différents types d'expositions. Trois groupes d'expositions liées au risque de cancer peuvent être définis :

Les expositions professionnelles : il a été mis en évidence à travers des études menées sur les expositions professionnelles que les travailleurs manuels sont davantage exposés aux agents cancérigènes⁵⁴. Au cours du temps, 70 agents ont été identifiés et classés comme agents cancérigènes professionnels⁵⁵. Ces substances vont pouvoir agir sur l'ADN et augmenter le risque de mutations génétiques qui peuvent par la suite être à l'origine du développement d'un cancer. Même si la fraction attribuable aux expositions professionnelles estimée par le centre international de recherche sur le cancer (CIRC) est modérée (4 à 8.5%), le rôle des expositions professionnelles peut être important dans les inégalités sociales d'incidence de cancers du poumon⁵⁶ et des cancers respiratoires⁵⁷.

Les expositions chimiques et environnementales : différents facteurs environnementaux ont été liés à l'apparition de certains cancers. Il s'agit de différents agents qui se trouvent dans l'atmosphère, l'eau, les sols et l'alimentation comme la pollution ou encore les pesticides. La pollution de l'air extérieur est classée par le CIRC dans le groupe 1 (cancérigène pour l'homme)⁵⁸, notamment par son implication dans le développement du cancer du poumon. Une revue globale de 2015 a mis en évidence des disparités dans les expositions à la pollution, soulignant que les communautés avec une faible PSE sont confrontées à des concentrations plus élevées des principaux polluants atmosphériques⁵⁹.

Les infections : d'après l'Institut National du Cancer (INCa), 7% des cancers seraient attribuables à des agents infectieux dans les pays développés. Parmi ces agents pathogènes il y a entre autres les *papillomavirus humains* (HPV) qui sont associés au développement des cancers de la zone anogénitale (col de l'utérus, anus, pénis, cavité orale), le *virus d'Epstein-Barr* (EBV) qui est associé aux lymphomes de Burkitt, aux lymphomes hodgkiniens et non-

hodgkiniens et aux cancers du nasopharynx. Il y a également la *bactérie Helicobacter pylori* qui est associée au risque de cancer de l'estomac et aux lymphomes gastriques. La rencontre avec ces pathogènes se déroule en général pendant l'enfance, notamment dans des milieux sociaux moins favorisés^{60,61}. La période durant laquelle l'organisme va entrer en contact avec le pathogène va avoir un impact sur la maturation du système immunitaire et va pouvoir affecter les fonctions immunitaires tout au long de la vie.

C. STRESS ET CANCER

La position sociale est associée à différents types de stress psychosociaux, et ce dès l'enfance. Nous retrouvons notamment dans la littérature une récente revue systématique qui a souligné l'association entre une faible PSE pendant l'enfance et des expériences de vie négatives (abus, négligence, etc)⁶². D'autres études ont par ailleurs mis en évidence une augmentation du stress perçu chez les personnes vivant dans des quartiers défavorisés⁶³ ainsi qu'une association entre une faible PSE et un score de détresse psychologique⁶⁴.

L'association entre le stress et le développement et la progression du cancer a été démontrée au niveau biologique chez l'humain et grâce à l'utilisation de modèles animaux. Une récente étude publiée en décembre 2020 souligne que de plus en plus de preuves scientifiques mettent en lumière que le stress psychologique chronique joue un rôle positif dans l'initiation, la progression et les métastases du cancer et un rôle négatif sur la fonction immunitaire anti-tumorale et la réponse thérapeutique⁶⁵.

Chez l'animal, il a été démontré que la modification des activités de l'axe HPA et du système sympathique-adrénal-médullaire (SAM), qui joue également dans les mécanismes de réponses au stress, altère de nombreux mécanismes biologiques impliqués dans la tumorigenèse, notamment la croissance tumorale, la migration et l'invasion cellulaire, les réponses inflammatoires et immunitaires⁶⁶. Ce phénomène a notamment été bien décrit *in vitro* et dans des modèles animaux de stress de contrainte chronique ou d'isolement social⁶⁶⁻⁷⁰. Il a notamment été mis en évidence chez le rat que l'isolement social sur le long terme aboutissait à la mise en place de mécanismes de réponses au stress prolongés lors de

l'exposition à des événements stressants. L'isolement a également joué un rôle sur le développement de tumeurs mammaires chez ces rats⁷¹.

Chez l'humain, une littérature de plus en plus abondante suggère des liens entre les facteurs psychosociaux, comme le stress, la dépression ou l'isolement social, et la progression du cancer par l'activation des systèmes HPA et SAM⁶⁶. Une altération des taux d'hormones du stress a ainsi été retrouvée chez des personnes atteintes d'un cancer⁶⁶, et un certain nombre de corrélations entre le stress psychosocial, les voies biologiques liées à la tumorigenèse et le cancer ont été rapportées chez l'homme. Par exemple, il a été observé que les patients soumis à un stress psychosocial ont moins de leucocytes, des niveaux élevés de cortisol sérique (basal), ou encore des concentrations plasmatiques accrues de cytokines inflammatoires et davantage de réponses inflammatoires, y compris des dommages à l'ADN, des facteurs de croissance et angiogéniques et des protéases⁶⁷.

De plus, il a précédemment été mis en évidence une association entre des événements stressants notamment pendant l'enfance et certains comportements comme le tabagisme ou l'alcoolisme, qui vont accroître d'autant plus le risque de cancer⁷²⁻⁷⁴.

Il est néanmoins important de souligner que les études disponibles chez l'humain sont relativement rares et principalement basées sur des corrélations et non sur des associations causales. De plus malgré les études qui viennent d'être citées, la littérature épidémiologique sur le lien entre le stress et le cancer est peu abondante et généralement peu concluante car les résultats sont contradictoires. Cela peut être dû à la nature du stress considéré (aigu ou chronique), à son moment (précoce ou adulte) et à sa mesure (rétrospective ou prospective), ainsi qu'à la conception de l'étude (cohorte ou études cas-témoins).

L'exposition à des stress chroniques pourrait donc être un des mécanismes par lesquels la position sociale influence le risque de cancer, mais qui mérite d'être étudié de manière plus approfondi.

2. DES MALADIES QUI N'EXPRIMENT PAS LES MEMES CARACTERISTIQUES INITIALES AU MOMENT DU DIAGNOSTIC

Différentes caractéristiques initiales, notamment les facteurs pronostiques, sont évaluées au moment du diagnostic du cancer. Il s'agit en particulier du stade, qui définit l'étendue du cancer ainsi que du grade qui détermine l'agressivité de la maladie. Ces deux caractéristiques sont détaillées au chapitre suivant, Partie 1, chapitre III, section 1B.

Les études qui ont été menées sur le lien entre la PSE et les caractéristiques des cancers montrent généralement que les personnes moins favorisées ont un cancer plus avancé au moment du diagnostic⁷⁵⁻⁷⁷ et également un grade plus élevé^{78,79}. La littérature sur le grade des cancers est moins riche que celle sur le stade au diagnostic et la plupart des études prennent en compte ces caractéristiques comme facteurs médiateurs potentiels du lien entre la PSE et la survie.

Le stade du cancer au moment du diagnostic peut être impacté par l'accès au système de santé ainsi que la participation au dépistage. Ce point sera détaillé dans la section suivante et au chapitre III de cette partie. Il peut également être la conséquence d'un cancer agressif dont les caractéristiques vont favoriser une croissance rapide de la tumeur.

Pour certains types de cancers, il existe également différents sous types qui vont être plus ou moins agressifs. C'est notamment le cas du cancer du sein, dont un sous type en particulier, le triple négatif que nous définirons plus tard dans la Partie 1, chapitre III, section 1C, est reconnu comme étant plus agressif que les autres types de cancer du sein⁸⁰.

A notre connaissance, les mécanismes par lesquels l'environnement social peut impacter les caractéristiques du cancer sont encore mal connus. Cependant, il a été précédemment suggéré que les mécanismes inflammatoires pourraient jouer sur les caractéristiques du cancer^{81,82}. Nous reviendrons plus en détail sur ce point dans la partie discussion.

3. UNE PRISE EN CHARGE DANS LES SYSTEMES DE SANTE DIFFERENCIEE

L'accès au dépistage, la précocité du diagnostic ainsi que l'accès à des traitements efficaces sont d'une importance capitale pour contrôler et réduire la mortalité par cancer. Les différences observées de mortalité par cancer peuvent être en partie dues à un accès au système de santé différent entre les groupes sociaux¹¹. Plusieurs facteurs peuvent limiter l'accès au système de santé.

Le premier facteur concerne le coût des soins. Comme cela a été souligné dans un rapport du haut conseil de santé publique (HCSP) en 2009, « La proportion de Français bénéficiant d'une assurance n'a cessé d'augmenter au fil des ans pour atteindre la quasi-totalité de la population. La couverture complémentaire est un élément déterminant de l'accès aux soins, et la part laissée à la charge du patient au cours des vingt dernières années a eu tendance à s'accroître et représente autour de 25 % des dépenses de santé »⁸³. En effet, l'absence de prise en charge des frais de santé ou la nécessité d'avoir recours à une assurance privée peut entraîner un renoncement ou un décalage dans la prise en charge, notamment pour les personnes avec de faibles revenus. Différents dispositifs ont été mis en place pour améliorer l'accès aux soins des personnes les moins favorisées, notamment la complémentaire santé solidaire⁸⁴. Dans le rapport qui a été cité précédemment, il est également fait état d'inégalités d'accès au soin de prévention, bien qu'il n'y ait pas dans ces circonstances de barrières financières⁸³. Cela pourrait être dû au fait que tout le monde n'a pas le même rapport au corps et à la maladie, ne possède pas un niveau de connaissance suffisant et toute l'information nécessaire pour prendre en charge sa santé et formuler ses attentes vis-à-vis du système de soins.

L'accès au système de santé peut également être limité par l'accessibilité géographique. Il a précédemment été souligné que « l'offre de soins a un impact sur la consommation de soins, puisqu'une faible densité médicale augmente le coût des soins, par le biais d'un coût de transport ou par le biais du coût d'opportunité du temps, lié au temps d'attente par exemple »⁸⁴. Il a par ailleurs été mis en évidence que plus les personnes sont éduquées, plus elles parcourent des distances plus importantes pour se faire soigner⁸⁴.

Ce retard dans la prise en charge va permettre à la maladie de se développer ce qui va aboutir à un stade au diagnostic plus avancé et cela va à terme impacter le traitement mis en place et

la survie des patients¹¹. D'autres facteurs peuvent également entrer en jeu comme les croyances culturelles et les expériences négatives vécues par le passé dans le système de soin¹¹.

Il a également été souligné dans un récent rapport publié par le CIRC que les personnes appartenant à des groupes moins favorisés reçoivent un traitement de moins bonne qualité que celles appartenant à des groupes plus favorisés¹¹. Ces différences peuvent s'expliquer en partie par l'accessibilité des services de santé et la qualité des soins reçus dans ces services mais également par des facteurs liés plus directement aux patients comme les comorbidités¹¹. En effet, les comorbidités, plus présentes chez les personnes ayant une faible PSE, vont également avoir un impact sur la stratégie thérapeutique mise en place, limitant le recours aux traitements les plus agressifs.

4. PSE ET CANCER : QUELLES SPECIFICITES ?

Les mécanismes que nous venons de voir sont communs pour la plupart des cancers. Comme nous l'avons abordé tout au long de ce chapitre, il est reconnu que les personnes moins favorisées tendent à avoir des comportements et des modes de vie plus délétères pour la santé, qu'elles sont plus susceptibles être exposées dans leurs environnements de vie (domicile et travail) à des polluants ou des pathogènes et qu'elles sont plus susceptibles d'être exposées à des événements stressants bien qu'ayant moins de ressources nécessaires pour y faire face. En ce sens, il a largement été retrouvé dans la littérature que les personnes moins favorisées sont plus à risque de développer des cancers du poumon⁸⁵⁻⁸⁸, de l'œsophages^{88,89}, des lèvres-bouche-pharynx⁸⁵, du larynx⁸⁵, de la vessie^{85,86,88}, de l'estomac⁸⁶, du foie⁸⁸, du pancréas⁸⁸, du col de l'utérus⁸⁸ et du rein^{86,88}. Cependant, quelques localisations cancéreuses, se caractérisant par la possibilité de dépistage ou de surveillance, comme les cancers du sein⁸⁶⁻⁸⁸, de la prostate⁸⁸, de la thyroïde^{86,88}, de la peau^{86,88} ou encore des ovaires⁸⁶, ont un schéma d'incidence inverse : les personnes plus favorisées ont plus de chance d'avoir un cancer diagnostiqué ; la PSE est ainsi liée à la précocité du diagnostic et à la prise en charge.

Il est intéressant de constater que lorsque nous nous intéressons à la survie, une moindre PSE est très largement associée, voire presque toujours à une moindre survie⁹⁰. Les inégalités

sociales de survie varient donc en fonction des types de cancer, et sont plus ou moins liées à des inégalités sociales d'incidence et de prise en charge⁹¹. Il est à noter que ces évidences en France proviennent de données écologiques en ce qui concerne la mesure de la PSE, du fait notamment de l'absence de données sociales individuelles dans les dossiers médicaux et donc dans les registres de cancers.

Compte tenu de sa fréquence, de son profil particulier au regard de la PSE, à savoir une moindre incidence et survie chez les femmes avec une PSE moins élevée, de son lien potentiel avec l'inflammation, il m'est apparu intéressant de se focaliser sur le cancer du sein pour le cœur de mon travail de thèse.

Il existe des inégalités sociales marquées et persistantes dans le continuum du cancer. Les inégalités sociales de mortalité observées peuvent être la conséquence d'inégalités d'incidence, potentiellement d'un profil biologique de tumeurs différent ainsi que d'un accès au système de santé socialement stratifié.

De nombreux mécanismes biologiques socialement stratifiés ont été identifiés comme faisant partie de la chaîne causale du risque de cancer : expositions matérielles (y compris les inertes comme l'amiante, ou les vivantes comme les papilloma virus) et mécanismes comportementaux (y compris la consommation de tabac et d'aliments gras). Il existe des évidences concernant le rôle du stress chronique dans le développement et la progression du cancer qui méritent d'être confirmées.

La PSE est associée à des caractéristiques biologiques du cancer différentes, notamment en ce qui concerne le développement de sous types plus agressifs, du grade et du stade. Les mécanismes impliqués dans le développement des sous types et du grade sont peu connus mais laissent suggérer que des mécanismes inflammatoires pourraient être impliqués. Les inégalités sociales dans le stade au diagnostic peuvent s'expliquer principalement par un accès différencié au système de santé et à la participation au dépistage. Le coût des soins, leur accessibilité géographique ainsi que des croyances et des expériences négatives pourraient entraîner cette différence au niveau de l'accès au système de santé.

Le cancer du sein est le type de cancer le plus fréquent chez les femmes, dont un certain nombre de facteurs de risque ont été identifiés. Par ailleurs il présente une incidence plus élevée chez les personnes favorisées associée à une meilleure survie et pourrait avoir un lien avec l'inflammation et le stress. Toutes ces spécificités m'ont amené à travailler principalement sur le cancer du sein pour ce travail de thèse.

CHAPITRE III. EPIDEMIOLOGIE DU CANCER DU SEIN

1. LE CANCER DU SEIN

A. CONTEXTE GENERAL

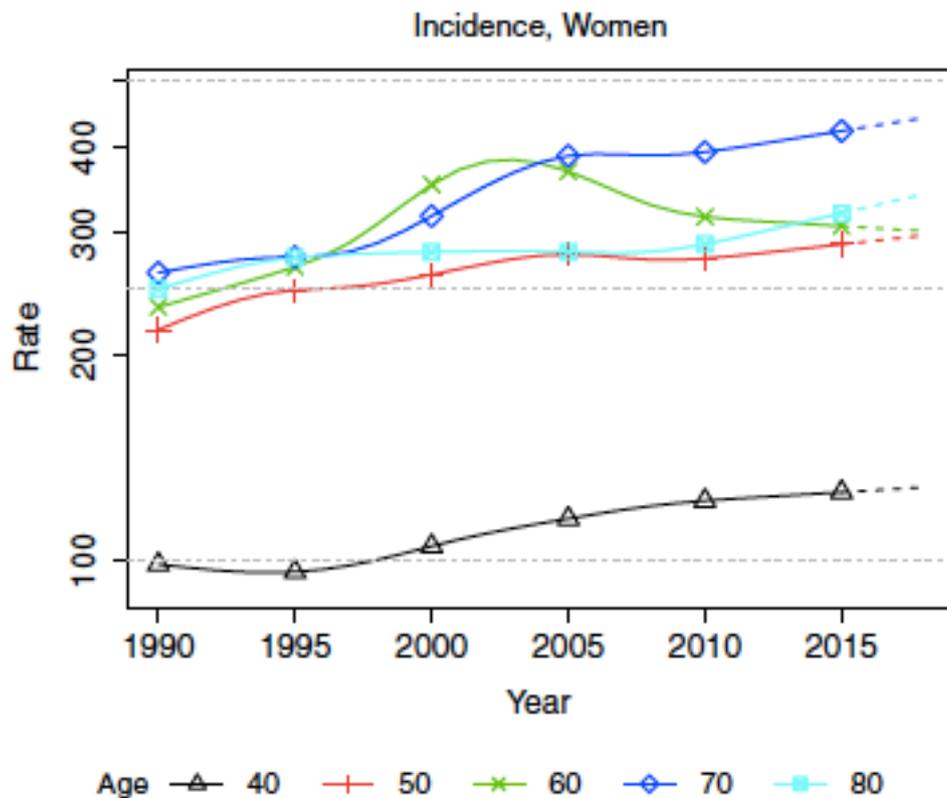
A. INCIDENCE ET MORTALITE

Avec plus de 2 millions de nouveaux cas diagnostiqués dans le monde en 2020, le cancer du sein est le cancer le plus représenté chez la femme, comptant pour près de 25% de la totalité des cancers diagnostiqués dans cette population cette même année.⁹² Ce constat s'applique pour la majorité des pays dans le monde : d'après un rapport publié en 2020 sur des données produites par le CIRC⁹², le cancer du sein est le principal cancer féminin dans 159 pays sur les 185 étudiés. C'est également le plus meurtrier pour 110 de ces pays.

B. TENDANCES OBSERVEES

L'évolution de l'incidence du cancer du sein a fluctué au cours de ces dernières décennies suivant l'avancée des connaissances, des traitements et des comportements. Après une forte augmentation de l'incidence dans les 1980/1990, suite à l'utilisation importante des traitements hormonaux de la ménopause ainsi que la généralisation du dépistage, un plateau semble avoir été atteint entre les années 2000-2010, reflétant une stabilisation du taux d'incidence, qui peut s'expliquer en partie par un effet de saturation du dépistage. Cependant, nous observons depuis les années 2010, une légère reprise de l'augmentation de l'incidence du cancer du sein, aussi bien en France⁹³ (figure 5) que dans le reste du monde.⁹⁴

FIGURE 5 : ÉVOLUTION DU TAUX D'INCIDENCE DES CANCERS DU SEIN PAR AGE CHEZ LES FEMMES, ENTRE 1990 ET 2018 EN FRANCE. SOURCE : RESEAU FRANÇAIS DES REGISTRES DU CANCER, 2019



B. FACTEURS PRONOSTICS

Le cancer du sein est une maladie hétérogène, caractérisée par un aspect morphologique, un comportement et des propriétés moléculaires variées au moment du diagnostic. Ces caractéristiques vont avoir un impact sur l'évolution de la maladie. Elles sont utilisées pour définir la stratégie thérapeutique la plus pertinente. Différents facteurs pronostics du cancer du sein ont précédemment été identifiés.

A. STADE

Le stade est l'un des principaux facteurs pronostics établis à la suite du diagnostic. Il correspond à l'étendue de la tumeur et l'extension de la maladie et il est déterminé à partir de la classification TNM (pour *Tumor, Nodes, Metastasis*) de l'Union Internationale Contre le Cancer (UICC)⁹⁵.

Comme l'indique le nom de la classification, le stade se caractérise par l'évaluation de trois critères :

- **la taille de la tumeur (T)** : la tumeur se développe à partir d'une cellule anormale qui va proliférer pour former, dans un premier temps, un amas de cellules localisé et qui va, à terme, s'étendre aux tissus voisins. Lorsque les cellules anormales restent localisées, on parle de carcinome in situ. Lorsqu'elles commencent à s'étendre, il s'agit d'une tumeur invasive. La taille est classée de 0 à 4 :
 - T0 = la tumeur primitive n'est pas palpable
 - T1 = tumeur ≤ 2 cm
 - T2 = $2 \text{ cm} < \text{tumeur} \leq 5$ cm
 - T3 = tumeur > 5 cm
 - T4 = tumeur avec une extension à la paroi thoracique ou à la peau, quelle que soit sa taille
- **l'étendue aux ganglions lymphatiques (N)** : les cellules cancéreuses peuvent, en plus de se développer dans le sein, aller se disséminer dans d'autres parties du corps et en premier lieu, les ganglions lymphatiques. La classification considère donc l'absence d'atteinte ganglionnaire régionale (N0), l'atteinte d'un ganglion proche de l'anomalie de départ par un amas de cellules de petites tailles (N1) ou l'extension à d'autres ganglions avec présence d'un amas plus ou moins important de cellules (N2 et N3).
- **la présence de métastases (M)** : si les cellules se sont disséminées dans d'autres parties du corps, alors on considère qu'il y a présence de métastases (M1).

Le stade est défini d'après l'UICC sous la forme d'un score déterminé à partir de ces trois critères : le *stade 0* correspond à un cancer in situ ou une anomalie précancéreuse, le *stade 1* correspond à une tumeur unique avec un amas de cellules de petites tailles, le *stade 2* correspond à une tumeur de taille plus importante tandis que le *stade 3* indique un développement régional en dehors de l'organe touché. Le *stade 4* correspond à une tumeur qui s'est propagée dans le sang et s'est disséminée jusqu'à une localisation éloignée.

D'autres regroupements de ces trois critères existent notamment celui proposé par le réseau européen des registres des cancers (ENCR). Ce regroupement, qualifié de plus simple et plus stable que celui de l'UICC, définit le stade des cancers suivant quatre groupes : local limité,

local étendu, régional et avancé⁹⁶. D'après ce regroupement, en France, sur la période 2009-2012, 59% des femmes étaient atteintes d'un cancer à un stade d'extension locale limitée, 29% à un stade d'extension régionale et 12% à un stade d'extension avancée⁹⁶.

B. GRADE

L'agressivité du cancer du sein est caractérisée par son grade. Le grade est défini grâce à un échantillon de la tumeur sur lequel le pathologiste va évaluer trois critères : l'apparence des cellules de la tumeur, la forme du noyau ainsi que le nombre de cellules en développement. Ces indicateurs vont permettre d'évaluer le degré de différenciation des cellules ainsi que le taux de prolifération cellulaire. De la même manière que pour le stade, un score est établi à partir de ces différents critères, allant d'un grade classé 1 (bas grade) pour les tumeurs les moins agressives à un grade classé 3 (haut grade) pour les tumeurs les plus agressives.

C. FACTEURS PREDICTIFS ET SOUS TYPES DE CANCER DU SEIN

A. L'EXPRESSION DE RECEPTEURS HORMONAUX ET DE LA PROTEINE HER2

Certaines cellules possèdent à leur surface ou au niveau de leur noyau des récepteurs hormonaux. Ces récepteurs vont être capables de reconnaître et de se lier aux hormones circulantes dans le sang ce qui va impacter le comportement et/ou la croissance des cellules. Les cellules du sein possèdent naturellement des récepteurs qui vont pouvoir se lier à deux hormones féminines : les œstrogènes et la progestérone. Cette caractéristique se retrouve sur certaines cellules du cancer du sein, ce qui implique que l'exposition à l'une ou l'autre de ces deux hormones va stimuler le développement de la tumeur. On parle alors de tumeurs hormonosensibles ou récepteurs hormonaux positives (RH+). La présence de récepteurs aux œstrogènes et/ou à la progestérone est déterminée par l'examen anatomopathologique d'une partie de la tumeur. Plus l'examen montre un nombre élevé de récepteurs plus le traitement par hormonothérapie aura une influence sur la prolifération des cellules tumorales.

Un autre type de récepteurs, le HER2 pour *human epidermal growth factor receptor 2*, est impliqué dans la croissance des cellules. L'expression de la protéine HER2 est dépendante d'un gène présent dans les cellules saines en un nombre de copies bien défini. Dans le cas de certaines cellules cancéreuses, l'amplification de ce gène conduit à la surexpression de la protéine HER2, ce qui va favoriser la croissance tumorale. Il existe deux moyens d'évaluer le statut HER2 des tumeurs, soit en dénombrant le nombre de copies du gène par la technique FISH, pour *fluorescent in situ hybridization*, soit en dénombrant le nombre de protéines HER2 par la technique IHC, pour *immunohistochimie*. Les cancers du sein présentant une surexpression de la protéine HER2 (environ 20% des cancers du sein) peuvent recevoir un traitement spécifique au trastuzumab (Herceptin) qui va se fixer aux récepteurs et interrompre la croissance des cellules de la tumeur.

Les tumeurs qui ne présentent pas de récepteurs hormonaux et celles ne sur-exprimant pas la protéine HER2 sont peu susceptibles de répondre aux traitements par hormonothérapie et par l'Herceptin respectivement.

B. LES SOUS-TYPES DE CANCER DU SEIN

Quatre sous types moléculaires de cancer du sein peuvent être identifiés à partir du statut de la tumeur vis-à-vis des récepteurs hormonaux et de la protéine HER2⁹⁷.

Les sous-types hormonaux dépendants (RH+ / ± 85% des cancers du sein)

Cette catégorie regroupe deux sous types :

- les tumeurs qui possèdent uniquement des récepteurs hormonaux mais pas de surexpression de la protéine HER2 (**RH⁺/HER⁻**, appelé luminal A, qui a été estimé à ±75% des cancers du sein^{98,99})
- les tumeurs qui possèdent à la fois des récepteurs hormonaux et une surexpression de la protéine HER2 (**RH⁺/HER2⁺**, appelé **luminal B**, qui a été estimé à ± 10% des cancers du sein^{98,99}).

Les tumeurs de type luminal A sont celles avec le meilleur pronostic, présentant en général un stade moins avancé et un grade bas, reflétant une plus faible agressivité par rapport aux autres sous types. Les tumeurs de type luminal B ont quant à elles en général un grade plus avancé que les tumeurs de type luminal A.

Les sous types non dépendants aux hormones (RH- / ± 15% des cancers du sein)

On distingue également deux sous types :

- Les tumeurs présentant une surexpression de la protéine HER2 mais ne possédant pas des récepteurs hormonaux (**RH⁻/HER2⁺**, appelé **HER-positif ou enrichi**, qui a été estimé à ± 5% des cancers du sein^{98,99}).
- les tumeurs ne présentant ni de récepteurs hormonaux ni de surexpression de la protéine HER2 (**RH⁻/HER2⁻**, appelé **triple négatif**, qui a été estimé à ± 10% des cancers du sein^{98,99})

Ces deux sous types présentent le moins bon pronostic parmi les quatre évoqués. Avec le développement des traitements ciblés permettant de neutraliser la protéine HER2, les tumeurs triples négatives sont devenues celles dont le pronostic est le moins bon, ce qui s'explique notamment par le fait qu'à l'heure actuelle, les traitements proposés pour ces tumeurs sont moins efficaces que ceux utilisés pour les autres sous types. Les tumeurs RH- sont en général plus agressives et diagnostiquées à un stade plus avancé¹⁰⁰.

Bien que de nouvelles classifications soient apparues au fil du temps, considérant les tumeurs triples négatives comme un sous-groupe hétérogène notamment par rapport à l'expression de certains gènes en plus des trois récepteurs dont nous avons parlé précédemment, nous nous sommes uniquement basés dans ce travail de thèse sur la présence ou non de récepteurs hormonaux et de l'oncoprotéine HER2.

D. TYPES HISTOLOGIQUES

Les cancers du sein sont également caractérisés suivant leur type histologique. Le sein est composé de lobules qui vont produire le lait en période d'allaitement ainsi que de canaux galactophores qui vont conduire le lait au mamelon. Le cancer du sein peut se développer à partir des cellules de ces deux structures : les cancers lobulaires représentent 15% des cancers du sein tandis que les cancers canaux comptent pour 85% des cancers diagnostiqués.

2. INCIDENCE ET FACTEURS DE RISQUE IDENTIFIES

Le cancer du sein est une maladie multifactorielle pour lequel un certain nombre de facteurs de risque ont précédemment été identifiés. Parmi eux, nous pouvons distinguer trois grands groupes : les déterminants génétiques et familiaux, les caractéristiques de la vie reproductive et enfin les comportements de santé et les mesures anthropométriques.

A. CARACTERISTIQUES GENETIQUES ET FAMILIALES

Facteurs génétiques : certaines mutations, notamment sur les gènes *BRCA1* et *BRCA2*, vont entraîner une prédisposition aux cancers du sein et des ovaires. La mutation en elle-même ne va pas forcément entraîner le développement d'un cancer, mais le risque chez les femmes qui en sont porteuses sera augmenté.

Antécédents familiaux de cancer : les femmes qui ont des antécédents de cancer du sein dans leur famille ont un risque plus important de développer la maladie. Le rapport de risque est de 1.80 pour les femmes qui ont un parent au premier degré atteint et augmente avec le nombre de parents touchés par la maladie¹⁰¹.

B. CARACTERISTIQUES DE LA VIE REPRODUCTIVE

Plusieurs caractéristiques relatives à la vie reproductive des femmes vont jouer sur le risque de cancer du sein notamment grâce à leur capacité à moduler l'exposition aux hormones sexuelles. L'exposition à ces hormones va, dans certains cas, favoriser la progression des cancers du sein.

Certains facteurs vont influencer directement la production d'hormones sexuelles endogènes. C'est notamment le cas pour l'âge aux premières règles, l'âge à la ménopause, l'âge à la première grossesse mais également le nombre total de grossesses.

La durée de vie reproductive, définie à partir de l'âge aux premières règles et de l'âge à la ménopause, va définir la durée globale d'exposition aux estrogènes. L'âge aux premières règles va déterminer l'entrée dans la puberté et le moment où les ovaires vont commencer à produire des œstrogènes. Il a été mis en évidence qu'une puberté précoce augmente le risque global de cancer du sein, ce qui peut s'expliquer par une augmentation de la durée

d'exposition aux œstrogènes. Chaque année supplémentaire entraînant un report de l'âge aux premières règles abouti à une diminution du risque de cancer du sein en post-ménopause de 4%^{102,103}. L'âge à la ménopause va déterminer le moment où les ovaires arrêtent de produire des œstrogènes. Un âge précoce à la ménopause est associé à un moindre risque de cancer du sein. Pour chaque année de plus retardant l'entrée en ménopause, il a été estimé une augmentation de 3% du risque de cancer du sein¹⁰³.

L'âge à la première grossesse va déterminer le moment où les cellules mammaires vont finir de se différencier, ce qui diminue le risque de transformation et de prolifération en cellules cancéreuses des cellules mammaires. Une étude menée dans la cohorte E3N a mis en évidence qu'une augmentation de l'âge à la première grossesse est associée à une augmentation du risque de cancer du sein ; ce sur-risque devient significatif pour les grossesses au-delà de 30 ans¹⁰². Le risque de cancer du sein augmente avec l'augmentation de l'âge à la première grossesse : il a été estimé à 5 % par an pour les cancers du sein diagnostiqués précocement ou avant la ménopause et de 3% pour les cancers diagnostiqués tardivement ou après la ménopause¹⁰⁴.

Le nombre d'enfants : le fait d'avoir peu ou pas d'enfants est également associé à un risque plus important de cancer du sein par rapport aux femmes qui ont eu plusieurs enfants. Il a été mis en évidence que le risque de cancer du sein en post-ménopause diminue de 12% pour chaque grossesse menée à terme¹⁰². Cet effet protecteur s'observe à long terme ; dans un premier temps, il existe un pic de risque 5 à 10 ans après la grossesse qui va progressivement diminuer jusqu'à devenir un effet protecteur¹⁰⁵. Sur le long terme, les femmes qui ont eu des enfants ont un risque plus faible de cancer du sein que les femmes du même âge qui n'ont jamais eu d'enfants. Il a été suggéré que les mécanismes les plus probables de ces effets sont que les concentrations élevées d'hormones pendant la grossesse stimulent la croissance de petites tumeurs précliniques, augmentant ainsi le risque de cancer du sein, mais que les hormones de grossesse diminuent le nombre de cellules sensibles dans le sein en provoquant une différenciation complète du tissu mammaire¹⁰⁵.

L'allaitement : il va moduler l'exposition aux œstrogènes en limitant le nombre total de cycle menstruel. L'allaitement ainsi que sa durée sont associés à un risque plus faible de cancer du sein¹⁰³.

D'autres facteurs comme la prise d'un contraceptif oral ou d'un traitement hormonal de la ménopause vont jouer sur l'exposition aux hormones en apportant des hormones sexuelles exogènes ce qui peut favoriser le risque de cancer du sein. Il a notamment été mis en évidence que les femmes qui prennent des contraceptifs oraux combinés ont une augmentation du risque de près de 20% de développer un cancer du sein. Ce sur-risque s'atténue après l'arrêt du traitement et disparaît au-delà de 10 ans¹⁰⁶.

C. COMPORTEMENTS DE SANTE ET FACTEURS ANTHROPOMETRIQUES

Alcool : la consommation d'alcool est associée à un risque plus important de développer un cancer du sein. D'après les chiffres rapportés par le CIRC, une femme qui consomme 1,5 verres d'alcool par jour a 1,10 fois plus de risque de développer un cancer du sein qu'une femme qui ne consomme pas du tout ou moins de 1,5 verres/jour.

Activité physique : une activité physique régulière diminue le risque de cancer du sein. Un rapport du World Cancer Research Fund (WCRF) conclut que l'activité physique de différents types, durées et intensités a une multitude d'effets physiologiques qui peuvent affecter à travers différents mécanismes le risque de cancer du sein. Parmi ces mécanismes se trouvent la prévention de l'obésité et un effet immuno-modulateur¹⁰⁷.

Tabac : les études qui se sont intéressées à l'association entre la consommation de tabac et le risque de cancer du sein rapportent des résultats hétérogènes malgré les mécanismes biologiques susceptibles de favoriser le développement des cancers. Certaines études ne rapportent pas de lien entre le tabac et le risque de cancer du sein¹⁰⁸ à l'inverse de résultats plus récents qui rapportent une association^{109,110}.

Alimentation : le lien entre l'alimentation et le risque de cancer du sein est difficile à estimer car l'effet des différents aliments est complexe à prendre en compte. Il a été cependant mis en évidence que certains types de régimes pouvaient modifier le risque de cancer du sein. En ce sens, un régime alimentaire riche en produits transformés, en viande et en alcool, comme les régimes de type occidental, favoriserait le risque de cancer du sein. Dans le rapport du WCRF, il est indiqué que les preuves sont limitées quant à l'association entre la consommation

de légumes, de produits à base de caroténoïdes, de produits laitiers et riche en calcium et la diminution du risque de cancer du sein.

Obésité : il semblerait que l'obésité ait un rôle différent en pré ou post-ménopause. En effet, l'obésité est susceptible d'augmenter le risque de cancer du sein chez les femmes en post-ménopause alors qu'elle diminuerait le risque chez les femmes en pré-ménopause^{103,107}. L'obésité en pré-ménopause est associée à un profil hormonal anormal et à des cycles anovulatoires ce qui pourrait expliquer le moindre risque de cancer du sein dans cette population bien que les mécanismes restent à approfondir^{103,107}. En post-ménopause, les hormones sexuelles et notamment les œstrogènes qui sont associés au risque de cancer du sein sont issus du tissu adipeux, ce qui peut expliquer la contribution de l'obésité dans le risque de cancer du sein chez les femmes en post-ménopause¹⁰⁷.

Taille : une taille élevée à l'âge adulte va aboutir à une augmentation du risque de cancer du sein, probablement due à l'exposition à différents facteurs nutritionnels ainsi qu'à des hormones et à des facteurs de croissance¹⁰⁷, particulièrement au début de la vie.

La densité mammaire : la densité mammaire est un facteur reconnu de cancer du sein. Il a été mis en évidence une augmentation du risque de cancer du sein avec l'augmentation de la densité mammaire¹¹¹. Cette étude suggère que la densité reflète la proportion de tissu épithélial et stromal dans le sein par rapport au tissu graisseux non dense. Une densité élevée pourrait refléter un plus grand nombre de cellules à risque de carcinogenèse et/ou un taux accru de prolifération des cellules épithéliales¹¹¹.

Les maladies bénignes du sein : elles sont associées à un risque plus important de développer un cancer du sein. Il a notamment été mis en évidence une augmentation du risque de cancer du sein de 4 à 5 fois chez les femmes qui présentent une hyperplasie épithéliale atypique grave¹¹².

Le cancer du sein est un cancer dont l'incidence augmente avec l'âge, définissant également l'âge comme un facteur de risque. La participation au dépistage ne peut pas en soi être considérée comme un facteur de risque de cancer du sein bien qu'elle favorise le diagnostic des cancers, ainsi que la découverte de lésions très précoces.

3. FACTEURS DE RISQUE DES CANCERS DE MAUVAIS PRONOSTICS

Comme nous l'avons évoqué dans la section 1B de ce chapitre, le cancer du sein est une maladie hétérogène, avec des caractéristiques biologiques distinctes. Les expositions qui influencent le risque de développer un cancer du sein pourraient être liées aux profils moléculaires des tumeurs qui affectent ensuite la biologie et le comportement clinique des cancers qui apparaissent. Les caractéristiques initiales de la maladie peuvent être la conséquence de facteurs étiologiques différents et l'impact de ses facteurs peut varier d'un sous type à l'autre. A l'inverse de la section précédente, nous allons donc aborder ici les facteurs de risque par sous-types de cancer du sein et facteurs pronostics plutôt que par grand groupe de facteurs de risque.

A. LES SOUS-TYPES HORMONAUX DEPENDANTS

Bien que les facteurs reproductifs soient susceptibles d'impacter tous les sous-types de cancer du sein, certaines études indiquent un effet majoritaire sur les tumeurs présentant des récepteurs hormonaux. De plus, il a été mis en évidence que les facteurs de risque de cancer du sein reflètent les facteurs de risque des tumeurs hormonales dépendantes et particulièrement celles de type luminal A^{113,114}. Parmi les facteurs de risque évoqués dans la section 2 de ce chapitre, une récente méta-analyse¹¹³ a mis en évidence qu'un âge précoce aux premières règles, un âge tardif au premier enfant et à la ménopause ainsi que les antécédents familiaux, la consommation d'alcool et l'utilisation d'un traitement hormonal substitutif de la ménopause étaient associés au risque de développer un cancer de type luminal A.

B. LES SOUS-TYPES NON DEPENDANTS AUX HORMONES

Certaines spécificités des facteurs de risque de tumeurs non dépendantes aux hormones ont été soulignées dans la littérature, particulièrement dans le cas des tumeurs triples négatives¹¹⁵. C'est notamment le cas pour le nombre de grossesses, ainsi que dans une moindre mesure pour l'âge à la première grossesse, et l'obésité : en effet, il semblerait qu'un nombre plus élevé de grossesses et l'obésité en pré-ménopause soient tous deux des facteurs

de risque de cancers du sein triples négatifs bien qu'ils aient été identifiés comme facteurs protecteurs du cancer du sein et du type luminal A¹¹³. D'autres caractéristiques apparaissent particulièrement importantes face au risque de tumeurs triples négatives, notamment un âge plutôt jeune, le fait d'être en pré-ménopause, l'appartenance ethnique avec un sur-risque observé chez les femmes afro-américaines ainsi que la dimension génétique avec la présence de la mutation du gène *BRCA1*¹¹⁶. La durée d'allaitement semble être un facteur protecteur important face au développement des tumeurs triples négatives.

Il a également été mis en évidence que certains facteurs de risque de cancer du sein de type luminal A, notamment concernant l'âge aux premières règles et le nombre de grossesses, n'étaient pas retrouvés pour les tumeurs HER2-enrichie¹¹⁷.

C. STADE ET GRADE

Le principal facteur influençant le stade du cancer est le délai au diagnostic. Plus le cancer sera dépisté tard, plus il aura eu le temps de se développer voire d'aller se disséminer à d'autres parties du corps. Un stade avancé au diagnostic peut également être dû à des anomalies moléculaires favorisant la prolifération cellulaire et également être lié au grade des tumeurs. Les mécanismes qui jouent sur le grade des cancers du sein sont peu connus ; il a néanmoins été mis en évidence que l'obésité pouvait être associée à un grade plus élevé de cancer du sein⁸². Il a été suggéré que les mécanismes biologiques liant l'excès d'adiposité à des caractéristiques tumorales de mauvais pronostic comprennent des altérations locales et systémiques des marqueurs inflammatoires, des hormones stéroïdiennes, des cytokines/adipokines, de l'insuline et des facteurs de croissance analogues à l'insuline (IGF - *insulin growth factor*)⁸².

4. INEGALITES SOCIALES FACE AU CANCER DU SEIN : INCIDENCE, FACTEURS PRONOSTICS ET MORTALITE

La PSE apparait de plus en plus dans la littérature comme un facteur de risque de cancer du sein¹¹⁸. Un nombre croissant d'études ont mis en évidence ces dernières années le rôle

important de l'environnement social sur l'incidence du cancer du sein mais également sur les facteurs pronostics et la survie.

A. POSITIONS SOCIO-ECONOMIQUES ET INCIDENCE GLOBALE

Les femmes qui ont une PSE plus favorisée ont un risque plus important de développer un cancer du sein par rapport aux femmes moins favorisées^{119,120}. Cela peut s'expliquer par une distribution socialement différenciée des facteurs de risque comme nous l'avons évoqué dans la section 2 de ce chapitre. Cependant, le sur-risque observé chez les femmes favorisées persiste dans certaines études¹²¹⁻¹²⁴ même après ajustement sur les facteurs reproductifs, notamment l'âge à la première grossesse et la parité, qui semblent être deux facteurs particulièrement importants pour expliquer le gradient social de cancer du sein. Cela peut s'expliquer par le fait que peu d'études ont, dans un même temps, pris en compte les comportements de santé et les mesures anthropométriques en plus des facteurs reproductifs pour expliquer cette association. De plus, peu d'études ont examiné la PSE à différents moments dans le cadre d'un parcours de vie^{122,124-126}. L'étude de l'effet prospectif et combiné de la PSE pendant l'enfance et à l'âge adulte est nécessaire car elle peut révéler différents mécanismes impliqués dans le gradient social de l'incidence du cancer du sein.

Paradoxalement, quelques études ont mis en évidence que les facteurs psychosociaux pouvaient jouer un rôle face au risque de ce type de cancer. Il a notamment été montré que la discrimination raciale était associée à un risque de cancer du sein plus important¹²⁷. Certaines études ont également mis en évidence que des facteurs à l'origine d'un stress sont associés à un risque de cancer du sein plus important^{128,129}. Cependant la littérature sur le rôle des facteurs psychosociaux et du stress dans l'incidence du cancer du sein n'est pas homogène¹³⁰ ce qui suggère que de nouvelles études sont nécessaires pour étudier davantage ce lien.

Comme nous l'avons précédemment abordé, les inégalités sociales peuvent modifier les processus biologiques à travers différents types d'expositions exogènes et endogènes, qui peuvent aboutir au développement de maladies. Certains biomarqueurs ont été identifiés dans le risque de cancer du sein. Nous avons discuté du lien entre inflammation et le risque de cancer du sein dans la Partie 1, chapitre I. Le cholestérol, notamment les lipoprotéines haute densité (HDL) et les triglycérides, ont également été précédemment associés à un

moins de risque de cancer du sein^{131,132}. L'hypertension et une pression artérielle systolique élevée sont aussi retrouvées associées à un risque accru de cancer du sein dans la littérature^{133,134}. Le risque de cancer du sein est également associé à un profil particulier d'hormones stéroïdes sexuelles : un niveau élevé d'œstradiol et/ou de testostérone ainsi qu'un niveau plus faible de SHBG (*sex hormone binding globulin*) sont associés à un risque plus important de cancer du sein¹³⁵⁻¹⁴².

Nous avons également souligné dans la Partie 1, chapitre I que certains de ces marqueurs biologiques (inflammatoires, métaboliques et cardiovasculaires) étaient socialement distribués. Quelques rares études ont également mis en évidence un effet des indicateurs de PSE sur le taux de testostérone et de SHBG, les femmes moins favorisées présentant des taux plus élevés de testostérone et moins élevés de SHBG. Cependant, à notre connaissance, il n'y a pas d'études qui ont considéré dans un même temps la PSE, les marqueurs biologiques et le risque de cancer du sein.

B. POSITIONS SOCIO-ECONOMIQUES, STADE, GRADE ET SOUS TYPES MOLECULAIRES

A. PSE, STADE ET GRADE DU CANCER DU SEIN

Les femmes qui ont une PSE moins favorisée sont plus à risque de recevoir un diagnostic de cancer du sein à un stade avancé¹⁴³⁻¹⁴⁷. Une étude basée sur des données françaises a mis en évidence que les femmes qui ont une faible PSE ont un risque multiplié par deux d'avoir un cancer du sein à un stade avancé, indépendamment des caractéristiques du cancer et du mode de détection (dépistage ou signes cliniques)¹⁴³.

A notre connaissance, un nombre plus restreint d'études ont été menées sur le lien entre la PSE et le grade du cancer du sein. Une étude transversale basée sur les données d'un registre du cancer au Royaume-Uni a mis en évidence une association entre les personnes les moins favorisées et un grade élevé de cancer du sein⁷⁸. Cette association a également été retrouvée dans une étude transversale réalisée dans une population du Worcestershire¹⁴⁸. Une autre étude, menée sur la base de données des cancers aux États-Unis, a mis en évidence que les femmes qui résidaient dans des zones ayant le niveau d'éducation le plus bas étaient plus

susceptibles de présenter un cancer du sein de haut grade que les femmes résidant dans des zones plus éduquées¹⁴⁹. Les mêmes tendances, bien que non significatives, sont également retrouvées dans une étude basée sur les données du registre des cancers de Nouvelle Zélande¹⁵⁰. Par ailleurs, une étude rétrospective de 1994 basée sur les données du registre du cancer et des dossiers de pathologies et de biochimie en Ecosse n'a pas mis en évidence d'association entre la défavorisation de la zone de résidence et le grade des cancers du sein¹⁵¹. Les résultats de ces différents travaux sont bruts et ne présentent pas de modèles ajustés sur les facteurs comportementaux, anthropométriques et reproductifs.

B. PSE ET SOUS TYPES MOLECULAIRES DU CANCER DU SEIN

Les femmes qui ont une PSE moins avantagee semblent être davantage touchées par des sous types de cancer du sein avec un moins bon pronostic^{148,149,152,153}. Cependant, la littérature fait état de résultats contradictoires concernant l'association entre la PSE et les sous-types de mauvais pronostic, certaines études ne retrouvant pas d'association^{154,155}. Compte tenu des disparités ethniques qui existent face au sous type triple négatif mises en évidence dans différentes études soulignant notamment une plus grande proportion de ce sous type chez les femmes afro-américaines¹⁵⁶, beaucoup d'études se sont intéressées à l'impact de l'appartenance ethnique sur le risque de développer ce sous type de cancer du sein. Certaines de ces études se sont intéressées à la PSE en tant que médiateur potentiel du lien entre l'appartenance ethnique et le développement de tumeurs triples négatives^{157,158}. Une étude de 2016 a souligné que bien que les facteurs hormonaux et reproductifs représentaient une proportion substantielle des disparités ethniques face aux cancers qui n'expriment pas de récepteurs hormonaux, dont font partie les tumeurs triples négatives, la PSE représentait une proportion égale ou supérieure de la disparité, et la médiation apparente par la PSE n'était pas due à son association avec la parité, l'âge à la première grossesse, ou d'autres facteurs hormonaux et reproductifs qui tendent à différer par la PSE¹⁵⁸.

En utilisant une grande enquête nationale basée sur la population danoise, une étude de 2006 s'est intéressée à l'association entre la PSE et la progression des tumeurs du sein, définie en deux groupes, *risque faible* ou *risque élevé*, en fonction de la taille de la tumeur, des ganglions lymphatiques, du grade histologique et de la présence ou non de récepteurs hormonaux. Les

résultats après ajustement ont mis en évidence une diminution de 12% du risque d'avoir un cancer de risque élevé chez les femmes qui ont un niveau d'éducation élevé¹⁵⁹.

Par ailleurs une étude française de 2017 menée sur la base de données du registre des cancers du sein de Côte d'Or s'est intéressée à l'impact de la PSE, approchée par un indice écologique de défavorisation, sur le sous type moléculaire des cancers du sein en distinguant les quatre sous types définis par le statut hormonal et la protéine HER2. Cette étude ne retrouve pas d'association entre la PSE et le sous type de cancer du sein¹⁵⁴. Elle souligne qu'une attention particulière doit être accordée aux facteurs de confusion et à l'utilisation de données socio-économiques individuelles pour les prochains travaux sur ce sujet, comme cela a également été suggéré par ailleurs¹⁵⁵.

C. MARQUEURS BIOLOGIQUES ET CARACTERISTIQUES DES CANCERS DU SEIN

La littérature sur l'association entre le niveau de biomarqueurs en amont du diagnostic et les caractéristiques d'agressivité et d'étendue des tumeurs ainsi que du développement des différents sous types est moins riche que celle sur l'incidence du cancer de manière globale comme nous l'avons abordé dans la section précédente.

Peu d'études se sont intéressées au lien entre certains marqueurs biologiques et les différentes caractéristiques des cancers du sein. Néanmoins, une étude s'est intéressée à l'association entre la charge allostatique et les caractéristiques clinico-pathologiques défavorables des cancers du sein. Cette étude a mis en évidence qu'une charge allostatique plus élevée, qui reflète une usure physiologique, est associée à une probabilité accrue d'avoir une tumeur avec des cellules moins différenciées et donc plus agressive, ainsi que d'avoir une tumeur de plus grande taille. Ces résultats pourraient également suggérer qu'une charge allostatique élevée est associée à un risque plus important de tumeurs n'exprimant pas de récepteurs hormonaux, bien que la force de cette association soit relativement faible¹⁶⁰.

C. POSITIONS SOCIO-ECONOMIQUES ET SURVIE

Les femmes qui ont une PSE moins favorisée présentent des durées de survie plus courtes et une mortalité plus élevée que les femmes plus avantagées. Une récente étude a évalué la contribution de l'incidence et de la létalité de différents cancers face aux inégalités sociales de mortalité. Dans le cas du cancer du sein, pour lequel les femmes favorisées ont une incidence plus élevée, il a été mis en évidence que l'excès de létalité chez les femmes moins favorisées explique entièrement la mortalité plus élevée observée dans cette population⁹¹.

Plusieurs études ont mis en évidence que le stade au moment du diagnostic pouvait être un médiateur important dans le lien entre la PSE et la survie^{144,145,161,162}. Cependant, il a été mis en évidence que le stade n'explique pas totalement les inégalités sociales de mortalité qui ont été observées^{146,161,163}. Une étude de 2003 a estimé que la proportion de l'effet de la classe sociale sur la survie attribuable à l'âge et au stade était de 28%¹⁶¹. Cette étude souligne que bien que le stade tende à expliquer une plus grande partie de l'effet, en particulier de la mesure de la classe sociale, il y a toujours une proportion substantielle de l'effet qui n'est pas expliquée par le stade. Une autre étude réalisée sur les données du SEER (Surveillance, Epidemiology and End Results) a également mis en évidence que le stade au moment du diagnostic explique en grande partie les inégalités sociales dans la survie du cancer du sein. Cette étude souligne également que le traitement de première intention, bien qu'en partie orienté par le stade au diagnostic, peut également jouer sur la survie¹⁶².

Comme nous en avons parlé dans la section 1.C de ce chapitre, le sous type de cancer du sein va avoir un impact important sur le type de traitement envisagé, ce qui va avoir un impact sur la survie¹⁶⁴. Il a précédemment été mis en évidence que l'incidence plus élevée de tumeurs ne présentant pas de récepteurs hormonaux chez les femmes moins favorisées explique environ un tiers de la différence de survie entre les femmes favorisées et les femmes moins favorisées¹⁵². Il a également été mis en évidence que les sous-types de tumeurs sont associés à des facteurs pronostics particuliers^{154,165}, ce qui peut jouer sur la survie.

Ces études soulignent l'importance des différentes caractéristiques et sous types des cancers du sein pour expliquer les inégalités sociales de mortalité. Il apparaît néanmoins dans certaines études que la prise en compte de ces différents facteurs n'explique pas entièrement les disparités observées. Une étude de 2014 menée en Irlande souligne qu'en utilisant les

informations disponibles sur les caractéristiques des patients et des tumeurs et sur les traitements (notamment le stade, le grade, la présence ou non de récepteurs hormonaux, la participation au dépistage, la tabagisme et les comorbidités), ils n'ont pas pu expliquer plus de la moitié de la disparité de survie de cancer du sein entre les groupes les plus et les moins défavorisés (le risque de mortalité entièrement ajusté est 18% plus élevé dans le groupe le plus défavorisé, bien que cela ne soit plus statistiquement significatif). Ce travail conclut que des futures études pourraient permettre d'évaluer un éventail plus complet de facteurs de risque, y compris des variables personnelles et communautaires, afin de mieux comprendre les causes de la disparité de survie¹⁶⁶. Une précédente étude de 1994 a suggéré que les femmes issues de zones défavorisées pourraient avoir une capacité réduite à ralentir la propagation à distance du cancer du sein en raison d'une réponse immunitaire différentielle. Cela pourrait être dû à de nombreuses causes possibles, telles que le régime alimentaire, le tabagisme ou des facteurs environnementaux¹⁵¹.

L'étude citée au début de cette section souligne également que le cancer du sein, entre autres, a un pourcentage de survie élevé par rapport aux autres localisations cancéreuses (survie nette à 5 ans de 88%). Il est donc possible que les inégalités sociales aient plus de temps pour se manifester dans le cas du cancer du sein durant le parcours de soins, ce qui se traduit par une plus grande importance des inégalités sociales dans la mortalité⁹¹.

5. PERSPECTIVES DE RECHERCHE SUR LES INEGALITES SOCIALES DE CANCER DU SEIN

Le cancer du sein se distingue de la majorité des autres cancers par le fait que son incidence est plus élevée chez les femmes socialement avantagées bien qu'elles aient une meilleure survie que les femmes moins avantagées. Cela pourrait s'expliquer par des mécanismes différents et socialement stratifiés jouant sur l'incidence et la survie. De nombreux facteurs de risque d'incidence ont été identifiés incluant notamment des comportements de santé, des mesures anthropométriques et des facteurs reproductifs. Il a par ailleurs été mis en évidence qu'une modification des marqueurs biologiques pouvait précéder le développement du cancer du sein. Cependant, il existe certaines limites dans la littérature actuelle et il apparait

nécessaire d'approfondir certaines questions pour mieux comprendre les mécanismes qui sont à l'œuvre dans le lien entre la PSE, l'incidence, la biologie et la survie du cancer du sein.

- Certaines études qui se sont intéressées au lien entre la PSE et le risque de cancer du sein n'expliquent pas totalement le gradient observé à travers les différents facteurs de risque pris en compte. Est-ce que cela peut s'expliquer par l'absence d'un facteur important parmi les variables d'ajustement ? La participation au dépistage, qui n'est pas toujours prise en compte dans ces études, peut-elle jouer un rôle important ? Par ailleurs peu d'études ont pris en compte simultanément un nombre important de médiateurs potentiels. Quel est le lien entre la PSE et le risque de cancer du sein lorsqu'une majorité des variables d'ajustement sont prises en compte ?
- Est-ce que la persistance du gradient social peut refléter d'autres mécanismes que ceux passant par les facteurs reproductifs ou par les comportements de santé et les facteurs anthropométriques ? En particulier existe-t-il un effet biologique de la PSE qui ne passerait pas par les facteurs reproductifs ou par les comportements de santé et les facteurs anthropométriques ? Il existe peu, si ce n'est pas, d'études sur le lien entre PSE, biomarqueurs et risque de cancer qui pourraient apporter des pistes de réflexion pour répondre à cette question.
- Il a été mis en évidence des associations contradictoires entre la PSE et les caractéristiques biologiques des cancers. Est-ce que la PSE est associée aux différentes caractéristiques de stade, de grade et aux différents sous types de cancer du sein ? Quels rôles pourraient avoir les comportements de santé, les mesures anthropométriques et les facteurs reproductifs sur chacune de ces différentes associations ?

CHAPITRE IV. OBJECTIFS ET HYPOTHESES DE CE TRAVAIL DE THESE

1. OBJECTIF GENERAL DE CE TRAVAIL DE THESE

Des inégalités sociales de mortalité par cancer du sein ont été mises en évidence dans la littérature. Différents facteurs peuvent contribuer à ces inégalités notamment sur l'incidence de la maladie, ses caractéristiques au moment du diagnostic mais également sur l'accès au système de santé.

Le cancer du sein est un cancer relativement fréquent, bien documenté, pour lequel un nombre important de facteurs de risque a été identifié. Au fil des années de plus en plus d'études se sont intéressées à l'impact des comportements de santé et des facteurs reproductifs sur le gradient social d'incidence de cancer du sein. Cependant, ces déterminants ne suffisent pas à eux seuls à expliquer les inégalités observées face à la survenue de la maladie. D'autres mécanismes, comme l'exposition à des stress psychosociaux, reposant sur le concept d'incorporation biologique de l'environnement social, suggèrent que les indicateurs de PSE pourraient impacter certains processus physiologiques associés au développement des maladies de manière indépendante des facteurs de risque classiques. Cependant ces mécanismes ont été peu étudiés dans l'incidence des cancers du sein.

Par ailleurs, le cancer du sein se distingue principalement par deux sous types moléculaires (RH) dont les facteurs étiologiques semblent différents. Il est alors possible que différents mécanismes puissent être impliqués dans le lien entre la PSE et le développement de ces différents sous types.

Nous faisons l'hypothèse que les inégalités sociales observées dans la mortalité par cancer du sein sont la conséquence de plusieurs mécanismes combinés, notamment une distribution socialement différenciée de l'incidence de la maladie et ses différents sous-types moléculaires mais aussi de ses caractéristiques initiales, et de l'accès au système de soin (dépistage, prise en charge).

Ce travail de thèse a pour objectif global d'étudier le lien entre la PSE à différents moments de la vie et l'incidence du cancer du sein et ses différents sous-types moléculaires ainsi que

ses caractéristiques cliniques en considérant un des ensembles les plus larges de variables de médiation potentielles, incluant des facteurs comportementaux, anthropométriques, reproductifs mais aussi biologiques.

2. OBJECTIFS ET HYPOTHESES DES DIFFERENTS TRAVAUX

Ce travail de thèse se décompose en trois travaux présentant chacun leurs hypothèses et leurs objectifs.

- TRAVAIL 3

La première hypothèse est qu'il existe des inégalités sociales face au risque de cancer du sein, les femmes plus socialement avantagées étant plus à risque que les femmes moins avantagées. Nous faisons également l'hypothèse que ces inégalités s'expliquent en partie par des facteurs de risque, notamment reproductifs, mais qu'une partie du lien pourrait persister, reflétant l'existence d'autres mécanismes. Le dépistage pourrait être un facteur important pour expliquer ce gradient social.

Le premier objectif de ce travail est d'**étudier le gradient social d'incidence de cancer du sein** à travers une mesure de la PSE à trois moments différents de la vie.

- TRAVAIL 4

La seconde hypothèse de ce travail est que certains biomarqueurs associés au risque de cancer du sein sont également socialement stratifiés. Nous faisons également l'hypothèse que l'éducation influence le niveau de ces biomarqueurs indépendamment des comportements de santé, des mesures anthropométriques et des facteurs reproductifs, ce qui pourrait constituer un mécanisme nouveau par lequel l'éducation pourrait influencer la survenue du cancer du sein.

Le second objectif se découpe en deux parties :

- **étudier la relation entre certains biomarqueurs et le risque de cancer du sein**, et

- **regarder l'impact de l'éducation sur le niveau de ces biomarqueurs**, conditionnellement aux comportements de santé, aux mesures anthropométriques et aux facteurs reproductifs.

- TRAVAIL 5

La troisième hypothèse de ce travail est que les femmes qui ont un faible niveau d'éducation ont un risque plus important d'avoir un cancer du sein à un stade avancé et à un grade élevé au moment du diagnostic ainsi qu'un risque plus élevé de développer un cancer de moins bon pronostic.

Le troisième objectif est **d'étudier l'impact de l'éducation sur l'extension du cancer au moment du diagnostic ainsi que son agressivité et la distribution des différents sous-types moléculaires.**

Ces différents travaux sont présentés dans la partie résultats, qui est séparée en deux chapitres : d'une part les travaux 3 et 4 relatifs à l'étude des mécanismes pouvant contribuer au gradient social d'incidence puis le travail 5 relatifs aux caractéristiques des cancers.

PARTIE 2 - MATERIEL ET METHODES

CHAPITRE I. LA COHORTE E3N

1. CARACTERISTIQUES GENERALES

L'étude épidémiologique auprès de femmes de la Mutuelle générale de l'Éducation nationale (MGEN), ou étude E3N, est une étude de cohorte prospective française qui a débuté en 1990. Cette cohorte est constituée de 98 995 femmes âgées de 40 à 65 ans, adhérentes de la MGEN, vivant en France métropolitaine et ayant répondu favorablement au questionnaire d'invitation. Cette étude est coordonnée par l'équipe « Exposome et hérédité » du Centre de Recherche en Épidémiologie et Santé des Populations au sein de l'institut Gustave Roussy à Villejuif (94).

Cette cohorte a pour objectif d'étudier les associations entre le mode de vie, les facteurs métaboliques et hormonaux, et certaines maladies chroniques, en particulier le cancer. Des données biologiques prélevées au cours du suivi ont par la suite permis d'évaluer les relations entre les maladies chroniques et plusieurs biomarqueurs, ainsi que de tester les interactions entre les facteurs de risque potentiels et les caractéristiques génétiques. Toutes les informations relatives à cette cohorte sont disponibles sur le site internet (<http://www.e3n.fr/>).

E3N est la composante française de l'étude européenne EPIC (*European Prospective Investigation into Cancer and nutrition*). L'étude EPIC inclut plus de 500 000 hommes et femmes issus de 10 pays différents (Allemagne, Danemark, Espagne, France, Grèce, Italie, Norvège, Pays-Bas, Royaume-Uni et Suède)¹⁶⁷.

2. RECUEIL DES DONNEES

Depuis le premier questionnaire (Q1) rempli par les participantes au moment de l'inclusion en 1990, 12 questionnaires ont été envoyés tous les deux à trois ans (dernier questionnaire Q12 envoyé en 2018). A travers ces différents questionnaires, un large éventail de données concernant la vie des femmes et leur santé a pu être recueilli. Le taux de participation est resté élevé tout au long du suivi (de 77% pour le questionnaire Q8 à 92% pour le questionnaire

Q4). Une faible proportion de perdus de vue a également été observée grâce au lien établi avec la MGEN. Enfin, peu de réponses manquantes ou aberrantes ont été retrouvées dans les différents questionnaires¹⁶⁸ avec un taux de réponses à chaque questionnaire de 85%. Les données obtenues au travers des différents questionnaires sont informatisées puis nettoyées par les data-managers qui vérifient également la cohérence des réponses aux questions qui reviennent plusieurs fois au cours du suivi. En parallèle des questionnaires, un prélèvement sanguin a été effectué entre 1994 et 1999 sur 25 000 femmes volontaires ($\pm 40\%$ de la cohorte initiale).

3. CONSENTEMENT DES PARTICIPANTS ET DECLARATION ETHIQUE

L'étude E3N a été approuvée par la Commission National Informatique et Libertés (CNIL / demande d'autorisation n°327346). Toutes les participantes ont signé un consentement éclairé conformément aux règles de la CNIL. Elles ont également donné leur accord pour que la MGEN puisse fournir des renseignements à l'équipe de recherche concernant leur état civil, changement d'adresse ainsi que les remboursements de leurs frais médicaux. De plus, un avis favorable du Comité de Protection des Personnes (CPP) concernant le projet de constitution d'échantillons biologiques a été obtenu en 1996 pour les prélèvements sanguins.

CHAPITRE II. LES CANCERS DANS E3N

1. VALIDATION DES PATHOLOGIES

Les 12 questionnaires comportent une section sur l'état de santé des femmes. C'est à travers cette partie que les participantes auto-déclarent les nouveaux éléments de santé survenus depuis le questionnaire précédent. Il est également demandé aux femmes de fournir, si elles l'acceptent, les coordonnées de leur médecin afin que le médecin de E3N puisse le contacter pour avoir des compléments d'information. Les médecins sont par la suite contactés par voie postale afin de récupérer les documents médicaux nécessaires à la confirmation des pathologies, notamment les cancers, comme les comptes rendus histologiques des tumeurs, les comptes rendus d'hospitalisation et ceux des réunions de concertations pluridisciplinaires. C'est notamment par cette démarche que les cas de cancer auto-rapportés sont validés, ce qui permet d'avoir près de 90% des cas de cancer confirmés histologiquement. Cependant, si ces documents ne peuvent pas être récupérés et que, par conséquent, le cancer ne peut pas être confirmé, aucune date de diagnostic n'est renseignée et le suivi de la femme s'arrête au dernier questionnaire rempli (les données sont censurées à cette date).

2. DEFINITION DES CAS DE CANCER

Les différents documents recueillis auprès des médecins ont permis de valider les tumeurs primitives, concomitantes (diagnostiquées à moins de trois mois d'intervalle) et les récidives, d'obtenir une date de diagnostic précise et de définir un codage à partir de différentes classifications internationales établies par l'OMS et l'Union Internationale de lutte Contre le Cancer (UICC) :

- la classification internationale des maladies 10^{ème} édition (CIM-10) pour la topographie des tumeurs et leur localisation exacte¹⁶⁹
- la classification internationale des maladies oncologiques 2^{ème} et 3^{ème} éditions pour les codes morphologiques des tumeurs¹⁷⁰

Pour sélectionner les cas de cancer du sein pour ce travail de thèse, nous nous sommes basés sur le code topographique validé à partir des documents récupérés.

TABLEAU 1 : CLASSIFICATION CIM-10 DES TUMEURS MALIGNES DU SEIN

Code CIM-10	Localisation	Répartition parmi les cas identifiés dans la cohorte E3N	
		N	(%)
C50.0	Mamelon - Aréole	29	0,37
C50.1	Région centrale du sein - Région rétromamelonnaire	410	5,21
C50.2	Quadrant supéro-interne du sein	1512	19,20
C50.3	Quadrant inféro-interne du sein	593	7,53
C50.4	Quadrant supéro-externe du sein	2752	34,94
C50.5	Quadrant inféro-externe du sein	523	6,64
C50.6	Prolongement axillaire du sein	97	1,23
C50.8	Lésion à localisations contiguës du sein	93	1,18
C50.9	Sein, sans précision	1868	23,71
		7877	100,00

3. CARACTERISTIQUES DES CANCERS DU SEIN

Les différents documents recueillis auprès des médecins ont également permis de récupérer des informations sur les caractéristiques tumorales et sur le bilan d'extension de la maladie qui ont pu être codées à partir des classifications évoquées précédemment.

Pour ce travail de thèse je me suis basée sur quatre caractéristiques : le stade, le grade histologique, l'hormonosensibilité éventuelle de la tumeur (expression positive ou négative des récepteurs aux œstrogènes et à la progestérone) et le statut de la protéine HER2. Le statut de la protéine HER2 n'a été implémenté en routine que pour les cas les plus récents, ce qui amène à une représentation inégale du statut au cours du suivi des femmes E3N : près de 13% des femmes diagnostiquées en 1999 avaient un statut HER2 renseigné et ce pourcentage a augmenté jusqu'à dépasser les 80% pour les cas diagnostiqués en 2006 (annexe 8.8).

TABLEAU 2 : PROPORTION DES DIFFERENTES CARACTERISTIQUES DE CANCER DU SEIN DANS LA POPULATION DE E3N (N=7877)

Stade			Grade			Récepteurs oestrogènes			Récepteurs progéstérone			HER2		
Catégories	N	%	Catégories	N	%	Catégories	N	%	Catégories	N	%	Catégories	N	%
0	1010	12,8	1	814	10,3	Positif	5047	64,1	Positif	3852	48,9	Surexpression	411	5,2
1	4094	52,1	2	2025	25,7	Négatif	1115	14,2	Négatif	2131	27,1	Pas de surexpression	2598	33
2	1945	24,7	3	3046	38,7	Manquants	1715	21,8	Manquants	1894	24	Manquants	4868	61,8
3	523	6,6	4	2	<0,001									
4	90	1,1	Manquants	1990	25,3									
Manquants	215	2,7												

La proportion des différentes caractéristiques de cancer du sein dans E3N est détaillée dans le tableau 2. Compte tenu de la faible proportion de stade 4 nous avons regroupé les catégories 3 et 4 pour nos analyses. Nous avons également regroupé ces mêmes catégories pour le grade. Le sous type de cancer du sein déterminé en fonction de la présence de récepteurs hormonaux (RH) a été défini suivant la présence de l'un ou l'autre des récepteurs. Les femmes ayant des récepteurs aux œstrogènes positifs (RE+) mais des récepteurs à la progéstérone négatifs (RP-), ou étant RE- et RP+ ou RE+ et RP+ sont considérées comme ayant une tumeur avec des récepteurs hormonaux (RH+). L'absence des deux récepteurs a ainsi abouti au sous type RH-. Le sous type triple négatif a été défini de la même façon, en fonction de la présence au non de l'un ou l'autre de récepteurs hormonaux, ainsi que par la surexpression ou non de la protéine HER2 : si le dossier fait état de récepteurs aux œstrogènes et à la progéstérone négatifs et qu'il n'y a pas de surexpression de la protéine HER2, alors il s'agit d'un cancer triple négatif.

CHAPITRE III. DONNEES SOCIALES AU COURS DE LA VIE ET MOBILITE INDIVIDUELLE

1. DONNEES SOCIALES INDIVIDUELLES

Le profil sociodémographique des participantes a été recueilli à partir des deux premiers questionnaires qui ont suivi l'inclusion. Les différents domaines abordés apportent une vision globale de l'environnement social dans lequel les femmes ont évolué grâce, d'une part, à des mesures reflétant différentes périodes de leur vie mais aussi à travers des questions élargies à leur conjoint et à leur père.

Les pourcentages, donnés à titre indicatif dans la partie ci-dessous, ont été estimés à partir de la population de femmes de la cohorte E3N utilisée dans le cadre du troisième travail de cette thèse dont l'objectif était d'étudier le gradient social de cancer du sein à partir d'une population de femmes ayant développé un cancer du sein ainsi que celles étant restées indemnes de cancer (N = 83 436, 84.3%), partie 1, chapitre IV.

A. L'ACTIVITE PROFESSIONNELLE DE LA PARTICIPANTE, DE SON PERE ET SON CONJOINT

La question relative à l'activité professionnelle a été abordée au cours du deuxième questionnaire (Q2 - 1992). Elle détaille 43 modalités regroupées en 3 catégories qui tiennent compte du fait que toutes les femmes sont affiliées à la MGEN, régime qui assure principalement des personnes issues de l'éducation nationale mais pas uniquement. On retrouve donc différentes professions suivant si la participante travaille dans l'éducation nationale dans le domaine de l'enseignement, hors enseignement ou bien si elle ne travaille pas dans l'éducation nationale. La question relative à l'activité professionnelle était la même pour les participantes, leur père et leur conjoint.

Afin d'avoir une mesure qui soit comparable avec les données d'autres cohortes européennes, j'ai créé une variable à trois modalités d'après les travaux de d'Errico & al.¹⁷¹ basés sur la classification socio-économique européenne (*European Socio-Economic Classification*, E-

SEC¹⁷²). Le regroupement des modalités pour la création de la variable 'activité professionnelle' est détaillé dans le Tableau 3.

Parmi les femmes dont l'information était disponible dans notre population d'étude, 16.8% avaient une PSE pendant l'enfance avantagée et 40.9% avaient une PSE parmi les moins avantagées. En ce qui concerne la PSE des femmes à l'âge adulte, 17.7% avaient une PSE avantagée et 17.6% avaient une PSE parmi les moins avantagées.

TABLEAU 3 : CATEGORISATION DE L'ACTIVITE PROFESSIONNELLE EN TROIS CLASSES

	Regroupement d'après l'E-SEC	Répartition suivant les différentes modalités à Q2
Activité professionnelle	Avantagée (Classes 1-3 de l'E-SEC) Cadres dirigeants, professions intellectuelles et scientifiques, professions intermédiaires salariées	<u>Education nationale - enseignement :</u> Directeur d'école; Maître auxiliaire; Professeur des Universités; Maître de conférence; Agrégé. <u>Education nationale - non enseignement :</u> Direction/inspection; Chef d'établissement 2nd degré. <u>Hors education nationale :</u> Chef d'entreprise (10 salariés ou +); Profession libérale; Cadre fonction publique/profession intellectuelle/artistique.
	Intermédiaire (Classes 4, 5, and 6 de l'E-SEC) Petits entrepreneurs (non-salariés), employés qualifiés, ouvriers qualifiés salariés	<u>Education nationale - enseignement :</u> Instituteur; Certifié; Adjoint d'enseignement; PEGC; Professeur de lycée professionnel. <u>Education nationale - non enseignement :</u> Conseillé d'éducation; Infirmier. <u>Hors education nationale :</u> Agriculteur; Artisan; Commerçant et assimilé; Cadre d'entreprise/profession intermédiaire en santé/fonction publique et assimilé; Professions intermédiaires administratives et commerciales des entreprises; Technicien.
	Moins avantagée (Classes 7-9 de l'E-SEC) Professions salariées peu qualifiées, retraités et personnes hors du marché du travail de 65 ans ou plus, autres personnes sans emploi	<u>Education nationale - enseignement :</u> / <u>Education nationale - non enseignement :</u> Attaché; Secrétaire; Commis/agent/adjoint administratif; Steno-dactylo; Ouvrier contremaître; Agent de bureau; Agent d'établissement d'enseignement; Autre <u>Hors education nationale :</u> Contremaître/agent de maîtrise; Employé de la fonction publique; Employé administratif d'entreprises; Employé de commerce; Personnel des services directs aux particuliers; Ouvrier qualifié; Ouvrier non qualifié; Ouvrier agricole.

B. L'ÉDUCATION DE LA PARTICIPANTE ET DE SON CONJOINT

La question relative au niveau d'éducation a été abordée lors du premier questionnaire (Q1 - 1990). Comme pour la variable d'activité professionnelle, le niveau d'instruction est disponible pour les participantes ainsi que pour leur conjoint.

De la même manière que pour l'activité professionnelle, j'ai créé une variable à trois modalités en regroupant les six catégories d'origine pour pouvoir faire une comparaison qui soit cohérente entre différents pays européens, comme cela a été proposé précédemment¹⁷¹. La variable 'éducation' est donc catégorisée en :

- **Élevé** : 3 années ou plus après le baccalauréat
- **Moyen** : de niveau baccalauréat jusqu'à 2 ans au-delà
- **Faible** : pas d'études, certificat d'études ou diplômes professionnels

Parmi les femmes pour lesquelles l'information était disponible dans notre population d'étude, 36.1% des femmes avaient un niveau d'éducation élevé et 13.9% avaient un faible niveau d'éducation.

2. MOBILITE SOCIALE : DE L'ENFANCE A L'AGE ADULTE

A travers les différentes questions qui ont été abordées, nous pouvons donc appréhender l'environnement social des participantes à trois moments distincts de leur vie : pendant leur enfance grâce à la profession de leur père, au début de leur vie d'adulte avec le niveau d'éducation et à l'âge adulte via l'activité professionnelle.

A partir de l'activité professionnelle du père et de celle des participantes, j'ai créé un indicateur de mobilité sociale défini de la façon suivante :

- **Stable moins avantageé** : les femmes ont grandi dans un environnement moins avantageé et sont restées dans le même type d'environnement
- **Mobilité sociale ascendante** : les femmes ont grandi dans un environnement moins avantageé et ont évolué vers un environnement plus favorisé
- **Mobilité sociale descendante** : les femmes ont grandi dans un environnement plutôt favorisé mais vivent dans un environnement moins avantageé à l'âge adulte

- **Stable intermédiaire** : les femmes ont grandi dans un environnement avec une position sociale intermédiaire et sont restées dans le même type d'environnement
- **Stable avantagee** : les femmes ont grandi dans un environnement plutôt favorisé et sont restées dans le même type d'environnement

Parmi les femmes dont l'information était disponible dans notre population d'étude, 4% avaient bénéficié d'une PSE avantagee tout au long de leur vie et 8% avaient vécu dans un environnement stable moins avantagee. 41% avaient vécu une mobilité ascendante entre leur enfance et l'âge adulte et 19% une mobilité descendante.

CHAPITRE IV. MARQUEURS BIOLOGIQUES

1. COLLECTE DES ECHANTILLONS

Les marqueurs biologiques utilisés dans cette étude ont été mesurés dans de précédentes études cas-témoins menées sur les 25000 femmes de la cohorte E3N ayant eu un prélèvement sanguin. Les mesures de certains d'entre eux ont été décrites ailleurs¹⁷³⁻¹⁷⁵. Les échantillons de sang des cas et des témoins appariés (sur l'âge, le statut ménopausique et la date au moment du prélèvement, l'âge à la ménopause et le centre de prélèvement) ont été analysés dans le même lot d'analyse et les techniciens de laboratoire n'ont pas été informés du statut de cas-témoin des sujets des études. La majorité des analyses ont été réalisées au laboratoire de biochimie de l'hôpital Bichat (Paris).

2. TRAITEMENT DES DONNEES BIOLOGIQUES

Compte tenu des paramètres suivants :

- 1) les biomarqueurs ont été collectés dans le cadre de différents projets type cas-témoins nichés dans E3N pour différentes maladies
- 2) ces études ont été réalisées entre 2004 et 2011, et comme le suivi a continué jusqu'en 2014, certains témoins ont pu évoluer en cas
- 3) dans la sous population des participantes ayant des mesures biologiques, les caractéristiques des femmes ayant développé un cancer du sein et celles étant restées indemnes de la maladie sont différentes
- 4) et qu'à ce moment je n'avais pas d'informations relatives à l'inclusion des participantes dans des études cas-témoins ni de leur statut cas ou témoin dans ces études,

J'ai traité les données biologiques indépendamment du design de leurs études d'origine.

3. DEFINITION DES BIOMARQUEURS UTILISES POUR LES ANALYSES

Notre idée première était de récupérer les différents biomarqueurs qui constituent la charge allostatique¹⁷⁶ afin d'utiliser ce score comme marqueur biologique. Parmi ces biomarqueurs, certains étaient disponibles dans E3N : un biomarqueur relatif au système inflammatoire (protéine C-réactive), quatre biomarqueurs issus du système métabolique (triglycérides, HDL, LDL, cholestérol), trois issus du système cardiovasculaire (la fréquence cardiaque et les pressions systoliques et diastoliques). Il était prévu que de nouvelles mesures, incluant des biomarqueurs représentatifs du système neuroendocrinien, notamment le cortisol, soient réalisées sur les échantillons des femmes de E3N durant la thèse. Cependant, les dosages ont été retardés comme nous en discuterons dans le chapitre VI de cette partie et je n'avais finalement aucun biomarqueur issu du système neuroendocrinien pour les analyses. Par conséquent j'ai considéré chacun des biomarqueurs indépendamment. Compte tenu du nombre important de facteurs associés au risque de cancer du sein dont les mécanismes d'action s'expliquent par la modification des hormones sexuelles, j'ai également utilisé trois hormones sexuelles qui étaient disponibles dans E3N (l'œstradiol, la testostérone et la SHBG). Tous ces biomarqueurs ont précédemment été suspectés dans la littérature d'être associés au risque de cancer du sein et nous intéresse donc pour répondre au deuxième objectif de cette thèse.

J'ai défini une variable à trois classes basée sur les terciles de la distribution de chaque biomarqueur dans la population totale, distinguant des femmes avec une faible concentration en biomarqueurs (tercile 1), une concentration intermédiaire (tercile 2) et une concentration élevée (tercile 3). La proportion de données manquantes pour chacun des biomarqueurs était faible, variant de 0.7% à 6.6%. Une approche conservatrice (biais maximum) a été utilisée pour les individus pour lesquels il y avait des données manquantes, en les considérant systématiquement dans le tercile avec le risque de cancer du sein le plus faible (tercile 1 sauf pour le HDL et la SHBG pour lesquels le tercile le moins à risque est le tercile le plus élevé).

CHAPITRE V. LES COVARIABLES

Les données collectées à partir des différents questionnaires de E3N sont présentées dans le tableau 4. Les covariables ont été choisies sur la base de la littérature selon leur potentiel à pouvoir agir comme facteur de confusion ou comme facteur de médiation dans la relation entre la PSE et le risque de cancer du sein. La majorité des données ont été collectées en début de suivi, au cours des deux premiers questionnaires. Certaines informations, notamment celles concernant les comportements de santé et les mesures anthropométriques, ont été actualisées à différents moments du suivi. Parmi les variables qui ont été collectées plusieurs fois, nous avons sélectionné celles qui étaient au plus proche de la variable d'exposition considérée dans chacun des travaux, c'est-à-dire la PSE pour le travail 3 et 5 (variables à Q1-Q2) et le prélèvement biologique pour le travail 4 (variables à Q5 dont les données manquantes ont été complétées à partir des variables des questionnaires précédents). La majorité des covariables utilisées ont été mesurées avant le diagnostic du cancer, hormis pour les variables de régime occidental et d'alcool, pour lesquelles les données étaient uniquement disponibles à Q3 (environ 10% des cas).

Les sections qui sont décrites dans ce chapitre présentent, pour chaque variable, dans un premier temps la manière dont elles ont été collectées et sont disponibles dans E3N et ensuite un focus est fait sur la manière dont elles sont prises en compte pour la thèse. Nous avons transformé certaines variables continues en variables catégorielles ; ces modifications ont été faites d'après la distribution des variables dans la cohorte générale. Les pourcentages donnés à titre indicatif sont estimés à partir de la population de femmes de la cohorte E3N utilisée pour répondre au troisième travail de cette thèse dont l'objectif était d'étudier le gradient social de cancer du sein à partir d'une population de femmes ayant développé un cancer du sein ainsi que celles étant restées indemnes de cancer (N = 83 436), partie 1, chapitre IV. La proportion de données manquantes est disponible sur la figure 6, partie 3, chapitre I.

TABLEAU 4 : DONNEES COLLECTEES LORS DE L'ENVOI DES DIFFERENTS QUESTIONNAIRES DANS LA COHORTE E3N (ADAPTEE DE CLAVEL-CHAPELON F.& AL, IJE, 2014). LA ZONE GRISEE CORRESPOND AUX DONNEES UTILISEES POUR CE TRAVAIL DE THESE.

	Q1 (06/1990)	Q2 (01/1992)	Q3 (06/1993)	Q4 (01/1995)	Q5 (04/1997)	Q6 (07/2000)	Q7 (07/2002)	Q8 (07/2005)	Q9 (06/2008)	Q10 (12/2011)
Facteurs socio-démographiques										
Niveau d'éducation	X									
Profession		X						X		
Age à l'arrêt de l'activité								X		X
Mesures anthropométriques										
Taille	X			X		X	X	X	X	X
Poids	X	X		X	X	X	X	X	X	X
Tour de hanche				X			X	X	X	X
Tour de taille				X			X	X	X	X
Facteurs reproductifs										
Facteurs menstruelles (âge aux premières règles, durée des cycles, ..)	X	X	X	X	X	X				
Histoire reproductive (nombre d'enfants, âge à chaque grossesse, durée et finalité de la grossesse, allaitement, infertilité, ..)	X	X								
Ménopause (âge, type)	X			X	X	X	X	X		
Traitements hormonaux										
Contraceptifs oraux, THM, ..)		X	X	X	X	X	X	X	X	
Comportements de santé										
Tabac	X	X	X	X	X	X	X	X		X
Alcool			X		X			X		
Activité physique	X				X		X	X		
Alimentation			X					X		
Santé globale										
Historique médical et opératoire	X	X	X	X	X	X	X	X	X	X
Traitements	X	X	X	X	X	X	X	X	X	X
Antécédents familiaux de maladies chroniques	X					X		X		
Santé mentale										
Echelle de dépression (CESD)								X		X
Dépression	X	X	X	X	X	X	X	X	X	X
Outcome de santé	X	X	X	X	X	X	X	X	X	X
Prélèvements biologiques										

1. COMPORTEMENTS DE SANTE

A. CONSOMMATION DE TABAC

Différentes questions relatives à la consommation de tabac des femmes, à la fréquence, type de cigarettes, l'âge au commencement... ont été posées au cours des différents questionnaires. Pour ce travail nous nous sommes principalement intéressés à la consommation tabagique des participantes et nous avons défini une variable à trois modalités : non fumeuse ; ancienne fumeuse ; fumeuse actuelle, à partir des données renseignées à Q1.

Dans notre population, 67.1% des femmes pour qui nous avons l'information n'avaient jamais fumé, 20.1% étaient d'anciennes fumeuses et 12.8% fumaient au moment de l'inclusion.

B. CONSOMMATION D'ALCOOL ET ADHERENCE AU REGIME ALIMENTAIRE OCCIDENTAL

Les questions relatives à la consommation d'alcool et au régime alimentaire sont issues des questionnaires alimentaires qui ont été envoyés à Q3 (1993) et Q8 (2005).

J'ai catégorisé la variable alcool, exprimée en grammes par jour (g/jr), en trois modalités : aucune (0g/jr) ; consommation modérée (≤ 20 g/jr) ; consommation élevée (> 20 g/jr). Cette classification est basée sur les recommandations¹⁷⁷ indiquant un seuil de consommation chez les femmes limité à 14 unités d'alcool par semaine, ce qui équivaut à 140 grammes par semaine, soit 20 g/jr. A Q3, parmi la population d'étude utilisée pour répondre au premier objectif, 9.8% des femmes ne consommaient pas d'alcool, 54.4% avaient une consommation modérée et 15.7% avaient une consommation élevée d'alcool.

Le profil alimentaire a été au préalable généré par une analyse en composantes principales (ACP) par les data-managers de l'équipe E3N et j'ai par la suite créé une variable en terciles d'après la distribution de femmes pour qui l'information était disponible.

C. ACTIVITE PHYSIQUE

L'activité physique a été recueillie aux questionnaires Q1, Q3, Q5, Q7 et Q8 à travers différentes questions relatives à la pratique de la marche, du vélo, du jardinage, d'un sport, du bricolage, de la montée d'étages ou encore du ménage. L'information recueillie à chaque questionnaire, pour chaque activité, a été transformée, par les data-managers de l'équipe E3N, en *metabolic equivalent task* (MET) par heure, ce qui correspond à une dépense énergétique par rapport à la dépense énergétique au repos, d'après les recommandations du Compendium of Physical Activities¹⁷⁸.

Pour la variable d'activité physique utilisée dans ce travail de thèse, j'ai combiné les différentes mesures du questionnaire Q1 afin de définir une variable globale d'activité physique et j'ai utilisé cette variable sous forme de terciles définis en fonction de la distribution des femmes pour lesquelles l'information était disponible. La catégorie allant de 0 à 4.66 MET représente un faible niveau d'activité physique ; celle allant de 4.66 à 7.57 MET représente un niveau moyen et celle allant de 7.57 à 35.9 MET représente un niveau élevé. La répartition des participantes incluses pour répondre au premier objectif était homogène dans chacune des catégories.

2. MESURES ANTHROPOMETRIQUES

A. TAILLE

La taille a été recueillie au cours de multiples questionnaires tout au long du suivi. Dans un premier temps, il était demandé à la participante d'écrire sa taille en clair, en centimètres (cm) puis de répéter l'information en noircissant les cases d'une grille.

Pour ce travail, j'ai créé une variable de taille en définissant des terciles à partir de la distribution des valeurs indiquées par les femmes à Q1. Dans la population sélectionnée pour répondre à l'objectif 1, 45.5% des femmes mesuraient moins de 160 cm, 22.3% mesuraient entre 160 et 164 cm et 32.2% mesuraient plus de 164 cm.

B. POIDS

Le poids a également été recueilli à différents moments au cours du suivi. De la même manière que pour la taille, il était demandé aux participantes d'écrire leur poids en kilogrammes (kg) en clair, puis de répéter l'information en noircissant les cases d'une grille. J'ai également catégorisé cette variable en terciles à partir de l'information renseignée à Q1. La répartition des participantes incluses pour répondre au premier objectif était homogène dans chacune des catégories.

3. FACTEURS REPRODUCTIFS, ANTECEDENTS FAMILIAUX ET DEPISTAGE

A. AGE A LA PREMIERE GROSSESSE MENE E A TERME

Lors du premier questionnaire, il a été demandé aux participantes de renseigner l'âge auquel elles ont eu leur premier enfant. A partir de cette information, j'ai créé une nouvelle variable catégorielle en définissant trois modalités d'après la distribution de l'âge dans la cohorte. Dans notre échantillon d'étude, 35.7% des femmes ont eu leur premier enfant entre 14 et 23 ans, 26.6% entre 23 et 26 ans et 25.4% entre 26 et 59 ans. Dans cette population, 12.3% des femmes n'ont pas eu d'enfants.

B. CONTRACEPTIF ORAL

Lors du premier questionnaire, il a également été demandé aux participantes si elles avaient déjà pris un contraceptif oral au cours de leur vie. Dans notre population d'étude pour répondre à l'objectif 1, 55.8% des femmes avaient déjà eu recours à un contraceptif oral.

C. ALLAITEMENT

Les questions relatives à l'allaitement ont été abordées au cours du premier questionnaire. Elles concernent, d'une part, le fait d'avoir déjà allaité pour au moins une grossesse (oui/non) et d'autre part, la durée cumulée d'allaitement. Dans notre échantillon d'étude, 59% des participantes déclaraient avoir déjà allaité.

D. DUREE DE VIE REPRODUCTIVE

La durée de vie reproductive a été définie à partir de la variable d'âge aux premières règles ainsi que de l'âge à la ménopause.

J'ai créé une variable à trois classes : les femmes avec une durée de vie reproductive longue (≥ 41 ans) qui représentaient 18% de notre effectif, les femmes avec une durée de vie reproductive moyenne (entre 27 et 41 ans) qui comptaient pour 80.4% de notre effectif, et celles avec une durée de vie reproductive très courte (< 27 ans) qui représentaient une très faible proportion de notre population ($\pm 1.6\%$)

E. STATUT MENOPAUSIQUE

J'ai utilisé l'âge à la ménopause ainsi que l'âge des femmes au moment du remplissage du questionnaire Q1 pour définir le statut ménopausique de chaque participante à l'inclusion. Parmi les femmes dont les données ont été utilisées pour l'objectif 1, 58.6% étaient en pré-ménopause.

F. UTILISATION D'UN TRAITEMENT DE LA MENOPAUSE

L'utilisation d'un traitement hormonal de la ménopause (THM), ainsi que la durée de consommation ont été régulièrement demandés au cours du suivi à partir de Q2. Afin de savoir si les participantes prenaient un traitement à Q1, j'ai combiné les deux variables. Si les femmes prenaient un traitement hormonal de la ménopause à Q2 depuis une durée plus longue que celle qui sépare leur réponse à Q1, alors elles étaient exposées au THM à Q1, ce qui représente près de 9% de notre échantillon de l'objectif 1. Si les participantes prenaient un traitement à Q2 mais dont nous ne connaissons pas la durée, alors une troisième modalité a été définie, comptant pour 3.5% de l'effectif.

G. ANTECEDENTS FAMILIAUX DE CANCERS HORMONAUX-DEPENDANTS

Les antécédents familiaux de maladies chroniques ont été actualisés à différents moments au cours du suivi. Je me suis principalement intéressée ici aux antécédents de cancers de l'ovaire et du sein concernant la famille au premier degré (mère, sœurs, filles).

Parmi la population d'étude de l'objectif 1, 1.2% des femmes avaient renseigné des antécédents familiaux de cancer de l'ovaire et 7.5% avaient un proche ayant développé un cancer du sein.

G. MAMMOGRAPHIE

La question relative à la participation au dépistage du cancer du sein a été abordée dès le premier questionnaire. Sur la population de femmes retenue pour le premier objectif, 70.4% avaient déjà subi une mammographie.

Les variables présentées dans cette partie n'ont pas toutes été incluses dans les trois travaux présentés dans la partie résultats. Seules les variables associées à la fois à la PSE et à l'outcome ont été incluses. Certaines variables disponibles dans la cohorte E3N et présentant un intérêt pour ces travaux n'ont pas non plus été incluses, en particulier dans les modèles totalement ajustés, car elles apparaissaient corrélées avec d'autres variables prises en compte. C'est notamment le cas pour le nombre d'enfants.

CHAPITRE VI. RECEPTION ET GESTION DES DONNEES

La récupération des données s'est faite par étapes, tout au long de la thèse, en fonction de l'avancement des différents objectifs.

La réunion de lancement du projet dans lequel s'inscrit ma thèse a eu lieu en janvier 2018. A partir de ce moment, nous avons procédé à une pré-sélection des données que nous souhaitons utiliser parmi l'ensemble très riche d'informations disponibles dans la cohorte E3N. Pour cela nous avons échangé de nombreuses fois avec les data-managers de l'équipe E3N et un gros travail a été nécessaire pour identifier les variables pertinentes à la réalisation du projet. Une fois que nous avons défini ces variables, un temps supplémentaire a été nécessaire pour que les data-managers préparent ces données et les valident car elles n'étaient pas toutes accessibles en l'état. Nous avons donc reçu une première vague de données en janvier 2019 ce qui a permis à ce moment-là de commencer à avancer sur les analyses dans la cohorte E3N. L'ensemble des données nécessaires à la réalisation du premier objectif de ma thèse a été disponible, après nettoyage, en juin 2019. En parallèle j'ai pu finaliser les différents travaux sur l'éducation et l'inflammation ainsi que sur l'inflammation et les cancers qui sont décrits dans la partie 1, chapitre I ce qui m'a permis de mieux développer mon modèle conceptuel liant PSE, fonctionnement biologique et cancer du sein et d'avancer sur les analyses dans la cohorte de réplique EPIC-Italie. Cependant, la réalisation du travail 3 a été plus tardive que prévu initialement.

Les données biologiques précédemment mesurées dans E3N devaient également être nettoyées, ce qui a retardé leur obtention. De plus, du fait de l'épidémie de Covid, les data-managers d'E3N ont été mobilisés sur des projets liés au virus et les nouvelles réorganisations professionnelles ont conduit à des délais supplémentaires dans l'accès pratique à ces données et leur diffusion à cette période. Toutes les données ont néanmoins été disponibles pour analyses en juin 2020. De plus, comme nous l'avons abordé dans le chapitre IV de cette partie, de nouvelles mesures biologiques devaient être réalisées et utilisées dans mes travaux. Cependant, pour des raisons administratives auxquelles c'est ajouté la crise sanitaire, les dosages n'ont pour le moment pas pu être menés.

La sélection des données ainsi que le délai pour leur obtention a décalé le planning que nous avons envisagé, ce qui a pour conséquence que nous n'avons pas pu traiter tous les objectifs

PARTIE 2 - MATERIEL ET METHODES

qui étaient initialement prévus, en particulier l'analyse de la survie, et que la valorisation des travaux déjà menés n'est à ce jour pas complètement aboutie.

CHAPITRE VII. DEMARCHE D'ANALYSES

1. POPULATIONS D'ETUDE

A. TRAVAIL 3 : GRADIENT SOCIAL DE CANCER DU SEIN

Le premier objectif de ma thèse est d'**étudier le gradient social de cancer du sein**.

A partir de la cohorte entière (N = 98 995), j'ai exclu les femmes avec un cancer prévalent (- 6 909) ainsi que les femmes diagnostiquées avec un autre cancer que celui du sein (- 8 650). La population d'étude est donc de **83 436 femmes** (84.3% de la population initiale).

Pour ce travail nous avons répliqué nos analyses dans la composante italienne de l'étude EPIC : EPIC-Italie. Cette cohorte regroupe des hommes et des femmes volontaires issus de cinq centres (Turin, Varèse, Naples, Florence et Ragusa) qui ont été inclus entre 1993-1998. Tous les participants ont signé un formulaire de consentement éclairé à leur entrée dans l'étude. Les comités d'examen éthique du CIRC et le comité d'éthique de l'autorité sanitaire locale de Florence en Italie ont approuvé l'étude. Tout comme pour E3N, de nombreuses informations relatives au style de vie, aux comportements, à la PSE et à l'état de santé des individus ont été recueillies au moment de l'inclusion dans la cohorte. A partir de la cohorte d'origine (N = 47 749), nous avons exclu les 13 597 participants issus du centre de Florence qui ne souhaitait pas participer à l'étude, 11 654 hommes, 4 femmes pour qui nous n'avons pas pu déterminer de temps de suivi, 1 636 femmes avec un cancer d'un autre type que le sein ou sans site précisé et 328 cancers prévalents. La population d'étude était donc de **20 530 femmes**.

B. TRAVAIL 4 : INCORPORATION BIOLOGIQUE DE L'EDUCATION ET RISQUE DE CANCER DU SEIN

Le second objectif se découpe en deux parties :

- (1) **étudier la relation entre certains biomarqueurs et le risque de cancer du sein**, et
- (2) **regarder l'impact de l'éducation sur le niveau de ces biomarqueurs**, conditionnellement aux comportements de santé et aux facteurs reproductifs.

A partir de la population de femmes sélectionnée pour l'objectif 1 (N = 83 436 femmes sans cancer ou ayant développé un cancer du sein), j'ai inclus toutes les femmes pour lesquelles nous disposons de la concentration d'au moins un biomarqueur (N = 3 389). Les femmes qui ont été diagnostiquées (- 9) ou perdues de vue (- 7) avant la collecte de sang ont été retirées des analyses. Les cancers du sein in situ ont également été exclus (- 39) du fait de la difficulté à classer ces anomalies ainsi que d'un manque d'exhaustivité dans leur collecte. De plus, les femmes atteintes de tumeurs en cours d'investigation et celles atteintes de tumeurs concomitantes (- 30) ont également été exclues des analyses. Les femmes en pré-ménopause (- 634) ou celles dont le statut ménopausique est inconnu au moment de la collecte de sang (- 14) et du diagnostic (- 53) ont également été exclues compte tenu de la particularité des cancers diagnostiqués tôt dans la vie (en pré-ménopause) et des modifications biologiques, notamment hormonales, survenant après la ménopause. La population d'étude comprenait **2603 femmes ménopausées sans cancer au moment de la collecte des échantillons sanguins dont certaines ont reçu un diagnostic de cancer du sein en post-ménopause au cours du suivi.**

Compte tenu de la petite taille d'échantillon de la cohorte EPIC-Italie pour lequel les participantes avaient à la fois une mesure des marqueurs biologiques systémiques et hormonaux, des données renseignées pour la PSE et les variables de confusion, ces analyses n'ont pas pu être répliquées.

C. TRAVAIL 5 : ENVIRONNEMENT SOCIAL ET CARACTERISTIQUES DES CANCERS DU SEIN

Le troisième objectif est **d'étudier l'impact de l'éducation sur l'extension du cancer au moment du diagnostic ainsi que son agressivité et la distribution des différents sous-types moléculaires.**

Pour ce travail, j'ai gardé tous les cas de cancer du sein identifiés dans la population définie pour l'objectif 1 (N = 7 877). A partir de cette population, nous avons exclu les femmes qui avaient des tumeurs concomitantes (- 183) et les cancers du sein in situ (- 993) pour ne garder que les tumeurs invasives. La population d'étude est donc de **6 701 femmes.**

2. ANALYSES DESCRIPTIVES

La description des variables utilisées pour répondre à chacun des objectifs a été réalisée à l'aide de la moyenne et de l'écart type pour les variables continues, de la médiane et de l'intervalle interquartiles pour le temps de suivi, ainsi que de la fréquence pour les variables catégorielles.

Les caractéristiques de chacune des populations de chacun des objectifs ont été comparées en fonction des indicateurs de PSE et par statut vis-à-vis du cancer du sein.

- Afin d'estimer les associations entre les variables et la/les PSE, le test du Chi₂ ou celui de Fisher ont été utilisés pour les variables catégorielles ainsi que le test de Student ou le Wilcoxon rank test pour les variables continues.
- Le test du log rank a été utilisé pour tester l'association entre chacune des variables et le risque de cancer du sein.

Pour les analyses multivariées, nous avons inclus les variables qui apparaissent significativement associées à la variable d'exposition et à l'outcome à un seuil de 20%. Certaines variables dont l'effet a été largement démontré dans la littérature ont également été incluses de manière systématique dans les modèles.

3. IMPUTATION MULTIPLE

Malgré un taux de réponse aux différents questionnaires très important, il existe un certain nombre de données manquantes dans les différentes variables utilisées, ce qui peut amener à une perte de précision et à des biais dans nos analyses. Pour éviter cela, j'ai imputé les données manquantes à l'aide des méthodes d'imputation multiple.

Principe

L'imputation multiple permet d'attribuer un certain nombre de valeurs plausibles pour chaque donnée manquante en utilisant pour cela toute l'information disponible et les relations existantes entre les variables de la base de données¹⁷⁹. L'attribution de différentes

valeurs pour une même donnée manquante va permettre de prendre en compte la variabilité autour de chaque donnée imputée.

Application

Pour chacun des travaux, quinze imputations aboutissant à la création de quinze nouvelles bases de données ont été effectuées en considérant l'hypothèse selon laquelle les données manquantes apparaissent de manière aléatoire dans la population étudiée. Toutes les variables que nous avons sélectionnées d'après la littérature et testées comme pouvant jouer un rôle dans l'association principale pour chacune des études ont été utilisées dans le modèle d'imputation de chaque étude (annexe 4).

Les analyses statistiques permettant de répondre à l'objectif ont été ensuite réalisées dans chacune des quinze bases indépendamment.

Enfin, nous avons utilisé les règles de combinaison de Rubin¹⁸⁰ pour obtenir le paramètre final des régressions à partir de ceux des différentes bases de données imputées.

4. MODELES MULTIVARIES

A. MODELES DE COX A RISQUES PROPORTIONNELS

Principe

Le modèle de Cox à risques proportionnels est l'un des modèles les plus utilisés pour la modélisation des données d'analyses de survie¹⁸¹. C'est un modèle semi-paramétrique. Les données de survie résultent d'un processus dynamique dont le but est de modéliser un phénomène dépendant du temps (décès, diagnostic, rechute, ...) en fonction de facteurs spécifiques. Le modèle de Cox exprime la fonction de risque qui représente la probabilité d'apparition de l'évènement dans un intervalle de temps sachant que l'évènement ne s'est pas réalisé avant et d'après les différentes covariables X prises en compte dans le modèle. C'est à partir de cette fonction de risque qu'il est possible de calculer les rapports de risque ou *Hazard Ratio* (HR) qui correspondent au risque que des évènements se produisent dans un groupe plutôt qu'un autre. Si le HR est supérieur à 1, alors le risque de faire l'évènement est plus élevé dans le groupe considéré par rapport au groupe de référence et si le HR est inférieur

à 1, alors le risque est plus faible. Si le HR est égal à 1, alors le risque instantané est le même entre les groupes.

La validité du modèle de Cox dépend de deux hypothèses :

- Hypothèse des risques proportionnels : le rapport de risque entre deux observations pour chacune des covariables est constant et indépendant du temps.
- Hypothèse de log-linéarité : l'augmentation du coefficient β doit augmenter linéairement avec l'augmentation d'une unité de la covariable considérée.

Application

Calcul du temps de suivi : Dans le cadre du troisième travail, pour la cohorte E3N, les femmes ont été suivies depuis l'inclusion dans la cohorte (Q1, 1990) jusqu'au moment du diagnostic de cancer ou du dernier questionnaire rempli pour les femmes n'ayant pas développé de cancer du sein. Pour la cohorte EPIC-Italie, le suivi a commencé à l'inclusion (1993-1998) jusqu'au dernier suivi : décembre 2010 pour les femmes de Naples, Turin et Ragusa, et jusqu'en décembre 2009 pour les femmes de Varèse. Dans le quatrième travail mené sur la cohorte E3N, le temps de suivi à partir de la collecte de sang a été utilisé comme échelle de temps.

Les femmes ayant quitté l'étude avant la fin du suivi ont été censurées à la date de dernières nouvelles et celles ayant répondu uniquement au questionnaire d'inclusion ont été censurées après un jour de suivi.

Les hypothèses du modèle de Cox ont été testées dans chacun des travaux et les variables ont été transformées en cas de non-respect des hypothèses.

J'ai utilisé des modèles de Cox dans les travaux 3 et 4 de cette thèse : d'abord pour étudier l'association entre la PSE et la mobilité sociale avec le risque futur de cancer du sein et pour regarder le lien entre différents marqueurs biologiques et le risque de cancer du sein. Dans ces travaux, nous avons pris comme référence une PSE moins avantagée et une faible concentration de biomarqueurs (tercile 1 comme cela a été décrit dans le chapitre IV).

B. MODELES DE REGRESSIONS LOGISTIQUES MULTINOMIALES

Principe

Le modèle de régression logistique multinomiale, ou polytomique, est une extension du modèle de régression logistique adaptée à une variable dépendante comportant plus de deux modalités¹⁸². C'est un modèle linéaire généralisé utilisant une fonction *logit* comme fonction de lien. Ce modèle permet de comparer le risque entre différents individus d'appartenir à une catégorie de la variable dépendante suivant leur niveau d'exposition par rapport à une catégorie de référence. La quantification du risque se fait à travers le rapport de côtes ou Odds Ratio (OR). Si l'OR est supérieur à 1, alors l'évènement est plus fréquent dans le groupe considéré par rapport au groupe de référence et si l'OR est inférieur à 1, alors l'évènement est moins fréquent. Si l'OR est égal à 1, alors la fréquence est la même dans les deux groupes. Dans le cadre des modèles de régressions logistiques multinomiales où la variable dépendante aurait trois modalités (référence R, la modalité M₁ et la modalité M₂), l'estimation serait faite en deux temps : M₁ par rapport à R et M₂ par rapport à R.

Application

J'ai utilisé ces modèles de régressions logistiques multinomiales dans les travaux 4 et 5 de cette thèse : d'abord pour étudier l'association entre l'éducation et différents marqueurs biologiques et ensuite pour regarder le lien entre l'éducation et les caractéristiques initiales des cancers du sein. Pour le cinquième travail, à l'inverse des travaux 3 et 4, nous avons mis en référence la PSE la plus avantagée. Pour ce travail, nous avons également mis en référence les caractéristiques de meilleur pronostic.

PARTIE 3 - RESULTATS

CHAPITRE I. IMPACT DE L'ENVIRONNEMENT SOCIAL SUR L'INCIDENCE DU CANCER DU SEIN

TRAVAIL 3 : DETERMINANTS SOCIAUX DU RISQUE DE CANCER DU SEIN

Dans ce chapitre, je présente l'étude du gradient social de cancer du sein dans la cohorte prospective E3N, mais également dans une cohorte de réplication, EPIC-Italie. Je me suis particulièrement intéressée à l'environnement social à trois moments de la vie et j'ai utilisé une approche « life course » afin de tenir compte de la chronologie des expositions. Je me suis focalisée principalement dans ce troisième travail sur les mécanismes indirects impliquant un rôle des facteurs de risque socialement stratifiés pour expliquer ces inégalités sociales.

Cette section a fait l'objet d'une publication « *The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk* », parue dans **BMC Cancer** [Annexe 5]

1. CONTEXTE

Des études précédentes ont mis en évidence que les femmes favorisées ont un risque de cancer du sein plus important^{119,120}. Ces inégalités sociales face au risque de cancer peuvent s'expliquer par la distribution socialement stratifiée des facteurs de risque connus de cancer du sein¹⁸³⁻¹⁸⁷, notamment l'âge à la première grossesse^{121-124,188-191}. Cependant, une association semble persister malgré la prise en compte des facteurs reproductifs¹²¹⁻¹²⁴ ce qui peut s'expliquer par le fait que peu d'études ont, dans un même temps, pris en compte les comportements de santé et les mesures anthropométriques en plus des facteurs reproductifs pour expliquer cette association. De plus, plusieurs études ont suggéré que la participation au dépistage pourrait médier l'association entre la PSE et le risque de cancer du sein^{121,125,189,192}, bien que les quelques études qui ont considéré ce facteur n'ont pas mis en évidence d'effet notable.

Il est également important de noter que la plupart des études se sont intéressées à la PSE à partir de l'entrée dans la vie adulte^{121,123,125,188-193}. Or, il apparaît important de prendre en

RESEARCH ARTICLE

Open Access

The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk



Eloïse Berger^{1*}, Noële Maitre^{2,3}, Francesca Romana Mancini^{2,3}, Laura Baglietto^{2,3,4}, Vittorio Perduca^{2,3,5}, Héléne Colineaux^{1,6}, Sabina Sieri⁷, Salvatore Panico⁸, Carlotta Sacerdote⁹, Rosario Tumino¹⁰, Paolo Vineis^{11,12}, Marie-Christine Boutron-Ruault^{2,3}, Gianluca Severi^{2,3,13}, Raphaële Castagné^{1†} and Cyrille Delpierre^{1†}

Abstract

Background: Women with an advantaged socioeconomic position (SEP) have a higher risk of developing breast cancer (BC). The reasons for this association do not seem to be limited to reproductive factors and remain to be understood. We aimed to investigate the impact of lifecourse SEP from childhood and social mobility on the risk of BC considering a broad set of potential mediators.

Methods: We used a discovery-replication strategy in two European prospective cohorts, E3N ($N = 83,436$) and EPIC-Italy ($N = 20,530$). In E3N, 7877 women were diagnosed with BC during a median 24.4 years of follow-up, while in EPIC-Italy, 893 BC cases were diagnosed within 15.1 years. Hazard ratios (HR) were estimated using Cox proportional hazard models on imputed data.

Results: In E3N, women with higher education had a higher risk of BC (HR [95%CI] = 1.21 [1.12, 1.30]). This association was attenuated by adjusting for reproductive factors, in particular age at first childbirth (HR[95%CI] = 1.13 [1.04, 1.22]). Health behaviours, anthropometric variables, and BC screening had a weaker effect on the association. Women who remained in a stable advantaged SEP had a higher risk of BC (HR [95%CI] = 1.24 [1.07; 1.43]) attenuated after adjustment for potential mediators (HR [95%CI] = 1.13 [0.98; 1.31]). These results were replicated in EPIC-Italy.

Conclusions: These results confirm the important role of reproductive factors in the social gradient in BC risk, which does not appear to be fully explained by the large set of potential mediators, including cancer screening, suggesting that further research is needed to identify additional mechanisms.

Keywords: Lifecourse socio-economic position, Social mobility, Breast cancer, Prospective cohorts

* Correspondence: eloise.berger@univ-tlse3.fr

†Raphaële Castagné and Cyrille Delpierre: these authors jointly supervised this work.

¹UMR LEASP, Université de Toulouse III, UPS, Inserm, Toulouse, France
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

compte l'effet combiné de la PSE au cours de la vie en plus de l'effet individuel des PSE car cela pourrait révéler différents mécanismes impliqués dans le gradient social de l'incidence de cancer du sein.

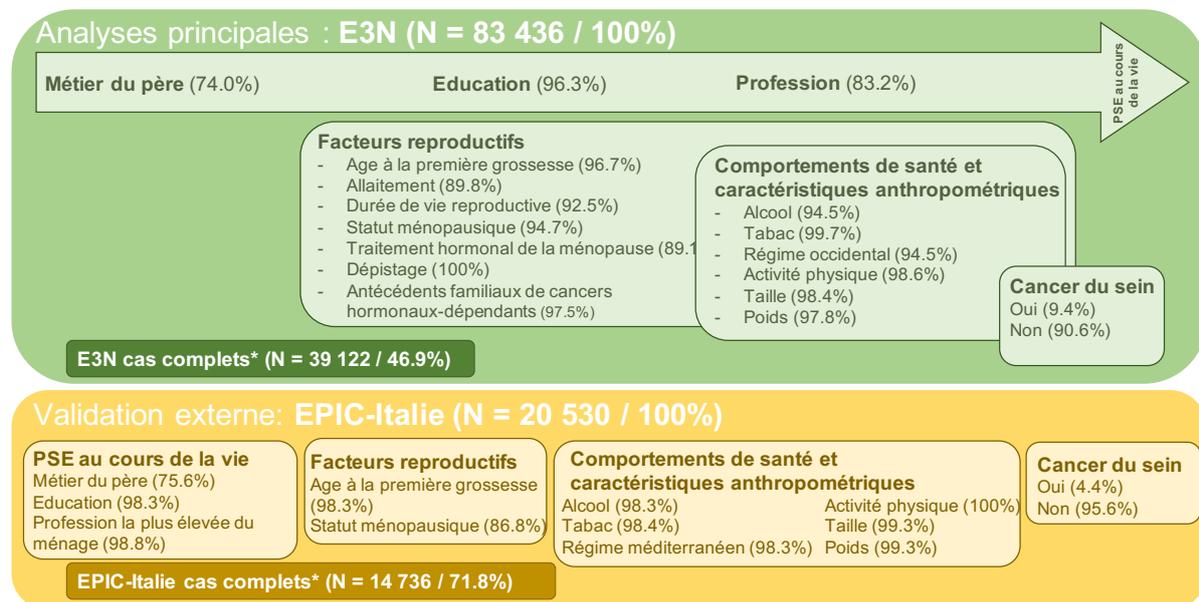
Objectifs de l'étude : Étudier l'association entre la PSE des femmes au cours de la vie (enfance, jeune adulte et âge adulte) ainsi que la mobilité sociale et le risque de cancer du sein. Il s'agira également d'étudier l'impact potentiel et mutuel d'un large ensemble de facteurs reproductifs, de comportements de santé et de mesures anthropométriques et du dépistage dans la cadre d'un parcours de vie.

Hypothèses : Il existe des inégalités sociales face au risque de cancer du sein, les femmes plus socialement avantagées étant plus à risque que les femmes moins avantagées. Nous faisons également l'hypothèse que ces inégalités s'expliquent en partie par des facteurs de risque, notamment reproductifs, mais qu'une partie du lien pourrait persister, reflétant l'existence d'autres mécanismes. Le dépistage pourrait être un facteur important pour expliquer ce gradient social.

2. METHODE ET SCHEMA D'ETUDE

Les analyses ont été réalisées sur la cohorte E3N et répliquées sur la cohorte EPIC-Italie. Les données utilisées dans le cadre de cette étude sont détaillées, pour chacune des cohortes, sur la figure 6. Malgré un nombre plus restreint de variables d'ajustement dans la cohorte EPIC-Italie, nous avons essayé d'être le plus cohérent possible entre les deux cohortes. Deux différences notables apparaissent cependant : au vu du nombre important de femmes au foyer dans la cohorte EPIC-Italie, la PSE à l'âge adulte a été approchée par la profession la plus élevée du ménage. L'adhérence au régime méditerranéen a été utilisée comme variable alimentaire.

FIGURE 6 : APERÇU DES DONNÉES SÉLECTIONNÉES DANS LES DEUX COHORTES ET PROPORTION DES DONNÉES DISPONIBLES POUR CHACUNE DES POPULATIONS ÉTUDIÉES. LE POURCENTAGE REPRÉSENTE LA PROPORTION DE FEMMES DISPOSANT DE DONNÉES DISPONIBLES POUR CHAQUE COVARIABLE DANS LA POPULATION AVANT IMPUTATION.



Nous avons mené les analyses multivariées en trois étapes, d'après les analyses statistiques détaillées dans la partie 2, chapitre VII :

- dans un premier temps, j'ai testé l'association entre chacun des trois indicateurs de PSE et le risque de cancer du sein et j'ai étudié l'impact des facteurs d'ajustement sur cette association à l'aide des modèles de Cox. Les associations globales entre les PSE et le risque de cancer du sein ont été ajustées sur l'âge (modèle 1). À partir du modèle 1, nous avons fait des modèles ajustés indépendamment sur chacune des covariables ainsi que des modèles ajustés par bloc de variables (d'une part sur tous les comportements de santé et les caractéristiques anthropométriques et d'autre part sur tous les facteurs reproductifs). Les modèles ont ensuite été ajustés sur l'ensemble des covariables (modèle 2).
- dans un second temps j'ai utilisé une approche *life course* en ajustant séquentiellement sur les trois PSE afin de tenir compte de l'ordre chronologique des expositions.
- enfin j'ai étudié l'impact de la mobilité sociale (ascendante, descendante ou stable) sur le risque de cancer du sein.

Ces analyses ont été réalisées dans la cohorte E3N, sur données imputées (N = 83 436) et une analyse de sensibilité a été réalisée sur les femmes n'ayant aucune donnée manquante pour tester la robustesse de nos résultats (N = 39 122). Pour tenir compte des spécificités potentielles des cancers du sein invasifs par rapport aux cancers in situ, nous avons également effectué ces analyses chez les femmes atteintes d'un cancer invasif uniquement (87.6% de tous les cas de cancer du sein incidents dans E3N). Ces analyses ont ensuite été répliquées dans EPIC-Italie, après imputation des données manquantes (N = 20 530).

3. RESULTATS

A. CARACTERISTIQUES DE LA POPULATION D'ETUDE

Le tableau 5 présente certaines caractéristiques de la population de femmes issue de la cohorte E3N selon le statut vis-à-vis du cancer du sein. Au cours d'une période de suivi médiane de 24,4 ans, 7 877 femmes ont reçu un diagnostic de cancer du sein. Par rapport aux femmes qui n'ont pas développé ce type de cancer, celles qui en ont été atteintes étaient plus susceptibles d'avoir une PSE avantagée au cours de leur vie, d'être plus âgées, d'avoir une consommation élevée d'alcool, de suivre un régime alimentaire occidental, de pratiquer peu ou pas d'activité physique, et d'être plus grandes et avec un poids plus élevé. Les femmes qui ont au cours du suivi été diagnostiquées d'un cancer du sein étaient également plus susceptibles d'avoir des antécédents familiaux de cancers hormonaux dépendants, d'avoir déjà utilisé un traitement hormonal de la ménopause, d'avoir eu moins d'enfants, d'avoir eu un premier accouchement à un âge plus avancé, de ne pas avoir allaité, d'avoir une plus longue durée de vie reproductive, d'être en pré-ménopause à l'inclusion et d'avoir déjà participé à une mammographie avant l'inclusion. Les caractéristiques par PSE sont disponibles dans l'annexe 6.1.

TABLEAU 5 : CARACTERISTIQUES DES FEMMES AVEC LES DONNEES DISPONIBLES DANS E3N SELON LE STATUT VIS-A-VIS DU CANCER DU SEIN.

Variables	Cancer du sein		p
	Non	Oui	
Métier du père (3cl)	n= 55766	n= 5957	0.025
Avantagé, n(%)	9294 (16.67)	1075 (18.05)	
Intermédiaire, n(%)	23619 (42.35)	2492 (41.83)	
Moins avantagé, n(%)	22853 (40.98)	2390 (40.12)	
Education (3cl)	n= 72773	n= 7577	<0.001
Elevé, n(%)	26096 (35.86)	2913 (38.45)	
Moyen, n(%)	36422 (50.05)	3745 (49.43)	
Faible, n(%)	10255 (14.09)	919 (12.13)	
Profession (3cl)	n= 62726	n= 6700	0.017
Avantagée, n(%)	10962 (17.48)	1250 (18.66)	
Intermédiaire, n(%)	40679 (64.85)	4332 (64.66)	
Moins avantagée, n(%)	11085 (17.67)	1118 (16.69)	
Age	n= 75559	n= 7877	<0.001
< 47.9 ans, n(%)	39491 (52.27)	3837 (48.71)	
≥ 47.9 ans, n(%)	36068 (47.73)	4040 (51.29)	
Consommation d'alcool	n= 71357	n= 7455	<0.001
Aucune, n(%)	7040 (9.87)	678 (9.09)	
Modérée, n(%)	38720 (54.26)	4167 (55.9)	
Élevée, n(%)	10975 (15.38)	1356 (18.19)	
Non répondante à Q3, n(%)	14622 (20.49)	1254 (16.82)	
Tabagisme	n= 75333	n= 7850	0.693
Jamais, n(%)	50531 (67.08)	5269 (67.12)	
Ancienne fumeuse, n(%)	15204 (20.18)	1560 (19.87)	
Fumeuse actuelle, n(%)	9598 (12.74)	1021 (13.01)	
Régime occidental	n= 71358	n= 7455	<0.001
[-2.93,-0.511], n(%)	18876 (26.45)	1921 (25.77)	
(-0.511,0.307], n(%)	18868 (26.44)	2089 (28.02)	
(0.307,7.78], n(%)	18991 (26.61)	2191 (29.39)	
Non répondante à Q3, n(%)	14623 (20.49)	1254 (16.82)	
Activité physique (MET)	n= 74540	n= 7768	<0.001
(7.57,35.9], n(%)	25040 (33.59)	2395 (30.83)	
(4.66,7.57], n(%)	24635 (33.05)	2630 (33.86)	
[0,4.66], n(%)	24865 (33.36)	2743 (35.31)	
Taille	n= 74316	n= 7759	<0.001
[135,160], n(%)	33983 (45.73)	3355 (43.24)	
(160,164], n(%)	16551 (22.27)	1732 (22.32)	
(164,190], n(%)	23782 (32)	2672 (34.44)	

PARTIE 3 - RESULTATS

Poids		n= 73847	n= 7715	0.076
	[29,55], n(%)	25135 (34.04)	2530 (32.79)	
	(55,62], n(%)	24737 (33.5)	2611 (33.84)	
	(62,163], n(%)	23975 (32.47)	2574 (33.36)	
Antécédents familiaux de cancer de l'ovaire (1grade)		n= 75559	n= 7877	0.002
	Non, n(%)	74709 (98.88)	7757 (98.48)	
	Oui, n(%)	850 (1.12)	120 (1.52)	
Antécédents familiaux de cancer du sein (1grade)		n= 73693	n= 7673	<0.001
	Non, n(%)	31622 (42.91)	3039 (39.61)	
	Oui, n(%)	5240 (7.11)	886 (11.55)	
	Indéfini, n(%)	36831 (49.98)	3748 (48.85)	
Traitement hormonal de la ménopause (THM)		n= 67245	n= 7111	<0.001
	Non, n(%)	58868 (87.54)	6021 (84.67)	
	Oui, n(%)	6029 (8.97)	806 (11.33)	
	Indéfini, n(%)	2348 (3.49)	284 (3.99)	
Nombre de grossesses menées à terme		n= 75551	n= 7877	<0.001
	3+, n(%)	22212 (29.4)	2000 (25.39)	
	1-2, n(%)	44491 (58.89)	4827 (61.28)	
	0, n(%)	8848 (11.71)	1050 (13.33)	
Age à la première grossesse		n= 73018	n= 7660	<0.001
	[14,23], n(%)	26390 (36.14)	2445 (31.92)	
	(23,26], n(%)	19487 (26.69)	1959 (25.57)	
	(26,59], n(%)	18293 (25.05)	2206 (28.8)	
	Pas de grossesse, n(%)	8848 (12.12)	1050 (13.71)	
Allaitement		n= 67708	n= 7189	0.019
	Oui, n(%)	40113 (59.24)	4156 (57.81)	
	Non, n(%)	27595 (40.76)	3033 (42.19)	
Durée de vie reproductive		n= 69697	n= 7508	<0.001
	[27; 41[, n(%)	56083 (80.47)	5964 (79.44)	
	<27, n(%)	1161 (1.67)	80 (1.07)	
	≥41, n(%)	12453 (17.87)	1464 (19.5)	
Dépistage du cancer du sein		n= 75559	n= 7877	<0.001
	Oui, n(%)	52720 (69.77)	6012 (76.32)	
	Non, n(%)	22839 (30.23)	1865 (23.68)	
Statut ménopausique		n= 71368	n= 7675	0.016
	Pré-ménopause, n(%)	41683 (58.41)	4593 (59.84)	
	Post-ménopause, n(%)	29685 (41.59)	3082 (40.16)	

B. IMPACT DE CHACUNE DES PSE SUR LE RISQUE DE CANCER DU SEIN

Les femmes dont le père avait une PSE plus avantagee présentaient un risque plus élevé de développer un cancer du sein par rapport aux femmes dont le père avait une PSE moins avantagee (M1 : HR [95%CI] = 1,10 [1,02, 1,18], tableau 6A). Les associations étaient légèrement atténuées après ajustement sur les comportements de santé et les caractéristiques anthropométriques (HR [95%CI] = 1,08 [1,01, 1,16], tableau 6A), les facteurs reproductifs (HR [95%CI] = 1,05 [0,98, 1,13], tableau 6A), ou sur l'ensemble des covariables (M2, HR [95%CI] = 1,04 [0,97, 1,12], tableau 6A). Les femmes dont le père avait une PSE intermédiaire n'étaient pas plus à risque de cancer du sein.

Une association positive était observée entre le niveau d'éducation des femmes et le risque de cancer du sein (M1 : HR [95%CI] = 1,21 [1,12, 1,30], tableau 6B). L'ajustement sur les comportements de santé et les caractéristiques anthropométriques n'a que légèrement atténué les associations (HR [95%CI] = 1,18 [1,09, 1,27], tableau 6B), tandis que l'ajustement sur les facteurs reproductifs, en particulier l'âge à la première grossesse, l'a affecté plus fortement (HR [95%CI] = 1,06 [0,99, 1,15], tableau 6B). Les HR ont davantage été réduits après l'ajustement sur toutes les covariables (M2 : HR [95%CI] = 1,06 [0,98, 1,14], tableau 6B).

Par rapport à leurs homologues moins avantagees, les femmes exerçant une profession plus avantagee présentaient un risque plus élevé de cancer du sein (HR [95%CI] = 1,08 [1,00, 1,17], tableau 6C). Les HR étaient légèrement atténués après ajustement sur chaque bloc de covariables indépendamment ou pour les deux (M2, HR [95%CI] = 1,05 [0,97, 1,14], tableau 6C). Les femmes avec une PSE intermédiaire n'étaient pas plus à risque de cancer du sein.

Parmi l'ensemble des autres covariables, un âge plus avancé, une consommation d'alcool élevée, une plus grande adhésion à un régime alimentaire occidental, une activité physique plus faible, une taille plus élevée, des antécédents familiaux de cancers hormonaux dépendants, l'utilisation d'un traitement de la ménopause et un âge plus avancé à la première grossesse ont été associés à un risque plus élevé de cancer du sein dans le modèle totalement ajusté. A l'inverse, les femmes ayant une durée de vie reproductive plus courte, celles qui n'ont jamais participé au dépistage du cancer du sein et celles qui sont en post-ménopause ont un risque plus faible de développer un cancer du sein. (annexe 6.2).

TABLEAU 6 : RESULTATS DES MODELES DE COX A RISQUES PROPORTIONNELS SUR LE RISQUE DE CANCER DU SEIN A PARTIR DES DONNEES IMPUTEES DE E3N (N = 83 436)

		A. Métier du père		B. Education ^b		C. Profession ^a	
		Intermédiaire	Avantagé	Moyen	Élevé	Intermédiaire	Avantagée
		HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
M1		1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.01, 1.17]	1.21 [1.12, 1.30]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
Comportements de santé et mesures anthropométriques	M1 + Alcool	1.00 [0.95, 1.06]	1.09 [1.02, 1.17]	1.09 [1.02, 1.17]	1.20 [1.11, 1.29]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Tabac	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.02, 1.17]	1.21 [1.12, 1.30]	1.03 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Régime occidental	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.10 [1.02, 1.18]	1.21 [1.13, 1.31]	1.03 [0.97, 1.09]	1.09 [1.00, 1.18]
	M1 + Activité physique	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.00, 1.16]	1.18 [1.10, 1.28]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + Taille	1.00 [0.94, 1.05]	1.09 [1.01, 1.17]	1.09 [1.01, 1.17]	1.19 [1.11, 1.28]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + Poids	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.10 [1.02, 1.18]	1.21 [1.13, 1.31]	1.03 [0.97, 1.09]	1.08 [1.00, 1.17]
	M1 + tous les comportements de santé et caractéristiques anthropométriques	1.00 [0.94, 1.05]	1.08 [1.01, 1.16]	1.08 [1.00, 1.16]	1.18 [1.09, 1.27]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
Facteurs reproductifs	M1 + Antécédents familiaux cancer des ovaires	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.01, 1.17]	1.20 [1.12, 1.30]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Antécédents familiaux cancer du sein	1.00 [0.94, 1.05]	1.09 [1.01, 1.17]	1.09 [1.01, 1.17]	1.19 [1.11, 1.29]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + THM	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.08 [1.01, 1.16]	1.20 [1.11, 1.29]	1.02 [0.96, 1.08]	1.08 [0.99, 1.16]
	M1 + Age à la première grossesse	0.99 [0.94, 1.05]	1.07 [1.00, 1.15]	1.07 [1.00, 1.15]	1.13 [1.04, 1.22]	1.01 [0.95, 1.07]	1.07 [0.98, 1.16]
	M1 + Allaitement	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.02, 1.17]	1.21 [1.12, 1.30]	1.03 [0.97, 1.09]	1.08 [1.00, 1.17]
	M1 + Durée de vie reproductive	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.01, 1.16]	1.19 [1.11, 1.29]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + Dépistage	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.00, 1.16]	1.19 [1.10, 1.28]	1.02 [0.96, 1.08]	1.08 [1.00, 1.17]
M1 + Statut ménopausique	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.08 [1.01, 1.16]	1.19 [1.11, 1.28]	1.02 [0.96, 1.08]	1.08 [1.00, 1.17]	
Modèle 1 + tous les facteurs reproductifs	0.98 [0.93, 1.04]	1.05 [0.98, 1.13]	1.03 [0.96, 1.11]	1.06 [0.99, 1.15]	0.99 [0.93, 1.05]	1.06 [0.98, 1.15]	
M2		0.98 [0.93, 1.04]	1.04 [0.97, 1.12]	1.03 [0.96, 1.11]	1.06 [0.98, 1.14]	0.99 [0.93, 1.05]	1.05 [0.97, 1.14]

M1 est ajusté sur l'âge.

^aRéférence: "Moins avantagée"

^bRéférence: "Faible niveau d'éducation"

M2 est ajusté sur l'ensemble des covariables

Hazard ratio (HR) et les intervalles de confiance sont reportés pour (A) le métier du père (B) l'éducation et (C) la profession des femmes.

C. PSE AU COURS DE LA VIE ET RISQUE DE CANCER DU SEIN

Lorsque nous avons ajusté séquentiellement sur chacune des PSE à différents moments de la vie, seule l'éducation était associée au risque de cancer du sein. Les HR pour l'éducation n'ont pas été affectés lors de l'ajustement sur la PSE pendant l'enfance (modèle B, HR [95 % IC] = 1,19 [1,11 ; 1,29], tableau 7) ou après prise en compte dans un même temps de la PSE de l'enfance et à l'âge adulte (modèle C, HR [95 % IC] = 1,23 [1,12 ; 1,35], tableau 7). L'association a été atténuée après ajustement sur l'âge à la première grossesse (HR [95%CI] = 1,14 [1,04 ; 1,25]) et les HR ont été davantage affaiblis dans le modèle totalement ajusté (modèle D, HR [95%CI] = 1,06 [0,97 ; 1,17], tableau 7).

D. MOBILITE SOCIALE ET RISQUE DE CANCER DU SEIN

Les femmes ayant bénéficié d'une PSE stable avantaagée tout au long de leur vie présentaient un risque plus élevé de développer un cancer du sein par rapport aux femmes ayant toujours vécu dans un environnement moins avantaagé (Modèle1 : HR [95%CI] = 1,24 [1,07 ; 1,43], tableau 8). Les HR ont été atténués, surtout après l'ajustement sur l'âge à la première grossesse (HR [95%CI] = 1,19 [1,03 ; 1,37]), et après avoir tenu compte de l'ensemble des covariables (modèle entièrement ajusté : HR [95%CI] = 1,13 [0,98 ; 1,31], tableau 8).

TABLEAU 7 : RESULTATS DES REGRESSIONS APRES AJUSTEMENT CHRONOLOGIQUE DES PSE A PARTIR DES DONNEES IMPUTEES DE E3N (N = 83 436)

Covariables	Références	Modalités	E3N			
			Modèle A	Modèle B	Modèle C	Modèle D
			HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Métier du père	Moins avantageé	Intermédiaire	1.00 [0.95; 1.06]	0.98 [0.93; 1.04]	0.98 [0.93; 1.04]	0.98 [0.93; 1.04]
		Avantageé	1.10 [1.02; 1.18]	1.06 [0.98; 1.14]	1.05 [0.98; 1.13]	1.03 [0.96; 1.11]
Education	Faible	Moyen	-	1.09 [1.01; 1.17]	1.12 [1.02; 1.22]	1.04 [0.95; 1.14]
		Élevé	-	1.19 [1.11; 1.29]	1.23 [1.12; 1.35]	1.06 [0.97; 1.17]
Profession	Moins avantageée	Intermédiaire	-	-	0.95 [0.88; 1.02]	0.97 [0.90; 1.04]
		Avantageée	-	-	1.00 [0.91; 1.09]	1.03 [0.94; 1.13]

Modèle A est ajusté sur l'âge et le métier du père

Modèle B est ajusté sur l'âge, le métier du père et l'éducation

Modèle C est ajusté sur l'âge et les trois PSE

Modèle D est ajusté sur l'âge, les trois PSE and toutes les covariables

TABLEAU 8 : ASSOCIATION ENTRE LA MOBILITE SOCIALE ET LE RISQUE DE CANCER DU SEIN A PARTIR DES DONNEES IMPUTEES DE E3N (N = 83 436)

Covariable	Référence	Modalités	E3N			
			Modèle 1	Modèle 1 + HB-A	Modèle 1 + RF	Modèle totalement ajusté
			HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Mobilité sociale	PSE stable moins avantagée	Mobilité descendante	1.06 [0.95; 1.19]	1.04 [0.93; 1.17]	1.00 [0.90; 1.12]	1.00 [0.89; 1.12]
		PSE stable intermédiaire	1.02 [0.92; 1.12]	1.01 [0.91; 1.11]	0.96 [0.87; 1.06]	0.96 [0.87; 1.06]
		Mobilité ascendante	1.04 [0.94; 1.14]	1.03 [0.93; 1.13]	1.00 [0.91; 1.10]	0.99 [0.90; 1.09]
		PSE stable avantagée	1.24 [1.07; 1.43]	1.20 [1.04; 1.39]	1.15 [0.99; 1.33]	1.13 [0.98; 1.31]

Modèle 1 est ajusté sur l'âge et la mobilité sociale

Modèle 1 + HB-A est ajusté sur l'âge, la mobilité sociale, l'alcool, le tabac, l'activité physique, le régime occidental, la taille, le poids

Modèle 1 + RF est ajusté sur l'âge, la mobilité sociale, les antécédents familiaux de cancer des ovaires ou du sein, la prise de THM, l'allaitement, le dépistage, la durée de vie reproductive, l'âge à la première grossesse et le statut à la ménopause

Le modèle complètement ajusté est ajusté sur l'ensemble des variables précédemment listées.

E. ANALYSES DE SENSIBILITE ET VALIDATION EXTERNE

Les caractéristiques de la population de cas complets sont données en annexe 6.3. Les analyses sur les données en cas complets ont montré des résultats similaires, bien que les associations avec le risque de cancer du sein soient légèrement plus fortes pour l'éducation et la profession des participantes (annexe 6.4). Les HR pour l'éducation ont été atténués mais pas entièrement expliqués après avoir pris en compte l'ensemble des covariables. Un schéma similaire a été observé pour la profession des participantes. Lorsque j'ai restreint nos analyses aux femmes ayant développé un cancer du sein invasif, les résultats obtenus étaient comparables (annexe 6.5).

Les résultats des analyses réalisées dans la cohorte EPIC-Italie sont fournis dans les annexes de 6.6 à 6.9. En résumé, les HR pour le risque de cancer du sein chez les femmes éduquées étaient similaires à ceux observés dans E3N (Modèle 1 : HR [95%CI] = 1,19 [0,96 ; 1,47], annexe 6.7). L'ajustement sur l'âge à la première grossesse a principalement atténué cette association (HR [95%CI] = 1,05 [0,84 ; 1,31]). L'association entre l'éducation et le risque de cancer du sein n'a pas été affectée par l'ajustement sur les autres indicateurs du PSE comme nous l'avons observé dans E3N (Modèle C : HR [95%CI] = 1,18 [0,94 ; 1,49], annexe 6.8). Les femmes qui ont toujours bénéficié d'une PSE intermédiaire ou avantagée avaient un risque plus élevé de développer un cancer du sein par rapport aux femmes qui ont toujours vécu dans un environnement moins avantagé, et les HR étaient marginalement affectés après ajustement sur l'ensemble des covariables (annexe 6.9).

4. DISCUSSION

Les résultats ont mis en évidence que les femmes ayant une PSE plus avantagée avaient un risque plus élevé de développer un cancer du sein et ce, quelle que soit la période considérée. Les analyses prenant en compte l'impact des PSE de manière globale dans le parcours de vie ont suggéré que l'éducation avait un effet plus important sur le risque de cancer du sein. Les analyses de mobilité sociale ont indiqué que les femmes qui restent dans une PSE stable et avantagée ont un risque accru par rapport aux femmes qui ont toujours vécu dans une PSE moins favorisée. Les associations sont en partie expliquées par les facteurs reproductifs et notamment l'âge à la première grossesse dans nos modèles. L'ajustement sur la participation

au dépistage du cancer du sein, sur les comportements de santé ou les caractéristiques anthropométriques ne modifiait que marginalement l'association entre l'éducation et le risque de cancer du sein. Les résultats étaient robustes face aux analyses de sensibilité que nous avons menées, et renforcés par les résultats obtenus dans la cohorte de réplication EPIC-Italie.

Notre étude s'ajoute aux précédents travaux qui ont été menés sur le gradient social de cancer du sein. En accord avec les études précédentes, l'éducation apparaît comme étant l'indicateur de PSE le plus fortement associé au risque de cancer du sein^{122,124,125} et l'association est principalement expliquée par les facteurs reproductifs^{122-124,188-191}. Mais nos résultats montrent également que cette association persiste même après avoir pris en compte les principaux autres médiateurs potentiels, ce qui constitue un résultat original de notre travail.

Nous avons d'abord émis l'hypothèse que la participation au dépistage pouvait jouer un rôle important dans le lien entre l'éducation et le risque de cancer du sein et que ce facteur pouvait contribuer à expliquer l'association indépendante qui persiste après avoir pris en compte les facteurs reproductifs. Cependant, l'ajustement sur la participation au dépistage n'a eu qu'un effet très modeste sur l'association, en accord avec les quelques études précédentes qui avaient tenu compte de ce facteur^{190,193}. L'ajustement sur les comportements de santé et les caractéristiques anthropométriques a également eu un effet très modeste sur l'association entre l'éducation et le risque de cancer du sein.

Selon la littérature, certains des facteurs de risque hormonaux de cancer du sein surviennent tôt dans la vie, tels que le poids à la naissance ou l'âge aux premières règles. Ces facteurs ont également été reconnus comme étant associés à la PSE¹⁹⁴⁻¹⁹⁶. Cela suggère que la PSE au début de la vie pourrait avoir un impact important sur le développement du cancer du sein à l'âge adulte. Nos résultats ne supportent pas l'hypothèse d'une période sensible précoce socialement différenciée dans le risque de cancer du sein. Les résultats sur la mobilité sociale montrent que les femmes qui ont bénéficié d'une PSE favorable tout au long de leur vie sont le groupe le plus à risque de cancer du sein, ce qui suggère que ces femmes pourraient cumuler les effets néfastes au cours de leur vie.

Dans l'ensemble, ce travail s'ajoute à la littérature précédente en soulignant l'importance des facteurs reproductifs dans le gradient social de cancer du sein, mais suggère également que d'autres voies et mécanismes sont impliqués et restent à caractériser.

Dans la cohorte E3N, les femmes avec une PSE avantagée présentaient un risque accru de cancer du sein. Parmi les trois indicateurs de PSE étudiés, l'éducation était le facteur le plus fortement lié au risque de cancer du sein. Compte tenu du grand nombre de médiateurs que nous avons étudié, l'âge à la première grossesse explique une part importante de l'association observée entre la PSE et le risque de cancer du sein. D'autres médiateurs potentiels, y compris le dépistage, ont eu un effet moindre sur l'association. L'association entre la PSE et le risque de cancer du sein n'a pas été entièrement expliquée, ce qui suggère que des recherches supplémentaires sont nécessaires pour identifier d'autres médiateurs et mécanismes potentiels.

**TRAVAIL 4 : INCORPORATION BIOLOGIQUE DE L'ENVIRONNEMENT SOCIAL ET
RISQUE DE CANCER DU SEIN**

Le premier travail de ce chapitre nous a permis de mettre en évidence la période de vie et l'indicateur de PSE le plus fortement associé au risque de cancer du sein : le niveau d'éducation. Au vu des résultats obtenus, il semblerait qu'une partie des inégalités observées persiste malgré la prise en compte d'un large panel de facteurs de risque. Dans cette seconde section, nous allons nous intéresser au mécanisme direct reliant l'environnement social au risque de cancer du sein, à travers l'étude de l'incorporation biologique de l'éducation dans la cohorte E3N.

L'article relatif à cette section est en cours de finalisation (version envoyée aux co-auteurs) « *Biological embodiment of education and future risk of breast cancer : findings from a French prospective cohort* ».

1. CONTEXTE

Au-delà du rôle des facteurs médiateurs identifiés, le lien entre l'éducation et le risque de cancer du sein peut être en partie dû à des modifications biologiques associées à l'éducation qui restent à élucider et à mieux comprendre. Le développement du cancer du sein est depuis longtemps associé à une dérégulation des biomarqueurs inflammatoires^{43,44}, métaboliques^{131,132} et cardiovasculaires^{133,134} ainsi qu'à un profil particulier d'hormones sexuelles¹³⁵⁻¹⁴² comme nous en avons discuté dans la partie 1, chapitre III. Les résultats d'études épidémiologiques individuelles soutiennent une association de ces biomarqueurs avec le risque de cancer du sein, mais les résultats de méta-analyses ont fait état d'une incohérence et d'une hétérogénéité, qui pourraient s'expliquer par divers facteurs, notamment la taille de l'échantillon et les différences entre les variables intermédiaires considérées. De plus, la plupart de ces études se sont principalement concentrées sur des biomarqueurs représentant un seul système biologique et très peu ont pris en compte des biomarqueurs de différents systèmes physiologiques dans la même population. Des études supplémentaires sont donc nécessaires pour mieux caractériser la relation entre ces biomarqueurs et le risque de cancer du sein.

Certaines études suggèrent que l'environnement social, en particulier le niveau d'éducation, peut susciter une réponse biologique qui peut être indépendante des comportements de santé et de l'IMC³³⁻⁴⁰, comme je l'ai montré et abordé dans la partie 1, chapitre I, section 5. Certaines composantes de cette réponse biologique peuvent à leurs tours avoir un impact sur le risque de cancer du sein. Mais à notre connaissance, aucune étude n'a regardé si une partie du risque de cancer du sein à long terme associé à l'éducation peut être attribuée à l'incorporation biologique de la PSE reflétée par la variation des biomarqueurs - qui ont eux-mêmes été liés de manière prospective au risque de cancer du sein.

Objectifs de l'étude : Ajouter de nouvelles preuves sur la relation entre certains biomarqueurs et le risque de cancer du sein, et étudier l'effet de l'éducation sur le niveau de ces biomarqueurs, conditionnellement aux comportements de santé et aux facteurs reproductifs.

Hypothèses : certains biomarqueurs associés au risque de cancer du sein sont également socialement stratifiés. L'éducation influence le niveau de ces biomarqueurs indépendamment des comportements de santé, des mesures anthropométriques et des facteurs reproductifs, ce qui pourrait constituer un mécanisme nouveau par lequel l'éducation pourrait influencer la survenue du cancer du sein.

2. METHODE ET SCHEMA D'ETUDE

J'ai mené des analyses multivariées (dont le détail se trouve dans la partie 2, chapitre VII) en deux étapes :

- dans un premier temps, des modèles de Cox à risques proportionnels ont été utilisés pour étudier la relation entre chaque biomarqueur et le risque de cancer du sein. Les associations globales entre chaque biomarqueur et le risque de cancer du sein ont été ajustées sur l'âge (modèle 1). Les modèles ont ensuite été ajustés sur toutes les covariables (modèle 2).
- ensuite, des modèles de régressions logistiques multinomiales ont été utilisés pour étudier l'association entre l'éducation et chacun des biomarqueurs précédemment identifiés comme étant associés au risque de cancer du sein. Les associations globales entre l'éducation et chaque biomarqueur ont été ajustées sur l'âge (modèle 1). À partir

du modèle 1, j'ai fait des modèles ajustés d'une part sur tous les comportements de santé et les mesures anthropométriques (modèle 1a), et d'autre part sur tous les facteurs reproductifs (modèle 1b). Les modèles ont ensuite été ajustés sur l'ensemble des covariables (modèle 2).

Ces analyses ont été réalisées dans la cohorte E3N, sur données imputées (N = 2 603) et une analyse de sensibilité a été conduite sur les femmes n'ayant aucune donnée manquante pour tester la robustesse de nos résultats (N = 1 835). En parallèle j'ai également regardé l'impact du statut ménopausique au moment du prélèvement biologique en faisant à nouveau des analyses sur les données imputées des femmes en pré-ménopause uniquement (N = 634) et sur la population sans distinction de statut ménopausique (N = 3 251). Le flowchart représentant la sélection de la population et les différentes variables utilisées est disponible en figure 7.

3. RESULTATS

A. CARACTERISTIQUES DE LA POPULATION D'ETUDE

La description des biomarqueurs et des caractéristiques comportementales et sociodémographiques de la sous-population de femmes de E3N, séparée par niveau d'éducation et par statut vis-à-vis du cancer du sein, est fournie dans le tableau 9 et dans l'annexe 7.1 respectivement.

Les femmes avec un niveau d'éducation plus élevé étaient plus jeunes, avaient plus de risque de développer un cancer du sein, étaient plus nombreuses à fumer, faisaient moins d'activités physiques, étaient plus grandes et plus minces et avaient une faible adhérence à un régime alimentaire occidental. Elles étaient également plus susceptibles d'avoir déjà utilisé un traitement hormonal de la ménopause et un contraceptif oral, d'avoir été plus âgées lors du premier accouchement et d'avoir une plus longue durée de vie reproductive (annexe 7.1.A). Les femmes qui ont développé un cancer du sein étaient plus susceptibles d'avoir une consommation élevée d'alcool et d'être plus grandes (annexe 7.1.B).

Les femmes plus éduquées avaient un niveau plus faible de CRP, TG, PAS, PAD ainsi que de testostérone et un niveau plus élevé de HDL, d'œstradiol et de SHBG (tableau 9).

Par rapport aux femmes qui sont restées indemnes de cancer, les femmes ayant développé un cancer du sein étaient plus susceptibles d'avoir un niveau plus élevé de CRP, HDL, FC, œstradiol et testostérone et un niveau plus faible de LDL, cholestérol et SHBG (tableau 9).

B. ASSOCIATION ENTRE CHAQUE BIOMARQUEUR ET LE RISQUE DE CANCER DU SEIN

Les résultats des modèles de Cox multivariés obtenus à partir des données imputées sont résumés dans le tableau 10. Des différences notables dans les associations avec le risque de cancer du sein ont été observées pour huit des onze biomarqueurs étudiés.

Des concentrations plus élevées de CRP, HDL, FC, œstradiol et testostérone sont associées à un risque plus élevé de cancer du sein en comparant le tercile le plus élevé au tercile le plus faible (CRP : $HR_{\text{tercile 3}} = 2,24 [1,74 ; 2,88]$, HDL : $HR_{\text{tercile 3}} = 1,72 [1,40 ; 2,11]$, FC : $HR_{\text{tercile 3}} = 1,24 [1,00 ; 1,53]$, œstradiol : $HR_{\text{tercile 3}} = 1,43 [1,14 ; 1,80]$ et testostérone : $HR_{\text{tercile 3}} = 2,59 [2,10 ; 3,19]$, Modèle 1, Tableau 10). En revanche, les concentrations de LDL, de cholestérol et de SHBG étaient inversement associées au risque de cancer du sein (LDL : $HR_{\text{tercile 3}} = 0,28 [0,22 ; 0,36]$, cholestérol : $HR_{\text{tercile 3}} = 0,57 [0,46 ; 0,71]$, et SHBG : $HR_{\text{tercile 3}} = 0,81 [0,66 ; 1,00]$, Modèle 1, tableau 10). Un schéma similaire a été observé en comparant le tercile 2 au tercile 1, bien que les associations aient été affaiblies pour le HDL, la FC, le cholestérol et le SHBG.

Ces associations sont restées remarquablement stables après l'ajustement sur les comportements de santé, les mesures anthropométriques et les facteurs reproductifs (Modèle 2, tableau 10).

FIGURE 7 : SELECTION DE LA POPULATION AVEC UNE MESURE DES BIOMARQUEURS ET VARIABLES UTILISEES POUR CETTE ETUDE

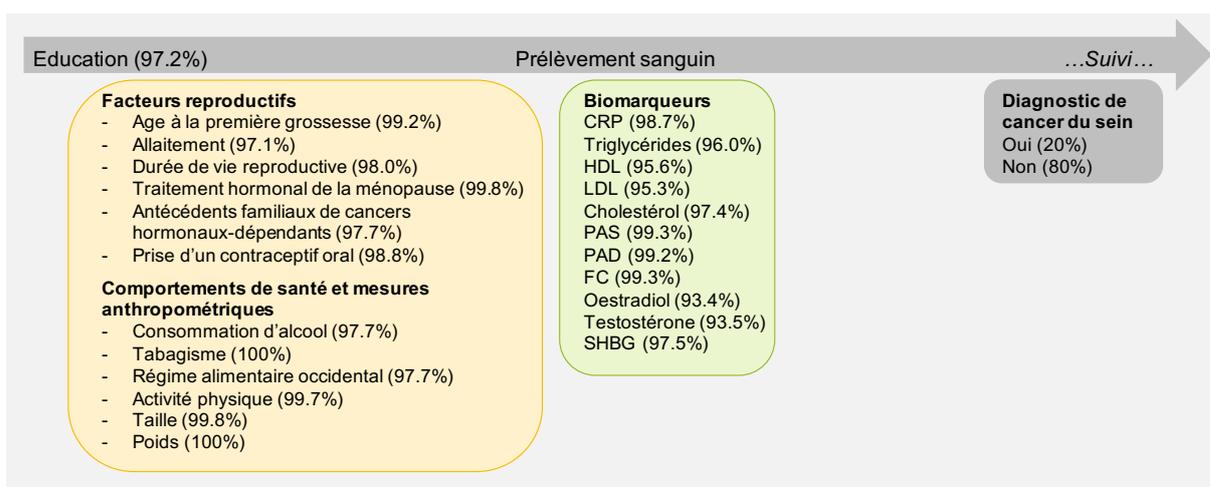
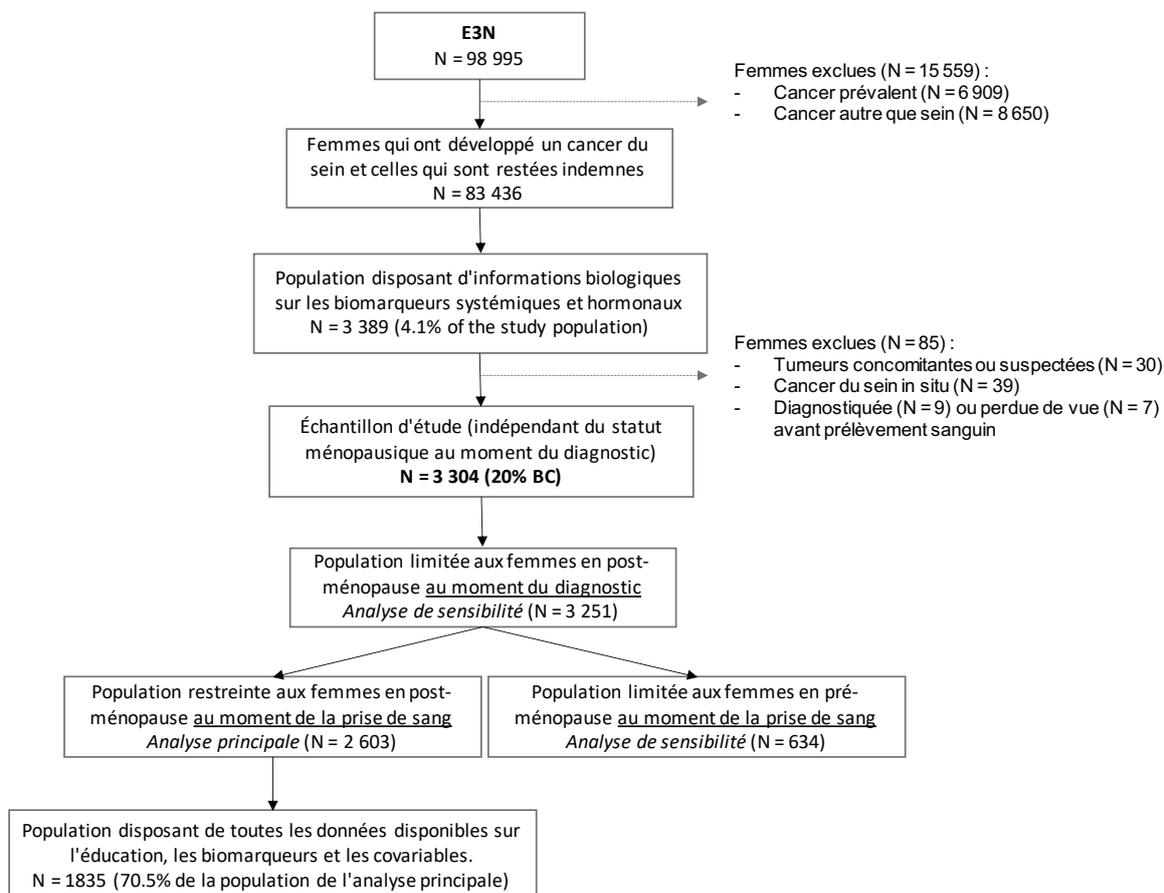


TABLEAU 9 : CARACTERISTIQUES BIOLOGIQUES DES FEMMES AVEC DES DONNEES DISPONIBLES DANS LA COHORTE E3N (N = 2603) PAR NIVEAU D'EDUCATION ET STATUT VIS-A-VIS DU CANCER DU SEIN

Variables	Description	(A) Education				p	(B) Cancer du sein		
		Moins éduquées	Moyennement éduquées	Plus éduquées			Non	Oui	p
CRP (mg/L)	n= 2603	n= 318	n= 1229	n= 984	0,003	n= 2084	n= 519	<0,001	
[0,0,91], n(%)	883 (33.92)	83 (26.1)	438 (35.64)	332 (33.74)		792 (38)	91 (17.53)		
(0.91,1.4], n(%)	835 (32.08)	99 (31.13)	385 (31.33)	333 (33.84)		594 (28.5)	241 (46.44)		
(1.4,56], n(%)	885 (34)	136 (42.77)	406 (33.03)	319 (32.42)		698 (33.49)	187 (36.03)		
Triglycérides (mmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,022	n= 2084	n= 519	0,800	
[0.06,0.77], n(%)	894 (34.34)	86 (27.04)	436 (35.48)	352 (35.77)		712 (34.17)	182 (35.07)		
(0.77,1.09], n(%)	829 (31.85)	102 (32.08)	382 (31.08)	314 (31.91)		661 (31.72)	168 (32.37)		
(1.09,5.53], n(%)	880 (33.81)	130 (40.88)	411 (33.44)	318 (32.32)		711 (34.12)	169 (32.56)		
HDL (mmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,009	n= 2084	n= 519	<0,001	
[0.46,1.61], n(%)	852 (32.73)	130 (40.88)	381 (31)	314 (31.91)		711 (34.12)	141 (27.17)		
(1.61,2.1], n(%)	806 (30.96)	77 (24.21)	396 (32.22)	310 (31.5)		681 (32.68)	125 (24.08)		
(2.1,5.11], n(%)	945 (36.3)	111 (34.91)	452 (36.78)	360 (36.59)		692 (33.21)	253 (48.75)		
LDL (mmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,253	n= 2084	n= 519	<0,001	
[0.36,2.53], n(%)	901 (34.61)	114 (35.85)	434 (35.31)	336 (34.15)		621 (29.8)	280 (53.95)		
(2.53,3.47], n(%)	810 (31.12)	98 (30.82)	355 (28.89)	327 (33.23)		662 (31.77)	148 (28.52)		
(3.47,8.3], n(%)	892 (34.27)	106 (33.33)	440 (35.8)	321 (32.62)		801 (38.44)	91 (17.53)		
Cholestérol (mmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,612	n= 2084	n= 519	<0,001	
[0.01,5], n(%)	835 (32.08)	103 (32.39)	394 (32.06)	314 (31.91)		633 (30.37)	202 (38.92)		
(5,5.81], n(%)	855 (32.85)	95 (29.87)	398 (32.38)	337 (34.25)		674 (32.34)	181 (34.87)		
(5.81,11.9], n(%)	913 (35.07)	120 (37.74)	437 (35.56)	333 (33.84)		777 (37.28)	136 (26.2)		
Fréquence cardiaque (bpm)	n= 2603	n= 318	n= 1229	n= 984	0,603	n= 2084	n= 519	0,200	
[37,65], n(%)	903 (34.69)	103 (32.39)	419 (34.09)	353 (35.87)		740 (35.51)	163 (31.41)		
(65,73], n(%)	878 (33.73)	106 (33.33)	427 (34.74)	318 (32.32)		701 (33.64)	177 (34.1)		
(73,138], n(%)	822 (31.58)	109 (34.28)	383 (31.16)	313 (31.81)		643 (30.85)	179 (34.49)		
PAS (mmHg)	n= 2603	n= 318	n= 1229	n= 984	<0,001	n= 2084	n= 519	0,700	
[81.5,116], n(%)	812 (31.19)	70 (22.01)	382 (31.08)	338 (34.35)		656 (31.48)	156 (30.06)		
(116,132], n(%)	849 (32.62)	102 (32.08)	384 (31.24)	341 (34.65)		677 (32.49)	172 (33.14)		
(132,220], n(%)	942 (36.19)	146 (45.91)	463 (37.67)	305 (31)		751 (36.04)	191 (36.8)		
PAD (mmHg)	n= 2603	n= 318	n= 1229	n= 984	0,047	n= 2084	n= 519	0,900	
[42.5,72], n(%)	884 (33.96)	96 (30.19)	413 (33.6)	352 (35.77)		707 (33.93)	177 (34.1)		
(72,80.5], n(%)	843 (32.39)	98 (30.82)	392 (31.9)	334 (33.94)		671 (32.2)	172 (33.14)		
(80.5,135], n(%)	876 (33.65)	124 (38.99)	424 (34.5)	298 (30.28)		706 (33.88)	170 (32.76)		
Oestradiol (pmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,006	n= 2084	n= 519	<0,001	
[18.4,69.1], n(%)	1065 (40.91)	145 (45.6)	531 (43.21)	365 (37.09)		895 (42.95)	170 (32.76)		
(69.1,238], n(%)	849 (32.62)	105 (33.02)	391 (31.81)	330 (33.54)		647 (31.05)	202 (38.92)		
(238,4970], n(%)	689 (26.47)	68 (21.38)	307 (24.98)	289 (29.37)		542 (26.01)	147 (28.32)		
Testostérone (nmol/L)	n= 2603	n= 318	n= 1229	n= 984	0,205	n= 2084	n= 519	<0,001	
[0.069,0.46], n(%)	1044 (40.11)	110 (34.59)	514 (41.82)	395 (40.14)		903 (43.33)	141 (27.17)		
(0.46,0.898], n(%)	792 (30.43)	103 (32.39)	358 (29.13)	304 (30.89)		649 (31.14)	143 (27.55)		
(0.898,47.7], n(%)	767 (29.47)	105 (33.02)	357 (29.05)	285 (28.96)		532 (25.53)	235 (45.28)		
SHBG (nmol/L)	n= 2603	n= 318	n= 1229	n= 984	<0,001	n= 2084	n= 519	0,100	
[0.35,49.5], n(%)	867 (33.31)	131 (41.19)	428 (34.83)	289 (29.37)		674 (32.34)	193 (37.19)		
(49.5,75.5], n(%)	824 (31.66)	102 (32.08)	374 (30.43)	322 (32.72)		665 (31.91)	159 (30.64)		
(75.5,305], n(%)	912 (35.04)	85 (26.73)	427 (34.74)	373 (37.91)		745 (35.75)	167 (32.18)		

TABLEAU 10 : ASSOCIATION ENTRE CHAQUE BIOMARQUEUR AVEC LE RISQUE DE CANCER DU SEIN DIAGNOSTIQUE EN POST-MENOPAUSE A PARTIR DE LA POPULATION DE E3N AVEC DONNEES IMPUTEES (N = 2 603)

		Modèle 1	Modèle 2
		HR(95%CI)	HR(95%CI)
CRP (mg/L)	Tercile 2	3,20 [2,51 ; 4,07]	3,24 [2,54 ; 4,13]
	Tercile 3	2,24 [1,74 ; 2,88]	2,46 [1,89 ; 3,19]
Triglycérides (mmol/L)	Tercile 2	1,00 [0,81 ; 1,23]	0,99 [0,80 ; 1,23]
	Tercile 3	0,93 [0,75 ; 1,15]	0,94 [0,75 ; 1,17]
HDL (mmol/L)	Tercile 2	0,92 [0,72 ; 1,17]	0,91 [0,71 ; 1,16]
	Tercile 3	1,72 [1,40 ; 2,11]	1,69 [1,37 ; 2,09]
LDL (mmol/L)	Tercile 2	0,54 [0,44 ; 0,66]	0,54 [0,44 ; 0,66]
	Tercile 3	0,28 [0,22 ; 0,36]	0,29 [0,23 ; 0,37]
Cholestérol (mmol/L)	Tercile 2	0,84 [0,69 ; 1,03]	0,87 [0,71 ; 1,06]
	Tercile 3	0,57 [0,46 ; 0,71]	0,59 [0,47 ; 0,73]
Fréquence cardiaque (bpm)	Tercile 2	1,13 [0,91 ; 1,40]	1,13 [0,91 ; 1,40]
	Tercile 3	1,24 [1,00 ; 1,53]	1,25 [1,01 ; 1,55]
PAS (mmHg)	Tercile 2	1,07 [0,86 ; 1,33]	1,08 [0,86 ; 1,34]
	Tercile 3	1,07 [0,86 ; 1,32]	1,10 [0,88 ; 1,38]
PAD (mmHg)	Tercile 2	1,01 [0,82 ; 1,25]	1,03 [0,83 ; 1,27]
	Tercile 3	0,97 [0,78 ; 1,19]	0,99 [0,80 ; 1,22]
Oestradiol (pmol/L)	Tercile 2	1,61 [1,31 ; 1,98]	1,60 [1,28 ; 1,99]
	Tercile 3	1,43 [1,14 ; 1,80]	1,43 [1,12 ; 1,82]
Testostérone (nmol/L)	Tercile 2	1,40 [1,11 ; 1,76]	1,37 [1,08 ; 1,74]
	Tercile 3	2,59 [2,10 ; 3,19]	2,57 [2,09 ; 3,18]
SHBG (nmol/L)	Tercile 2	0,85 [0,69 ; 1,04]	0,81 [0,65 ; 1,01]
	Tercile 3	0,81 [0,66 ; 1,00]	0,73 [0,58 ; 0,92]

Tercile 1 (ref) CRP : [0 ; 0,91]; TG : [0,06 ; 0,77]; HDL : [0,46 ; 1,61]; LDL : [0,36 ; 2,53]; Cholestérol : [0,01 ; 5]; FC : [37 ; 65]; PAS : [81,5 ; 116]; PAD : [42,5 ; 72]; Oestradiol : [18,4 ; 69,1]; Testostérone : [0,069 ; 0,46]; SHBG : [0,35 ; 49,5]

Tercile 2 CRP : (0,91 ; 1,40); TG : (0,77 ; 1,09); HDL : (1,61 ; 2,10); LDL : (2,53 ; 3,47); Cholestérol : (5,00 ; 5,81); FC : (65 ; 73); PAS : (116 ; 132); PAD : (72 ; 80,5); Oestradiol : (69,1 ; 238); Testostérone : (0,46 ; 0,898); SHBG : (49,5 ; 75,5)

Tercile 3 CRP : (1,40 ; 5,6); TG : (1,09 ; 5,53); HDL : (2,10 ; 5,11); LDL : (3,47 ; 8,30); Cholestérol : (5,81 ; 11,9); FC : (73 ; 138); PAS : (132 ; 220); PAD : (80,5 ; 135); Oestradiol : (238 ; 4970); Testostérone : (0,898 ; 47,7); SHBG : (75,5 ; 305)

Modèle 1 : biomarqueurs + âge

Modèle 2 : **Modèle 1** + alcool, tabac, régime occidental, activité physique, poids, taille, antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement.

C. ASSOCIATION ENTRE L'EDUCATION ET CHAQUE BIOMARQUEUR

Les résultats des analyses des régressions logistiques multinomiales sont présentés dans le tableau 11 pour chaque biomarqueur selon le niveau d'éducation.

Sur les huit biomarqueurs associés au risque de cancer du sein, cinq étaient également liés au niveau d'éducation (CRP, HDL, œstradiol, testostérone et SHBG). Les femmes plus éduquées étaient plus susceptibles d'avoir un faible niveau de CRP par rapport aux femmes moins éduquées ($OR_{\text{tercile 3, moyennement éduquées}} = 0,57 [0,42 ; 0,78]$ et $OR_{\text{tercile 3, plus éduquées}} = 0,59 [0,43 ; 0,81]$ modèle 1, tableau 11B, figure 8). La même tendance était observée pour la testostérone ($OR_{\text{tercile 3, moyennement éduquées}} = 0,74 [0,55 ; 1,00]$ et $OR_{\text{tercile 3, plus éduquées}} = 0,76 [0,56 ; 1,03]$, modèle 1, tableau 11B, figure 8). En revanche, les femmes plus éduquées étaient plus susceptibles d'avoir un niveau plus élevé de HDL ($OR_{\text{tercile 3, moyennement éduquées}} = 1,38 [1,04 ; 1,84]$ et $OR_{\text{tercile 3, plus éduquées}} = 1,34 [1,00 ; 1,80]$, modèle 1, tableau 11B, figure 8), d'œstradiol ($OR_{\text{tercile 3, moyennement éduquées}} = 1,20 [0,87 ; 1,67]$ et $OR_{\text{tercile 3, plus éduquées}} = 1,55 [1,11 ; 2,16]$, modèle 1, tableau 11B) et de SHBG ($OR_{\text{tercile 3, moyennement éduquées}} = 1,51 [1,11 ; 2,04]$ et $OR_{\text{tercile 3, plus éduquées}} = 1,93 [1,41 ; 2,63]$, modèle 1, tableau 11B, figure 8).

Les associations avec la CRP, le HDL, l'œstradiol et la SHBG ont été affaiblies après l'ajustement sur les covariables, principalement sur les comportements de santé et les mesures anthropométriques (modèle 1a, annexe 7.2). Cependant, une association persiste dans le modèle totalement ajusté pour la CRP ($OR_{\text{tercile 3, moyennement éduquées}} = 0,70 [0,50 ; 0,97]$, $OR_{\text{tercile 3, plus éduquées}} = 0,83 [0,59 ; 1,18]$, modèle 2, tableau 11B, figure 8) et, dans une moindre mesure, pour la SHBG ($OR_{\text{tercile 3, moyennement éduquées}} = 1,21 [0,87 ; 1,69]$ et $OR_{\text{tercile 3, plus éduquées}} = 1,37 [0,96 ; 1,95]$, modèle 2, tableau 11B, figure 8).

L'association avec la testostérone n'était que marginalement affectée par l'ajustement sur les comportements de santé et les mesures anthropométriques, les facteurs reproductifs, ou les deux dans le modèle totalement ajusté ($OR_{\text{tercile 3, moyennement éduquées}} = 0,75 [0,55 ; 1,02]$ et $OR_{\text{tercile 3, plus éduquées}} = 0,76 [0,55 ; 1,05]$, modèle 2, tableau 11B, figure 8).

Un schéma similaire a été observé lors de la comparaison entre le tercile 2 et le tercile 1, bien que les associations aient été affaiblies pour la plupart des biomarqueurs, à l'exception du HDL pour lequel il y avait une association plus forte qui a persistée dans le modèle totalement ajusté (modèle 2, tableau 11A, figure 8).

TABLEAU 11 : ANALYSES DE REGRESSIONS LOGISTIQUES MULTINOMIALES DE L'ASSOCIATION ENTRE L'ÉDUCATION ET CHAQUE BIOMARQUEUR PRÉCÉDEMMENT IDENTIFIÉ COMME ÉTANT ASSOCIÉ AU RISQUE DE CANCER DU SEIN À PARTIR DE LA POPULATION AVEC DONNÉES IMPUTÉES DANS E3N (N = 2 603)

Éducation		(A) Biomarqueurs Tercile 2		(B) Biomarqueurs Tercile 3	
		Modèle 1	Modèle 2	Modèle 1	Modèle 2
		OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)
CRP (mg/L)	Moyennement éduquées	0,74 [0,54 ; 1,03]	0,82 [0,59 ; 1,14]	0,57 [0,42 ; 0,78]	0,70 [0,50 ; 0,97]
	Plus éduquées	0,85 [0,61 ; 1,18]	0,98 [0,69 ; 1,39]	0,59 [0,43 ; 0,81]	0,83 [0,59 ; 1,18]
HDL (mmol/L)	Moyennement éduquées	1,70 [1,25 ; 2,33]	1,60 [1,16 ; 2,22]	1,38 [1,04 ; 1,84]	1,22 [0,90 ; 1,66]
	Plus éduquées	1,63 [1,18 ; 2,25]	1,61 [1,14 ; 2,28]	1,34 [1,00 ; 1,80]	1,16 [0,84 ; 1,60]
LDL (mmol/L)	Moyennement éduquées	0,95 [0,70 ; 1,28]	0,98 [0,72 ; 1,34]	1,08 [0,81 ; 1,46]	1,16 [0,85 ; 1,59]
	Plus éduquées	1,12 [0,82 ; 1,53]	1,21 [0,87 ; 1,68]	1,04 [0,76 ; 1,41]	1,22 [0,88 ; 1,70]
Cholestérol I (mmol/L)	Moyennement éduquées	1,10 [0,81 ; 1,50]	1,10 [0,80 ; 1,51]	0,96 [0,72 ; 1,29]	1,03 [0,76 ; 1,40]
	Plus éduquées	1,18 [0,86 ; 1,62]	1,16 [0,83 ; 1,63]	0,94 [0,69 ; 1,27]	1,09 [0,78 ; 1,52]
FC (bpm)	Moyennement éduquées	0,97 [0,72 ; 1,32]	0,97 [0,71 ; 1,33]	0,86 [0,63 ; 1,16]	0,84 [0,61 ; 1,15]
	Plus éduquées	0,85 [0,62 ; 1,16]	0,86 [0,62 ; 1,20]	0,82 [0,60 ; 1,12]	0,80 [0,57 ; 1,12]
Oestradiol (pmol/L)	Moyennement éduquées	1,01 [0,76 ; 1,35]	0,88 [0,65 ; 1,20]	1,20 [0,87 ; 1,67]	1,03 [0,72 ; 1,47]
	Plus éduquées	1,20 [0,89 ; 1,61]	0,98 [0,70 ; 1,36]	1,55 [1,11 ; 2,16]	1,29 [0,88 ; 1,88]
Testostérone (nmol/L)	Moyennement éduquées	0,75 [0,56 ; 1,02]	0,81 [0,59 ; 1,10]	0,74 [0,55 ; 1,00]	0,75 [0,55 ; 1,02]
	Plus éduquées	0,82 [0,60 ; 1,11]	0,91 [0,65 ; 1,26]	0,76 [0,56 ; 1,03]	0,76 [0,55 ; 1,05]
SHBG (nmol/L)	Moyennement éduquées	1,14 [0,85 ; 1,53]	0,96 [0,70 ; 1,32]	1,51 [1,11 ; 2,04]	1,21 [0,87 ; 1,69]
	Plus éduquées	1,47 [1,08 ; 1,99]	1,19 [0,85 ; 1,66]	1,93 [1,41 ; 2,63]	1,37 [0,96 ; 1,95]

Modèle 1 : éducation + âge

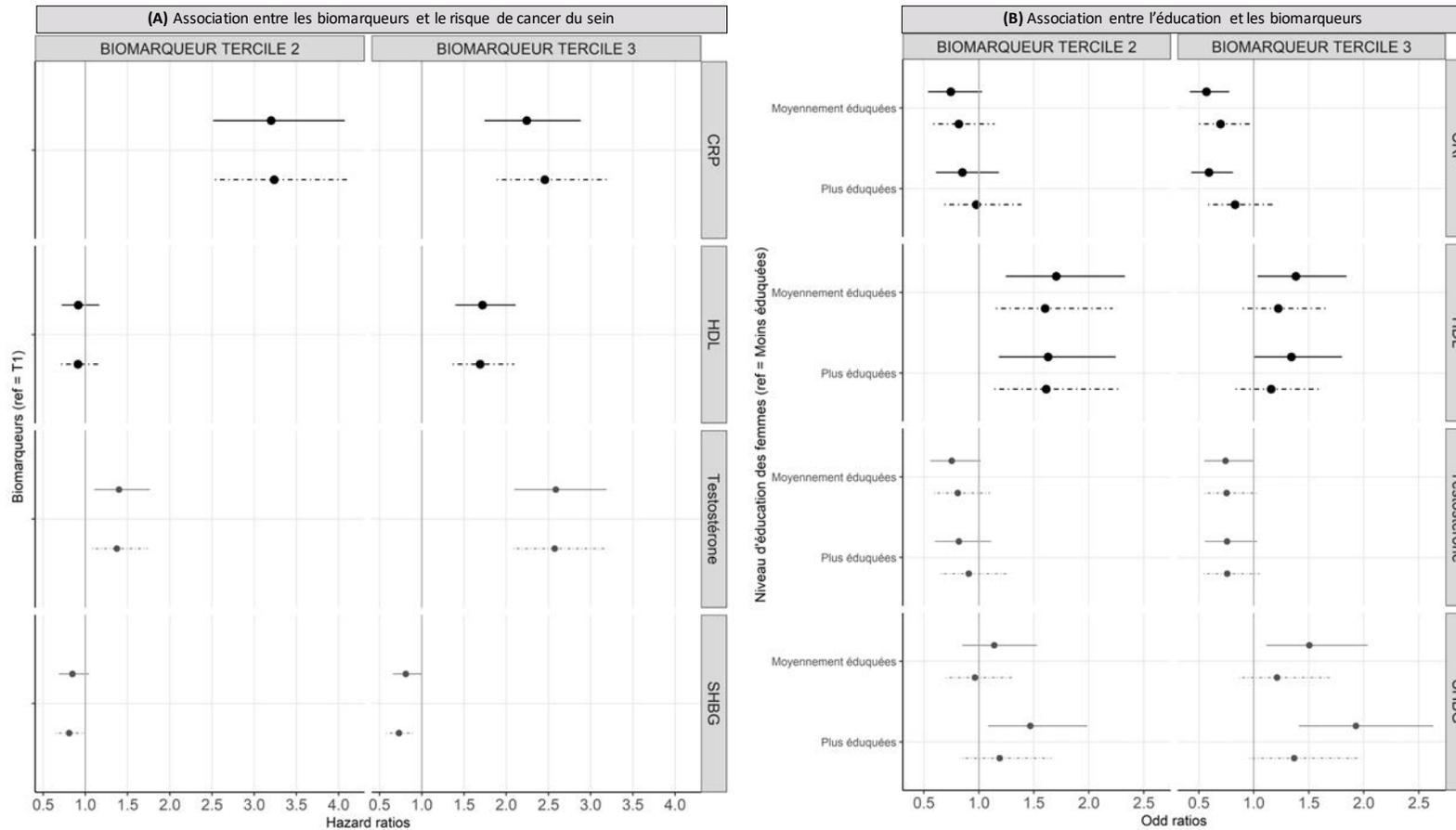
Modèle 2 : **Modèle 1** + alcool, tabac, régime occidental, activité physique, poids, taille, antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement.

Tercile 1 (ref) CRP : [0 ; 0,91]; HDL : [0,46 ; 1,61]; LDL : [0,36 ; 2,53]; Cholestérol : [0,01 ; 5]; FC : [37 ; 65]; Oestradiol : [18,4 ; 69,1]; Testostérone : [0,069 ; 0,46]; SHBG : [0,35 ; 49,5]

Tercile 2 CRP : (0,91 ; 1,40); HDL : (1,61 ; 2,10); LDL : (2,53 ; 3,47); Cholestérol : (5,00 ; 5,81); FC : (65 ; 73); Oestradiol : (69,1 ; 238); Testostérone : (0,46 ; 0,898); SHBG : (49,5 ; 75,5)

Tercile 3 CRP : (1,40 ; 56); HDL : (2,10 ; 5,11); LDL : (3,47 ; 8,30); Cholestérol : (5,81 ; 11,9); FC : (73 ; 138); Oestradiol : (238 ; 4970); Testostérone : (0,898 ; 47,7); SHBG : (75,5 ; 305)

FIGURE 8 - RESULTATS PRESENTANT LES BIOMARQUEURS ASSOCIES DE MANIERE INDEPENDANTE A LA FOIS AU CANCER DU SEIN ET A L'EDUCATION. (A) GRAPHIQUE REPRESENTANT LES RESULTATS DES ANALYSES DE REGRESSIONS DES MODELES DE COX A RISQUES PROPORTIONNELS (HR [INTERVALLE DE CONFIANCE A 95 %]) POUR L'ASSOCIATION ENTRE LES BIOMARQUEURS ET LES CANCERS DU SEIN EN POST-MENOPAUSE ET (B) GRAPHIQUE REPRESENTANT LES RESULTATS DES ANALYSES DE REGRESSIONS MULTINOMIALES (OR [INTERVALLE DE CONFIANCE A 95 %]) POUR L'ASSOCIATION ENTRE L'EDUCATION ET LES MARQUEURS BIOLOGIQUES. LES TRAITIS PLEINS REPRESENTENT LES RESULTATS DES MODELES M1 ET LES TRAITIS POINTILLES REPRESENTENT LES MODELES TOTALEMENT AJUSTES.



D. ANALYSES DE SENSIBILITE ET ANALYSES EN SOUS-GROUPE

Les caractéristiques de la population en cas complets sont fournies en annexe 7.3. Les analyses réalisées sur les données en cas complets ont donné des résultats similaires, bien que les associations entre chaque biomarqueur et le risque de cancer du sein aient été légèrement atténuées, principalement pour la FC et la SHBG (annexe 7.4). De plus, nous observons les mêmes associations entre l'éducation et chaque biomarqueur bien que certains effets semblent être renforcés, notamment pour la testostérone pour laquelle il y avait une association inverse significative non affectée par les différents ajustements (annexe 7.5).

Lorsque la population était limitée aux femmes en pré-ménopause au moment de la collecte de sang, ou lorsque nous avons défini une population sans tenir compte du statut ménopausique, les résultats de l'association entre les biomarqueurs et le risque de cancer du sein étaient cohérents dans les différentes populations (annexe 7.6). Cependant, certaines discordances concernant la direction de l'effet ont été observées pour l'association entre l'éducation et les biomarqueurs. Contrairement aux femmes en post-ménopause, les femmes plus éduquées en pré-ménopause lors de la collecte de sang avaient un niveau de CRP plus élevé par rapport aux femmes moins éduquées sans que cette association soit statistiquement significative (CRP: $OR_{\text{tercile 3, plus éduquées}} = 1,63 [0,72 ; 3,70]$, modèle 2, annexe 7.7). De même, les femmes plus éduquées en pré-ménopause avaient un niveau d'œstradiol moins élevé, sans que cette association soit statistiquement significative (œstradiol : $OR_{\text{tercile 3, plus éduquées}} = 0,58 [0,25 ; 1,30]$, modèle 2, annexe 7.7).

4. DISCUSSION

A partir d'un sous-échantillon de la cohorte prospective française E3N, les résultats montrent que des niveaux plus élevés de CRP, HDL, FC, œstradiol et testostérone étaient associés à un risque plus élevé de cancer du sein alors que des niveaux plus élevés de LDL, cholestérol et SHBG étaient inversement liés au développement de la maladie. Aucun des facteurs de confusion potentiels examinés dans cette étude n'a modifié les associations. Parmi ces biomarqueurs, cinq se sont également avérés être associés au niveau d'éducation : les femmes plus éduquées avaient des niveaux plus faibles de CRP et de testostérone mais des niveaux

plus élevés de HDL, d'œstradiol et de SHBG. À l'exception de la testostérone, toutes les associations entre l'éducation et les biomarqueurs ont été atténuées après l'ajustement sur les comportements de santé et les mesures anthropométriques, ainsi que sur les facteurs reproductifs. Cependant, l'association persistait malgré l'ajustement sur l'ensemble des facteurs pour la CRP et le HDL.

A. BIOMARQUEURS ET RISQUE DE CANCER DU SEIN

Notre étude s'ajoute à la littérature sur l'effet des marqueurs biologiques sur le développement du cancer du sein. L'association positive entre la CRP et le risque de cancer du sein est conforme à la littérature précédente. L'inflammation est l'une des caractéristiques reconnues du cancer¹⁹⁷ et des recherches antérieures ont fourni des preuves solides que l'inflammation peut augmenter le risque de cancer et favoriser la progression de la tumeur¹⁹⁸.

Nos résultats montrant un risque plus élevé de cancer du sein chez les femmes ayant un taux élevé d'œstradiol et de testostérone ou un faible taux de SHBG ont également été précédemment suspectés. Les œstrogènes contribuent à stimuler la croissance et la division des cellules du sein¹⁴¹ et pourraient accroître la prolifération cellulaire pour favoriser la progression du cancer. L'impact de la testostérone est moins clair, mais il a été précédemment suspecté que les précurseurs de stéroïdes pourraient être convertis en œstradiol par une enzyme, l'aromatase. En revanche, la SHBG pourrait jouer un rôle protecteur contre le développement du cancer du sein en régulant la disponibilité de la fraction libre des hormones sexuelles, identifiée comme étant la partie active des hormones.

Toutefois, nos résultats sur les marqueurs métaboliques ne sont pas en accord avec les résultats des méta-analyses précédentes^{131,132}. Bien que nous observions une association inverse entre le niveau de cholestérol total et le risque de cancer du sein, comme cela a été précédemment rapporté, les résultats des deux principaux composants du cholestérol, le HDL et le LDL, divergent. Ces méta-analyses ont mis en évidence une association inverse entre le HDL et le risque de cancer du sein, ce qui implique que l'association observée pour le cholestérol total serait principalement due au HDL. Cependant, nos résultats pour le cholestérol total semblent être les mêmes que pour le LDL et, nous avons mis en évidence une association positive entre le HDL et le développement de cancer du sein comme cela a été montré dans certaines études individuelles précédentes²⁰⁰⁻²⁰². De plus, nous n'avons pas

observé d'association entre les TG et le risque de cancer du sein, contrairement à une méta-analyse précédente¹³² qui a identifié une association inverse limite, bien que les études individuelles n'aient pas non plus trouvé d'associations significatives.

Des études antérieures ont fait état d'une association positive entre l'hypertension et, dans une moindre mesure, la PAS avec le risque de cancer du sein^{133,134}. Nous n'avons observé aucune association avec la PAS et la PAD. Toutefois, nous avons constaté une association positive pour la FC, qui a précédemment été identifiée comme étant corrélée à l'hypertension²⁰³.

Nos analyses ont été ajustées sur un large ensemble de covariables qui pourraient potentiellement agir comme facteurs de confusion sur l'association entre chaque biomarqueur et le risque de cancer du sein. Nos résultats n'ont été que marginalement affectés par ces ajustements, ce qui va à l'encontre des résultats précédents qui montraient une association variable entre certains biomarqueurs et le risque de cancer du sein après avoir pris en compte certains facteurs de confusion, en particulier par niveau d'IMC comme cela a été observé dans un sous-échantillon de E3N¹⁷³. L'excès de poids et l'obésité augmentent de manière significative le risque de cancer du sein et sont des facteurs établis qui contribuent, entre autres, à l'inflammation chronique, à l'hypertension et à l'augmentation de la concentration d'œstrogènes chez les femmes ménopausées par le biais des aromatases disponibles dans le tissu adipeux²⁰⁴. Néanmoins, nos associations entre les biomarqueurs et le risque de cancer du sein n'étaient pas influencées par l'ajustement sur le poids ou tout autre facteur de confusion.

En outre, certaines études ont fait état d'une association incohérente entre les analyses chez les femmes en pré et post-ménopause, ce qui suggère que la ménopause peut modifier la relation entre les biomarqueurs et le risque de cancer du sein. Nos résultats sur une sous-population de femmes pré-ménopausées au moment de la collecte de sang sont cohérents avec ceux des femmes post-ménopausées, bien qu'il y ait une plus grande variabilité probablement due à la taille réduite de l'échantillon dans ce groupe.

B. EDUCATION ET MARQUEURS BIOLOGIQUES

Nos résultats sur l'association entre l'éducation et chaque biomarqueur sont conformes à la littérature. Nos résultats sont en accord avec de précédents travaux sur le lien entre PSE et CRP^{38,205}, et avec mon premier travail⁴⁰ dans lequel nous retrouvons une association inverse entre la PSE et le niveau de CRP. Une association avec le HDL a également été identifiée³⁵ dans les quelques études qui l'utilisent seul, indépendamment d'un score biologique. Nos résultats sur les hormones sexuelles étaient également conformes aux quelques études publiées précédemment, avec une association inverse retrouvée pour la testostérone²⁰⁶⁻²⁰⁸ et une association positive avec la SHBG^{209,210}, bien que la force des associations soit relativement limitée. Contrairement à la littérature précédente, nous avons constaté une association positive entre l'éducation et l'œstradiol.

Les associations avec la CRP, le HDL, l'œstradiol et la SHBG étaient atténuées après l'ajustement sur les différentes covariables, ce qui peut refléter des mécanismes indirects impliqués dans le lien entre l'éducation et ces biomarqueurs. Même après avoir pris en compte un large éventail de médiateurs potentiels, l'association entre l'éducation et la CRP et le HDL persistait. De plus, les différents ajustements ne modifient que légèrement le lien entre l'éducation et la testostérone et ne semblent pas expliquer entièrement le lien entre l'éducation et la SHBG. Cela pourrait souligner que ces marqueurs biologiques capturent d'autres caractéristiques de l'incorporation sociale qui n'ont pas été capturées par les facteurs de risque précédemment établis.

Bien que ces résultats doivent être considérés avec prudence en raison de la petite taille de l'échantillon dans ce groupe, une certaine contradiction semble apparaître entre les associations chez les femmes pré et post-ménopausées. Les femmes plus éduquées en pré-ménopause étaient plus susceptibles d'avoir un niveau de CRP élevé. La plupart des études n'ont pas pris en compte le statut ménopausique ou se sont uniquement concentrées sur les femmes plus âgées. Des études supplémentaires devraient être réalisées pour confirmer cette relation particulière. De plus, les femmes plus éduquées en pré-ménopause semblaient également plus susceptibles d'avoir un niveau d'œstradiol plus faible. Ces résultats sont conformes à ceux d'une étude précédente qui faisait état d'une baisse de la concentration d'œstrogènes avec un niveau d'éducation plus élevé²¹¹. D'autres études sont nécessaires pour mettre en évidence les mécanismes impliqués dans ces différences.

Dernières avancées

L'article sur ce travail a été envoyé aux co-auteurs fin mars. Après relecture, nous avons fait une réunion mi-avril durant laquelle nous avons discuté du design utilisé, avec notamment des chercheurs ayant par le passé collecté et utilisé les données biologiques de la cohorte E3N. Notre approche initiale était d'inclure toutes les femmes ayant des données biologiques pour un ensemble de biomarqueurs pertinents pour étudier l'effet de l'incorporation biologique de l'environnement social sur le risque de cancer du sein et parmi cette sous-population identifier les participantes ayant développé un cancer entre la date de collection de l'échantillon sanguin et la dernière date de suivi, justifiant le choix d'utiliser un modèle de Cox. Compte tenu du cadre dans lequel ont été collectés les données biologiques (i.e. à travers différentes études cas-témoins de taille variable et pathologies diverses), et qui nous a été détaillé lors de cette réunion, il est apparu qu'il serait nécessaire de réaliser un ensemble d'analyses supplémentaires en utilisant des modèles de régression logistique sur la sous-population de l'étude cas-témoins ayant pour pathologie d'intérêt les cancers du sein¹⁷⁵.

Je viens de recevoir les identifiants des paires de cas-témoins et suis donc en train de revoir la stratégie de l'étude dans le but d'utiliser des modèles de régression logistique conditionnelle en utilisant ces nouvelles informations. J'ai néanmoins refait d'ores et déjà les analyses en utilisant une régression logistique classique pour tester le lien entre biomarqueurs et cancer du sein, et les résultats de la régression logistique sont concordants avec ceux des modèles de Cox.

Nos résultats soulignent que certains biomarqueurs associés au risque de cancer du sein sont également distribués socialement, ce qui suggère que ces biomarqueurs peuvent servir de médiateurs entre l'éducation et le risque de développer la maladie. À l'exception de la testostérone, toutes les associations entre l'éducation et les biomarqueurs étaient atténuées après l'ajustement sur les comportements de santé et les mesures anthropométriques, ainsi que sur les facteurs reproductifs, sans être totalement expliquées pour la CRP et le HDL. Ces résultats suggèrent l'existence de mécanismes autres que les facteurs considérés dans ce travail pour expliquer le lien

entre éducation et testostérone, et dans une moindre mesure entre éducation et CRP/HDL. Des travaux de recherche supplémentaires sont nécessaires pour mieux comprendre le lien entre les biomarqueurs socialement stratifiés et l'apparition du cancer du sein, ce qui pourrait permettre d'identifier les femmes présentant un risque accru de cancer du sein.

CHAPITRE II : DETERMINANTS SOCIAUX ET CARACTERISTIQUES DES CANCERS DU SEIN

TRAVAIL 5 : IMPACT DE LA PSE SUR LES FACTEURS PRONOSTICS DES CANCERS DU SEIN

1. CONTEXTE

Le cancer du sein est une maladie hétérogène qui peut se définir par différents facteurs pronostics et prédictifs au moment du diagnostic comme nous l'avons abordé dans la partie 1, chapitre III. Parmi ces facteurs il y a notamment le stade, le grade et différents sous types en fonction de la présence de récepteurs hormonaux (RH) et de la protéine HER2 (permettant notamment de distinguer les cancers triples négatifs n'exprimant ni de récepteurs hormonaux, ni de surexpression de la protéine HER2). Ces différentes caractéristiques et sous types peuvent avoir une influence sur la survie. Il a également été mis en évidence une association entre des marqueurs de PSE avec le stade, le grade et les sous types, majoritairement dans le sens où les femmes moins favorisées auraient de moins bons facteurs pronostics et prédictifs comme nous l'avons vu dans la partie 1, chapitre III. Cependant les résultats des précédents travaux sont parfois contradictoires. Il est possible que la PSE influence l'exposition à plusieurs facteurs de risque, ce qui pourrait modifier la biologie des tumeurs. L'impact de ces facteurs, qui pourrait apporter des pistes sur les mécanismes qui sont en jeu, n'est pas toujours pris en compte dans les précédentes études. Il a également été suggéré que de futurs travaux devraient être menés en utilisant les données de PSE individuelles et non écologiques comme dans certains des travaux existants.

Objectifs de l'étude : étudier l'impact de l'éducation sur l'extension du cancer au moment du diagnostic ainsi que son agressivité et la distribution des différents sous-types moléculaires.

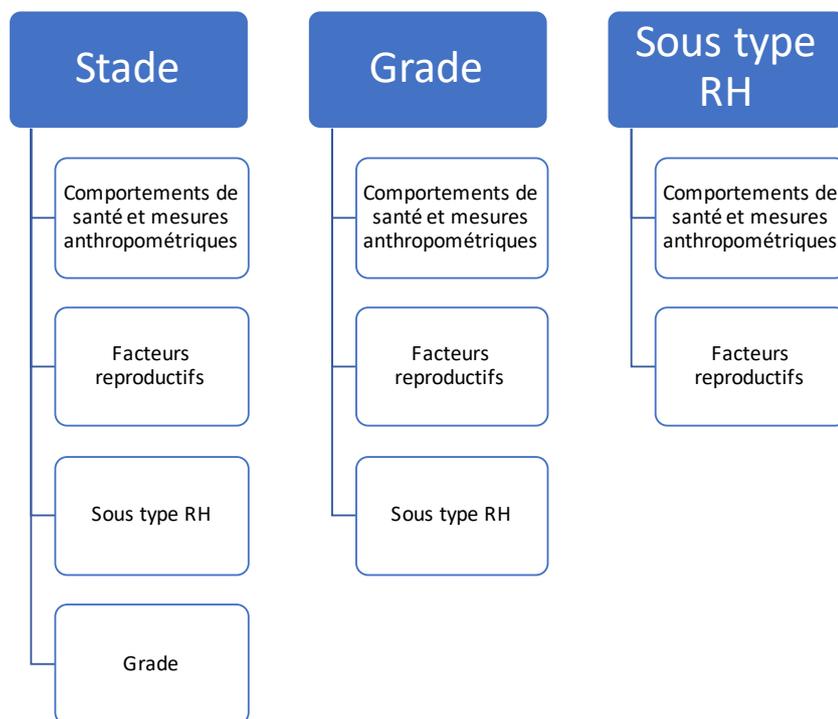
Hypothèses : les femmes qui ont un faible niveau d'éducation ont un risque plus important d'avoir un cancer du sein à un stade avancé et à un grade élevé au moment du diagnostic ainsi qu'un risque plus élevé de développer un cancer de moins bon pronostic.

2. METHODE ET SCHEMA D'ETUDE

Nous avons mené des analyses multivariées que nous avons répétées pour chacune des caractéristiques du cancer du sein (stade, grade et sous type) dont le codage a été décrit dans la partie 2, chapitre II. Dans un premier temps, j'ai regardé l'association entre l'éducation et les caractéristiques du cancer du sein et étudié l'impact de facteurs d'ajustement sur cette association à l'aide de modèles de régressions logistiques. Les associations globales entre l'éducation et les caractéristiques de cancer du sein ont été ajustées sur l'âge (modèle 1). À partir du modèle 1, nous avons fait des modèles ajustés indépendamment sur un certain nombre de covariables ainsi que des modèles ajustés par bloc de variables (d'une part sur tous les comportements de santé et les mesures anthropométriques et d'autre part sur tous les facteurs reproductifs). Les modèles ont ensuite été ajustés sur l'ensemble des covariables (modèle 2).

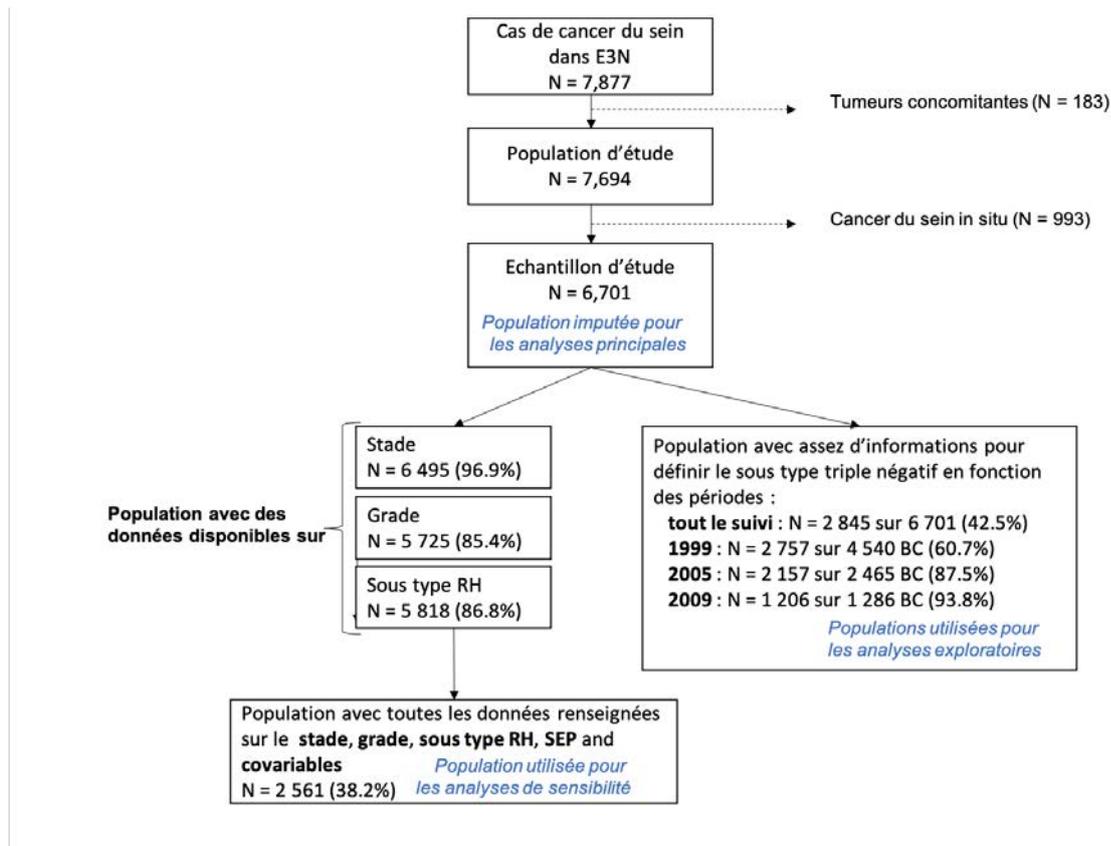
Les analyses sur le stade et le grade ont également été ajustées sur le sous type de cancer défini en fonction de la présence ou non de récepteurs hormonaux. Les analyses sur le stade ont en plus été ajustées sur le grade des cancers (figure 9).

FIGURE 9 : AJUSTEMENTS EN FONCTION DE LA CARACTERISTIQUE DU CANCER DU SEIN CONSIDEREE



Ces analyses ont été réalisées dans la cohorte E3N, sur données imputées (N = 6701) et une analyse de sensibilité a été reproduite sur les femmes n’ayant aucune donnée manquante pour tester la fiabilité de nos résultats (N = 2561). Une analyse exploratoire a également été menée pour étudier le lien entre la PSE et le développement de tumeurs triples négatives. La surexpression de la protéine HER2 n’a pas été systématiquement recherchée sur tous les cas de cancer du sein des femmes de la cohorte E3N. La proportion de données renseignées pour la protéine HER2 a commencé à augmenter à partir de 1999, pour arriver à 50% en 2005 et près de 90% en 2009. Les analyses réalisées sur les périodes les plus récentes réduisent considérablement la taille de l’échantillon. Le flowchart décrivant la sélection de la population est disponible en figure 10.

FIGURE 10 : SELECTION DE LA POPULATION



3. RESULTATS

A. CARACTERISTIQUES DE LA POPULATION D'ETUDE

La description des caractéristiques reproductives, comportementales, anthropométriques et sociodémographiques de la sous-population de femmes de E3N, séparée par caractéristiques des cancers du sein et par PSE, est fournie dans le tableau 12 et dans l'annexe 8.1 respectivement.

Les femmes qui ont un cancer à un stade avancé ont un cancer avec un grade plus élevé, plus souvent RH- ou triple négatif, sont plus jeunes, sont plus susceptibles d'avoir une consommation élevée d'alcool, d'avoir un poids plus élevé, d'avoir une durée de vie reproductive dans la norme, de ne pas participer au dépistage du cancer et d'être en préménopause.

Les femmes qui ont un cancer de grade élevé ont un cancer de stade plus avancé, plus souvent RH- ou triple négatif, sont moins susceptibles d'avoir des antécédents familiaux de cancer des ovaires et de l'utérus.

Les femmes qui présentent une tumeur RH- ont des cancers avec un stade plus avancé et de haut grade, ont un nombre d'enfants plus élevé et un âge à la première grossesse plus jeune. Elles sont également plus susceptibles d'avoir allaité, d'avoir une durée de vie reproductive dans la norme, et de ne pas avoir participé au dépistage.

B. ASSOCIATION ENTRE L'EDUCATION ET LE STADE DU CANCER DU SEIN

Aucune association n'a été retrouvée entre l'éducation et le stade du cancer du sein (M1, stade 3-4 versus 1, OR[95%CI] : 1,02 [0,85 ; 1,22], tableau 13).

TABLEAU 12 : CARACTERISTIQUES DES FEMMES AVEC LES DONNEES DISPONIBLES DANS E3N SELON LA CARACTERISTIQUE DU CANCER DU SEIN. ND = NON DEFINISSABLE

Variable	Stade					Grade					Sous type récepteurs hormonaux			
	NA	1	2	3-4	p	NA	1	2	3-4	p	NA	RH+	RH-	p
Education	253	n= 3851	n= 1814	n= 577	0,903	226	n= 750	n= 1886	n= 2863	0,059	222	n= 4776	n= 820	0,106
High, n(%)		1491 (38.72)	691 (38.09)	222 (38.47)			315 (42)	700 (37.12)	1084 (37.86)			1840 (38.53)	291 (35.49)	
Low, n(%)		2360 (61.28)	1123 (61.91)	355 (61.53)			435 (58)	1186 (62.88)	1779 (62.14)			2936 (61.47)	529 (64.51)	
Stade						141	n= 771	n= 1930	n= 2883	<0,001	150	n= 4835	n= 833	<0,001
1, n(%)			-				634 (82.23)	1243 (64.4)	1573 (54.56)			3044 (62.96)	415 (49.82)	
2, n(%)							120 (15.56)	562 (29.12)	972 (33.71)			1402 (29)	289 (34.69)	
3-4, n(%)							17 (2.2)	125 (6.48)	338 (11.72)			389 (8.05)	129 (15.49)	
Grade	911	n= 3450	n= 1654	n= 480	<0,001						737	n= 4363	n= 718	<0,001
1, n(%)		634 (18.38)	120 (7.26)	17 (3.54)								645 (14.78)	29 (4.04)	
2, n(%)		1243 (36.03)	562 (33.98)	125 (26.04)								1550 (35.53)	159 (22.14)	
3-4, n(%)		1573 (45.59)	972 (58.77)	338 (70.42)								2168 (49.69)	530 (73.82)	
Sous type récepteurs hormonaux	827	n= 3459	n= 1691	n= 518	<0,001	644	n= 674	n= 1709	n= 2698	<0,001				
RH+, n(%)		3044 (88)	1402 (82.91)	389 (75.1)			645 (95.7)	1550 (90.7)	2168 (80.36)					
RH-, n(%)		415 (12)	289 (17.09)	129 (24.9)			29 (4.3)	159 (9.3)	530 (19.64)					
Sous type triples négatifs	3719	n= 1674	n= 834	n= 268	<0,001	3136	n= 258	n= 768	n= 1563	<0,001	2973	n= 2443	n= 402	<0,001
Non, n(%)		1556 (92.95)	742 (88.97)	232 (86.57)			253 (98.06)	730 (95.05)	1377 (88.1)			2443 (100)	146 (36.32)	
Oui, n(%)		118 (7.05)	92 (11.03)	36 (13.43)			5 (1.94)	38 (4.95)	186 (11.9)			0 (0)	256 (63.68)	
Age	0	n= 3998	n= 1898	n= 599	0,085	0	n= 782	n= 1969	n= 2974	0,082	0	n= 4955	n= 863	0,166
Low, n(%)		1896 (47.42)	928 (48.89)	312 (52.09)			397 (50.77)	913 (46.37)	1448 (48.69)			2402 (48.48)	441 (51.1)	
High, n(%)		2102 (52.58)	970 (51.11)	287 (47.91)			385 (49.23)	1056 (53.63)	1526 (51.31)			2553 (51.52)	422 (48.9)	
Alcool	1394	n= 3150	n= 1484	n= 467	0,018	1213	n= 625	n= 1508	n= 2379	0,449	1224	n= 3901	n= 693	0,958
Aucune, n(%)		343 (10.89)	147 (9.91)	64 (13.7)			72 (11.52)	165 (10.94)	245 (10.3)			432 (11.07)	76 (10.97)	
Modérée, n(%)		2101 (66.7)	1029 (69.34)	285 (61.03)			415 (66.4)	1028 (68.17)	1580 (66.41)			2579 (66.11)	462 (66.67)	
Elevée, n(%)		706 (22.41)	308 (20.75)	118 (25.27)			138 (22.08)	315 (20.89)	554 (23.29)			890 (22.81)	155 (22.37)	
Tabac	23	n= 3984	n= 1892	n= 596	0,755	21	n= 780	n= 1962	n= 2962	0,400	24	n= 4936	n= 858	0,815
Jamais, n(%)		2668 (66.97)	1278 (67.55)	402 (67.45)			515 (66.03)	1350 (68.81)	1996 (67.39)			3311 (67.08)	585 (68.18)	
Ancien, n(%)		792 (19.88)	356 (18.82)	122 (20.47)			151 (19.36)	378 (19.27)	582 (19.65)			974 (19.73)	163 (19)	
Actuel, n(%)		524 (13.15)	258 (13.64)	72 (12.08)			114 (14.62)	234 (11.93)	384 (12.96)			651 (13.19)	110 (12.82)	
Régime occidental	1394	n= 3150	n= 1484	n= 467	0,538	1213	n= 625	n= 1508	n= 2379	0,892	1224	n= 3901	n= 693	0,109
[-2.93,-0.511], n(%)		957 (30.38)	475 (32.01)	131 (28.05)			188 (30.08)	458 (30.37)	727 (30.56)			1162 (29.79)	234 (33.77)	

Chapitre II : Déterminants sociaux et caractéristiques des cancers du sein

(-0.511,0.307], n(%)	1056 (33.52)	494 (33.29)	165 (35.33)		210 (33.6)	491 (32.56)	806 (33.88)		1318 (33.79)	219 (31.6)				
(0.307,7.78], n(%)	1137 (36.1)	515 (34.7)	171 (36.62)		227 (36.32)	559 (37.07)	846 (35.56)		1421 (36.43)	240 (34.63)				
Activité physique (MET)	88	n= 3946	n= 1875	n= 586	0,518	73	n= 775	n= 1942	n= 2935	0,354	75	n= 4893	n= 850	0,699
(7.57,35.9], n(%)		1187 (30.08)	597 (31.84)	181 (30.89)			245 (31.61)	579 (29.81)	925 (31.52)			1501 (30.68)	272 (32)	
(4.66,7.57], n(%)		1362 (34.52)	613 (32.69)	207 (35.32)			258 (33.29)	645 (33.21)	1008 (34.34)			1670 (34.13)	280 (32.94)	
[0,4.66], n(%)		1397 (35.4)	665 (35.47)	198 (33.79)			272 (35.1)	718 (36.97)	1002 (34.14)			1722 (35.19)	298 (35.06)	
Taille	95	n= 3943	n= 1871	n= 586	0,701	87	n= 771	n= 1939	n= 2928	0,650	83	n= 4886	n= 849	0,328
[135,160], n(%)		1685 (42.73)	801 (42.81)	259 (44.2)			325 (42.15)	816 (42.08)	1283 (43.82)			2087 (42.71)	386 (45.47)	
(160,164], n(%)		887 (22.5)	415 (22.18)	140 (23.89)			172 (22.31)	446 (23)	670 (22.88)			1112 (22.76)	184 (21.67)	
(164,190], n(%)		1371 (34.77)	655 (35.01)	187 (31.91)			274 (35.54)	677 (34.91)	975 (33.3)			1687 (34.53)	279 (32.86)	
Poids	126	n= 3919	n= 1862	n= 588	0,007	110	n= 773	n= 1928	n= 2914	0,115	113	n= 4862	n= 843	0,109
[29,55], n(%)		1303 (33.25)	564 (30.29)	188 (31.97)			268 (34.67)	591 (30.65)	954 (32.74)			1535 (31.57)	293 (34.76)	
(55,62], n(%)		1368 (34.91)	619 (33.24)	190 (32.31)			270 (34.93)	656 (34.02)	988 (33.91)			1655 (34.04)	287 (34.05)	
(62,163], n(%)		1248 (31.84)	679 (36.47)	210 (35.71)			235 (30.4)	681 (35.32)	972 (33.36)			1672 (34.39)	263 (31.2)	
Age aux premières règles	144	n= 3903	n= 1862	n= 586	0,325	126	n= 767	n= 1933	n= 2899	0,794	127	n= 4849	n= 842	0,920
≥12, n(%)		3078 (78.86)	1437 (77.18)	455 (77.65)			603 (78.62)	1522 (78.74)	2260 (77.96)			3790 (78.16)	660 (78.38)	
<12, n(%)		825 (21.14)	425 (22.82)	131 (22.35)			164 (21.38)	411 (21.26)	639 (22.04)			1059 (21.84)	182 (21.62)	
Antécédents familiaux cancers ovaires/utérus	0	n= 3998	n= 1898	n= 599	0,206	0	n= 782	n= 1969	n= 2974	0,063	0	n= 4955	n= 863	0,368
Non, n(%)		3784 (94.65)	1780 (93.78)	572 (95.49)			727 (92.97)	1861 (94.51)	2828 (95.09)			4686 (94.57)	809 (93.74)	
Oui, n(%)		214 (5.35)	118 (6.22)	27 (4.51)			55 (7.03)	108 (5.49)	146 (4.91)			269 (5.43)	54 (6.26)	
Antécédents familiaux cancers du sein	164	n= 3903	n= 1850	n= 578	0,089	137	n= 768	n= 1914	n= 2906	0,492	149	n= 4831	n= 838	0,275
Non, n(%)		1530 (39.2)	754 (40.76)	221 (38.24)			307 (39.97)	743 (38.82)	1170 (40.26)			1912 (39.58)	349 (41.65)	
Oui, n(%)		442 (11.32)	233 (12.59)	56 (9.69)			75 (9.77)	229 (11.96)	328 (11.29)			560 (11.59)	83 (9.9)	
ND, n(%)		1931 (49.47)	863 (46.65)	301 (52.08)			386 (50.26)	942 (49.22)	1408 (48.45)			2359 (48.83)	406 (48.45)	
Utilisation d'un contraceptif oral	0	n= 3998	n= 1898	n= 599	0,661	0	n= 782	n= 1969	n= 2974	0,055	0	n= 4955	n= 863	0,068
Jamais, n(%)		1811 (45.3)	838 (44.15)	274 (45.74)			337 (43.09)	933 (47.38)	1323 (44.49)			2271 (45.83)	366 (42.41)	
Oui - déjà, n(%)		2187 (54.7)	1060 (55.85)	325 (54.26)			445 (56.91)	1036 (52.62)	1651 (55.51)			2684 (54.17)	497 (57.59)	

PARTIE 3 - RESULTATS

Utilisation d'un THM	631	n= 3609	n= 1717	n= 538	0,001	555	n= 717	n= 1762	n= 2691	0,913	565	n= 4487	n= 766	0,192
Non, n(%)		3026 (83.85)	1465 (85.32)	486 (90.33)			612 (85.36)	1483 (84.17)	2287 (84.99)			3798 (84.64)	664 (86.68)	
Oui, n(%)		425 (11.78)	189 (11.01)	43 (7.99)			76 (10.6)	206 (11.69)	301 (11.19)			516 (11.5)	71 (9.27)	
ND, n(%)		158 (4.38)	63 (3.67)	9 (1.67)			29 (4.04)	73 (4.14)	103 (3.83)			173 (3.86)	31 (4.05)	
Nombre d'enfants	0	n= 3998	n= 1898	n= 599	0,764	0	n= 782	n= 1969	n= 2974	0,264	0	n= 4955	n= 863	0,013
3+, n(%)		1016 (25.41)	502 (26.45)	162 (27.05)			196 (25.06)	521 (26.46)	767 (25.79)			1281 (25.85)	237 (27.46)	
1-2, n(%)		2459 (61.51)	1141 (60.12)	364 (60.77)			482 (61.64)	1172 (59.52)	1847 (62.1)			3007 (60.69)	541 (62.69)	
0, n(%)		523 (13.08)	255 (13.44)	73 (12.19)			104 (13.3)	276 (14.02)	360 (12.1)			667 (13.46)	85 (9.85)	
Age à la première grossesse	174	n= 3895	n= 1847	n= 579	0,742	153	n= 759	n= 1909	n= 2904	0,394	158	n= 4821	n= 839	<0,001
[14,23], n(%)		1279 (32.84)	575 (31.13)	178 (30.74)			250 (32.94)	628 (32.9)	943 (32.47)			1524 (31.61)	297 (35.4)	
(23,26], n(%)		994 (25.52)	479 (25.93)	149 (25.73)			196 (25.82)	466 (24.41)	773 (26.62)			1216 (25.22)	245 (29.2)	
(26,59], n(%)		1099 (28.22)	538 (29.13)	179 (30.92)			209 (27.54)	539 (28.23)	828 (28.51)			1414 (29.33)	212 (25.27)	
Pas d'enfants, n(%)		523 (13.43)	255 (13.81)	73 (12.61)			104 (13.7)	276 (14.46)	360 (12.4)			667 (13.84)	85 (10.13)	
Durée d'allaitement	612	n= 3631	n= 1724	n= 528	0,523	535	n= 718	n= 1775	n= 2697	0,742	548	n= 4481	n= 789	0,049
(5.5,120], n(%)		605 (16.66)	326 (18.91)	94 (17.8)			118 (16.43)	309 (17.41)	466 (17.28)			783 (17.47)	141 (17.87)	
(2.5,5.5], n(%)		606 (16.69)	271 (15.72)	88 (16.67)			113 (15.74)	279 (15.72)	459 (17.02)			730 (16.29)	148 (18.76)	
[1,2.5], n(%)		876 (24.13)	399 (23.14)	131 (24.81)			179 (24.93)	414 (23.32)	655 (24.29)			1056 (23.57)	203 (25.73)	
0, n(%)		1544 (42.52)	728 (42.23)	215 (40.72)			308 (42.9)	773 (43.55)	1117 (41.42)			1912 (42.67)	297 (37.64)	
Durée de vie reproductive	300	n= 3822	n= 1819	n= 554	0,026	256	n= 757	n= 1880	n= 2832	0,122	270	n= 4738	n= 810	<0,001
[27; 41[, n(%)		3070 (80.32)	1458 (80.15)	471 (85.02)			590 (77.94)	1531 (81.44)	2273 (80.26)			3777 (79.72)	689 (85.06)	
≥41, n(%)		752 (19.68)	361 (19.85)	83 (14.98)			167 (22.06)	349 (18.56)	559 (19.74)			961 (20.28)	121 (14.94)	
Participation au dépistage	0	n= 3998	n= 1898	n= 599	<0,001	0	n= 782	n= 1969	n= 2974	0,175	0	n= 4955	n= 863	0,013
Oui, n(%)		3107 (77.71)	1423 (74.97)	419 (69.95)			615 (78.64)	1492 (75.77)	2245 (75.49)			3776 (76.21)	623 (72.19)	
Non, n(%)		891 (22.29)	475 (25.03)	180 (30.05)			167 (21.36)	477 (24.23)	729 (24.51)			1179 (23.79)	240 (27.81)	
Statut ménopausique au moment du diagnostic	45	n= 3984	n= 1885	n= 581	<0,001	38	n= 779	n= 1953	n= 2955	<0,001	40	n= 4932	n= 846	0,268
Premenopause, n(%)		546 (13.7)	329 (17.45)	128 (22.03)			118 (15.15)	343 (17.56)	353 (11.95)			639 (12.96)	122 (14.42)	
Postmenopause, n(%)		3438 (86.3)	1556 (82.55)	453 (77.97)			661 (84.85)	1610 (82.44)	2602 (88.05)			4293 (87.04)	724 (85.58)	

TABLEAU 13 : RESULTATS DES MODELES DES REGRESSIONS LOGISTIQUES MULTIVARIEES DE L'ASSOCIATION ENTRE L'EDUCATION ET LE STADE DU CANCER DU SEIN A PARTIR DES DONNEES IMPUTEES DE E3N (N = 6 701)

		Education	
		a. Stade 2 versus Stade 1	b. Stade 3-4 versus Stade 1
		OR (95%CI)	OR (95%CI)
M1		1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
Comportements et mesures anthropométriques	M1 + Alcool	1,03 [0,91 ; 1,15]	1,03 [0,86 ; 1,23]
	M1 + Tabac	1,03 [0,91 ; 1,15]	1,02 [0,85 ; 1,22]
	M1 + Régime occidental	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
	M1 + Activité physique	1,03 [0,91 ; 1,15]	1,02 [0,85 ; 1,22]
	M1 + Taille	1,03 [0,92 ; 1,16]	1,01 [0,85 ; 1,22]
	M1 + Poids	1,02 [0,91 ; 1,15]	1,02 [0,85 ; 1,22]
M1a : M1 + tous les comportements et mesures anthropométriques		1,01 [0,89 ; 1,13]	0,99 [0,82 ; 1,19]
Facteurs reproductifs	M1 + Age aux premières règles	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,23]
	M1 + Antécédents familiaux de cancers utérus/ovaires	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
	M1 + Antécédents familiaux de cancers du sein	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
	M1 + Utilisation de TMH	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
	M1 + Age à la première grossesse	1,05 [0,93 ; 1,18]	1,05 [0,87 ; 1,26]
	M1 + Durée d'allaitement	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,23]
	M1 + Durée de vie reproductive	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]
	M1 + Participation au dépistage	1,02 [0,91 ; 1,15]	1,00 [0,84 ; 1,20]
	M1 + Statut ménopausique	1,03 [0,92 ; 1,16]	1,03 [0,86 ; 1,23]
M1 + Utilisation d'un contraceptif oral	1,03 [0,92 ; 1,16]	1,02 [0,85 ; 1,22]	
M1b : M1 + tous les facteurs reproductifs		1,06 [0,94 ; 1,19]	1,01 [0,84 ; 1,23]
	M1 + Sous type moléculaire	1,02 [0,91 ; 1,15]	1,01 [0,84 ; 1,21]
	M1 + Grade	1,01 [0,90 ; 1,14]	1,00 [0,83 ; 1,20]
M2 : Modèle ajusté sur l'ensemble des facteurs		1,02 [0,90 ; 1,15]	0,96 [0,79 ; 1,17]

TMH : traitement hormonal de la ménopause

C. ASSOCIATION ENTRE L'ÉDUCATION ET LE GRADE DU CANCER DU SEIN

Un faible niveau d'éducation est associé à un grade plus élevé de cancer du sein (M1, grade 3-4 versus 1, OR[95%CI] : 1,17 [1,00 ; 1,37], tableau 14). Cette association est très peu affectée par l'ajustement sur les facteurs comportementaux et anthropométriques (M1a, grade 3-4 versus 1, OR[95%CI] : 1,15 [0,98 ; 1,34], tableau 14) ainsi que sur les facteurs reproductifs (M1b, grade 3-4 versus 1, OR[95%CI] : 1,18 [1,00 ; 1,38], tableau 14). La taille de l'effet est également très peu modifiée dans le modèle totalement ajusté (M2, grade 3-4 versus 1, OR[95%CI] : 1,15 [0,98 ; 1,36], tableau 14). Les mêmes associations avec la même taille d'effet ont été retrouvées pour les analyses sur le grade 2 versus le grade 1.

D. ASSOCIATION ENTRE L'ÉDUCATION ET SOUS TYPE RH DU CANCER DU SEIN

Un faible niveau d'éducation est associé à un risque plus important de tumeurs RH- (M1, RH- versus RH+, OR[95%CI] : 1,13 [0,96 ; 1,33], tableau 15). Cette association est très peu affectée par l'ajustement sur les facteurs comportementaux et anthropométriques (M1a, RH- versus RH+, OR[95%CI] : 1,13 [0,96 ; 1,33], tableau 15). Cependant elle semble plus largement impactée par l'ajustement sur les facteurs reproductifs (M1b, RH- versus RH+, OR[95%CI] : 1,06 [0,90 ; 1,26], tableau 15), ce qui est principalement dû à l'ajustement sur l'âge à la première grossesse (M1 + âge à la première grossesse, RH- versus RH+, OR[95%CI] : 1,07 [0,91 ; 1,26], tableau 15). La taille de l'effet est également atténuée dans le modèle totalement ajusté (M2, RH- versus RH+, OR[95%CI] : 1,06 [0,89 ; 1,27], tableau 15).

E. ANALYSES DE SENSIBILITE ET ANALYSES EXPLORATOIRES

Les caractéristiques de la population de cas complets sont fournies en annexe 8.2. Les analyses de sensibilité réalisées sur les données en cas complets (annexes 8.3 à 8.5) ont montré des tendances similaires en terme de la taille de l'effet pour le grade (annexe 8.4) et le sous type RH (annexe 8.5) bien que nous manquons de puissance dû à la réduction de l'échantillon (38% de la population de l'échantillon d'étude).

Les analyses exploratoires sur les tumeurs triples négatives ne montrent pas d'association entre l'éducation et le fait de développer une tumeur triple négative (annexe 8.6).

TABLEAU 14 : RESULTATS DES MODELES DES REGRESSIONS LOGISTIQUES MULTIVARIEES DE L'ASSOCIATION ENTRE L'EDUCATION ET LE GRADE DU CANCER DU SEIN A PARTIR DES DONNEES IMPUTEES DE E3N (N = 6 701)

		Education	
		a. Grade 2 versus Grade 1	b. Grade 3-4 versus Grade 1
		OR (95%CI)	OR (95%CI)
M1		1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
Comportements et mesures anthropométriques	M1 + Alcool	1,17 [0,99 ; 1,37]	1,17 [1,00 ; 1,37]
	M1 + Tabac	1,16 [0,98 ; 1,36]	1,16 [0,99 ; 1,36]
	M1 + Régime occidental	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Activité physique	1,15 [0,98 ; 1,36]	1,17 [1,00 ; 1,37]
	M1 + Taille	1,17 [0,99 ; 1,38]	1,16 [1,00 ; 1,36]
	M1 + Poids	1,16 [0,98 ; 1,36]	1,16 [0,99 ; 1,36]
M1a : M1 + tous les comportements et mesures anthropométriques		1,15 [0,97 ; 1,36]	1,15 [0,98 ; 1,34]
Facteurs reproductifs	M1 + Age aux premières règles	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Antécédents familiaux de cancers utérus/ovaires	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Antécédents familiaux de cancers du sein	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Utilisation d'un traitement hormonal de la ménopause	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Age à la première grossesse	1,19 [1,01 ; 1,41]	1,18 [1,01 ; 1,38]
	M1 + Durée d'allaitement	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,37]
	M1 + Durée de vie reproductive	1,16 [0,99 ; 1,37]	1,17 [1,00 ; 1,36]
	M1 + Participation au dépistage	1,16 [0,98 ; 1,37]	1,16 [0,99 ; 1,36]
	M1 + Statut ménopausique	1,17 [0,99 ; 1,38]	1,17 [1,00 ; 1,36]
M1 + Utilisation d'un contraceptif oral	1,16 [0,98 ; 1,37]	1,17 [1,00 ; 1,37]	
M1b : M1 + tous les facteurs reproductifs		1,18 [0,99 ; 1,40]	1,18 [1,00 ; 1,38]
	M1 + Sous type moléculaire	1,16 [0,98 ; 1,37]	1,15 [0,98 ; 1,35]
M2 : Modèle ajusté sur l'ensemble des facteurs		1,16 [0,98 ; 1,38]	1,15 [0,98 ; 1,36]

TABLEAU 15 : RESULTATS DES MODELES DES REGRESSIONS LOGISTIQUES MULTIVARIEES DE L'ASSOCIATION ENTRE L'EDUCATION ET LE SOUS TYPE DU CANCER DU SEIN A PARTIR DES DONNEES IMPUTEES DE E3N (N = 6 701)

		Education
		OR (95%CI)
M1		1,13 [0,96 ; 1,33]
Comportements et mesures anthropométriques	M1 + Alcool	1,13 [0,97 ; 1,33]
	M1 + Tabac	1,13 [0,96 ; 1,32]
	M1 + Régime occidental	1,14 [0,97 ; 1,33]
	M1 + Activité physique	1,13 [0,96 ; 1,33]
	M1 + Taille	1,12 [0,96 ; 1,32]
	M1 + Poids	1,14 [0,97 ; 1,34]
M1a : M1 + tous les comportements et mesures anthropométriques		1,13 [0,96 ; 1,33]
Facteurs reproductifs	M1 + Age aux premières règles	1,13 [0,96 ; 1,33]
	M1 + Antécédents familiaux de cancers utérus/ovaires	1,13 [0,96 ; 1,33]
	M1 + Antécédents familiaux de cancers du sein	1,13 [0,96 ; 1,33]
	M1 + Utilisation d'un traitement hormonal de la ménopause	1,13 [0,96 ; 1,33]
	M1 + Age à la première grossesse	1,07 [0,91 ; 1,26]
	M1 + Durée d'allaitement	1,14 [0,97 ; 1,33]
	M1 + Durée de vie reproductive	1,12 [0,96 ; 1,32]
	M1 + Participation au dépistage	1,12 [0,96 ; 1,32]
	M1 + Statut ménopausique	1,13 [0,97 ; 1,33]
M1 + Utilisation d'un contraceptif oral	1,14 [0,97 ; 1,34]	
M1b : M1 + tous les facteurs reproductifs		1,06 [0,90 ; 1,26]
M2 : Modèle ajusté sur l'ensemble des facteurs		1,06 [0,89 ; 1,27]

4. DISCUSSION

A partir des données d'une cohorte prospective française, nous avons observé qu'un faible niveau d'éducation est associé à des tumeurs de plus haut grade et au sous type n'exprimant pas de récepteurs hormonaux (RH-). L'ajustement sur les facteurs reproductifs, et notamment l'âge à la première grossesse, a largement atténué l'association entre l'éducation et le sous type de cancer du sein, alors qu'aucun ajustement, ni sur les facteurs reproductifs, ni sur les comportements de santé n'a modifié l'association observée avec le grade. Aucune association n'a été mise en évidence dans cette étude entre l'éducation et le stade du cancer au moment du diagnostic et le sous type de cancer du sein triple négatif.

A. PSE, STADE ET GRADE DU CANCER DU SEIN

Contrairement à la plupart des précédentes études qui ont été menées, nous ne retrouvons pas d'association entre la PSE et le stade du cancer du sein au moment du diagnostic. Un des principaux facteurs qui peut jouer sur le stade est la participation au dépistage, qui a été retrouvée comme plus fréquente chez les femmes plus favorisées, ce qui peut amener à un sur-diagnostic des cancers à un stade moins avancé dans cette population²¹². La population de la cohorte E3N a été sélectionnée parmi les femmes affiliées à un système national de santé (MGEN) qui assure principalement les personnes travaillant dans le système éducatif français et leur conjoint ce qui en fait une population plus éduquée que la population générale et qui pourrait être plus alerte sur les messages de prévention et sur les premiers signes de l'apparition de la maladie, et ainsi participer plus au dépistage. Une précédente étude a d'ailleurs mis en évidence que les pratiques de dépistage du cancer du sein dans la cohorte E3N étaient très élevées par rapport aux taux qui ont été rapportés dans une étude transversale sur une population représentative de la population générale²¹³. Dans notre population, 78% des femmes avec un niveau d'éducation élevé déclarent avoir participé au dépistage. Elles sont 75% chez les femmes avec un niveau d'éducation plus faible (annexe 8.1). Une autre explication pourrait venir du fait que dans notre étude nous avons sélectionné la population pour ne garder que les cas de cancer du sein invasifs et nous avons donc exclu les cas de cancer in situ.

Notre étude a mis en évidence une association entre l'éducation et le grade du cancer du sein. Ce résultat est cohérent avec ceux de précédentes études qui ont retrouvé une association entre la défavorisation approchée par le score de Townsend^{78,148} ou le niveau d'éducation de la zone de résidence¹⁴⁹ et un grade élevé. Notre étude complète les données de la littérature en mettant en évidence un lien avec la PSE à l'échelle individuelle et en montrant que l'ajustement sur différents facteurs comportementaux, anthropométriques et reproductifs ne modifie pas l'association observée ce qui implique que d'autres facteurs, notamment biologiques, pourraient intervenir dans cette association.

B. PSE ET SOUS TYPES DE CANCER DU SEIN

L'association que nous observons entre l'éducation et le sous type de cancer défini par la présence ou non de récepteurs hormonaux est conforme aux résultats observés dans de précédentes études^{148,149,152,214}. Nous observons que cette association est largement expliquée par la prise en compte de l'âge à la première grossesse et dans une moindre mesure, par le nombre de grossesses (annexe 8.7). Ces phénomènes, qui sont socialement distribués, sont des facteurs de risque reconnus de cancer du sein de moins bon pronostics.

Notre étude n'a pas mis en évidence de lien entre la PSE et le risque de développer des tumeurs triples négatives comme cela a précédemment été identifié dans la littérature. Pour ces analyses nous avons distingué les cas diagnostiqués en fonction de différentes périodes car un des marqueurs permettant de définir le sous type, la protéine HER2, n'a été dosé en routine qu'à partir des années 2000. Dans les analyses sans distinction de périodes ou celles à partir des années 1999, une faible proportion de sous type défini en fonction des récepteurs hormonaux et de la protéine HER2 a pu être attribuée et une grande majorité des données a donc été imputée (figure 10). Dans les analyses réalisées sur les périodes plus récentes, une plus grande proportion de sous types était renseignée.

Nos résultats soulignent qu'un faible niveau d'éducation est associé à des cancers de haut grade et au développement de sous types n'exprimant pas de récepteurs

hormonaux, tous deux étant de moins bon pronostic. Le lien entre éducation et sous type s'explique majoritairement par l'impact des facteurs reproductifs, notamment un âge précoce au premier enfant et le nombre élevé d'enfants. Cependant l'association entre l'éducation et le grade n'est pas expliquée par l'ensemble des facteurs pris en compte dans cette étude. De futurs travaux devraient être menés pour identifier les mécanismes impliqués dans le lien entre éducation et grade des cancers, en incluant notamment des facteurs biologiques. L'association entre l'éducation et les sous types n'exprimant pas de récepteurs hormonaux, notamment dans le cas des cancers triples-négatifs, devrait être investiguée dans une population plus large.

PARTIE 4 - DISCUSSION GENERALE

CHAPITRE I. DISCUSSION DES RESULTATS

1. SYNTHÈSE DES RESULTATS

Plusieurs mécanismes ont été proposés pour expliquer comment l'environnement social peut influencer le développement de maladies et l'état de santé en général. Il a été suggéré que la position sociale peut aboutir à différents types d'expositions qui vont pouvoir impacter les processus biologiques et à terme altérer l'état de santé. Mon travail de thèse a mis en évidence, à travers des travaux menés sur différentes cohortes européennes, que l'éducation est associée au niveau d'inflammation à l'âge adulte, indépendamment des comportements de santé comme le tabagisme, la consommation d'alcool et la sédentarité (article 1). Ces premiers travaux soutiennent l'hypothèse selon laquelle l'environnement social peut être incorporé biologiquement par d'autres mécanismes. Par ailleurs l'inflammation apparaît comme étant le socle commun d'un ensemble de pathologies, notamment les cancers. Dans la suite de mes travaux, j'ai étudié l'association entre l'inflammation et le développement de deux types de cancer ainsi que l'impact de différents comportements de santé et du niveau d'éducation sur cette association. Nos résultats ont mis en évidence une association entre le niveau d'inflammation et le statut cas ou contrôles de cancer du sein et de lymphomes non hodgkinien, très peu impactée par les différents ajustements (article 2). Le cancer est une maladie particulièrement intéressante à étudier, d'une part car il joue un rôle majeur dans les inégalités sociales de mortalité, mais également car c'est une maladie qui relève d'expositions et processus étiologiques variés, et de mécanismes biologiques différents. Il peut être un modèle de maladies particulièrement pertinent pour étudier comment le social devient biologique, via une approche fondée sur le parcours de vie. Or les mécanismes d'incorporation biologique et l'approche au cours de la vie ont été peu utilisés dans l'étude des inégalités sociales de cancer. Le développement du cancer du sein, pour lequel nous avons retrouvé une association avec l'inflammation, est d'autant plus intéressant que son lien avec la PSE est différent de la majorité des autres types de cancers.

Suite à ces différents travaux j'ai adopté une approche globale afin d'étudier le lien entre la PSE à différents moments de la vie et l'incidence du cancer du sein et ses différents sous-types moléculaires ainsi que ses caractéristiques cliniques au moment du diagnostic en considérant

un des ensembles les plus larges de variables de médiation potentielles, incluant des facteurs comportementaux, anthropométriques, reproductifs mais aussi biologiques. Dans le travail 3 détaillé dans la *Partie 3, chapitre I* (article 3) je me suis intéressée au gradient social de cancer du sein dans deux cohortes prospectives. Parmi les trois marqueurs de PSE étudiés, qui représentent respectivement le début de la vie des femmes pendant leur enfance (profession du père), leur entrée dans la vie adulte (niveau d'éducation), puis leur vie adulte (dernier métier exercé), les femmes avec un niveau d'éducation élevé avaient un risque accru de développer un cancer du sein. Cet excès de risque chez les femmes éduquées s'explique principalement par un âge plus avancé au moment de la première grossesse. Contrairement à notre hypothèse de départ, la participation au dépistage ne semble pas avoir d'effet sur l'association que nous observons dans notre échantillon d'étude. De plus, bien que nous ayons considéré un large ensemble de variables de natures différentes, notamment des comportements de santé, des mesures anthropométriques, des facteurs reproductifs incluant également les antécédents familiaux qui pourraient expliquer cette différence, une partie de l'association persiste. L'ensemble de ces résultats suggère que d'autres facteurs pourraient intervenir dans le lien entre l'éducation et le risque de cancer du sein, ou que d'autres mécanismes pourraient être en jeu. Ces résultats sont renforcés par le fait que nous avons obtenu des résultats qui sont cohérents en terme de direction et de taille de l'effet à partir d'une cohorte indépendante, EPIC-Italie, bien qu'avec une puissance moindre en raison de la taille réduite de la cohorte.

Dans le travail 4 détaillé dans la *Partie 3, chapitre I*, (article 4 - en cours de finalisation) nous avons voulu intégrer la dimension biologique en regardant comment le niveau d'éducation peut avoir un impact sur le fonctionnement biologique qui à terme pourrait contribuer au développement du cancer du sein. Les mécanismes d'incorporation biologique sont généralement étudiés à travers des biomarqueurs inflammatoires, métaboliques et cardiovasculaires ou à travers des scores biologiques comme la charge allostatique. Comme nous en avons discuté dans la partie 1, chapitre III, section 2, il existe un nombre important de facteurs associés au risque de cancer du sein dont les mécanismes d'action s'expliquent par la modification des hormones sexuelles. C'est pourquoi nous avons considéré des hormones sexuelles en plus des biomarqueurs communément utilisés pour étudier l'incorporation biologique de l'environnement social et le risque de cancer du sein. Pour ce

travail j'ai procédé en deux étapes : dans un premier temps, j'ai regardé quels marqueurs biologiques étaient associés au risque de cancer du sein puis j'ai étudié le lien entre l'éducation et les biomarqueurs précédemment identifiés comme associés au risque de cancer du sein. Les analyses de la première étape ont mis en évidence qu'un niveau élevé de CRP, d'HDL, de FC, d'œstradiol et de testostérone ainsi qu'un faible niveau de LDL, de cholestérol et de SHBG étaient associés à un risque plus élevé de cancer du sein. Ces associations étaient très peu impactées par l'ajustement sur les comportements de santé, les mesures anthropométriques et les facteurs reproductifs. Par ailleurs, les analyses réalisées dans la seconde étape ont permis d'observer que les femmes ayant atteint un niveau d'éducation élevé ont des niveaux de CRP et de testostérone plus faibles et des taux de HDL, œstradiol et SHBG plus élevés. L'ajustement sur les comportements de santé, les mesures anthropométriques et les facteurs reproductifs a atténué les associations observées entre l'éducation et la CRP, le HDL, l'œstradiol et la SHBG. Le groupe de facteurs intermédiaires qui a le plus influencé ces associations est celui des comportements de santé et des mesures anthropométriques. Les facteurs reproductifs ont également eu un effet sur les associations entre l'éducation et l'œstradiol et la SHBG. Les associations entre l'éducation et la CRP et le HDL sont restées significatives malgré les ajustements. Bien que la force de l'association se soit affaiblie, l'association entre l'éducation et la SHBG n'est pas totalement expliquée par les variables d'ajustement. A l'inverse des autres biomarqueurs, l'association entre l'éducation et la testostérone n'est pas modifiée par l'ajustement sur les comportements de santé, les mesures anthropométriques et les facteurs reproductifs suggérant que d'autres mécanismes d'incorporation peuvent être à l'œuvre.

Dans le travail 5 détaillé dans la *Partie 3, chapitre II*, (article 5 - en cours de préparation) j'ai regardé le lien entre l'éducation et les caractéristiques initiales du cancer et les différents sous-types moléculaires. Les résultats n'ont pas mis en évidence de lien entre l'éducation et le stade du cancer au moment du diagnostic. Par contre, les femmes moins éduquées ont des cancers plus agressifs (grade 2 et 3) et développent des sous types de moins bons pronostics (RH-) bien que l'association pour les sous types de cancers ne soit pas significative. L'association entre l'éducation et le grade n'est pas expliquée par les facteurs de risque considérés dans l'analyse. Cependant, ils expliquent une part conséquente du lien entre

l'éducation et le sous type déterminé en fonction de la présence ou non de récepteurs hormonaux, notamment les facteurs reproductifs en particulier l'âge à la première grossesse.

2. FORCES ET LIMITES

Une des premières forces de ce travail est que les analyses mises en place dans les trois travaux relatifs au cancer du sein reposent sur des mécanismes identifiés dans le cadre de précédentes études que nous avons menées. En effet, j'ai montré que l'éducation peut être incorporée biologiquement en modifiant le niveau d'inflammation indépendamment des comportements de santé. J'ai également identifié que l'inflammation est liée au développement du cancer, notamment les cancers du sein. Les résultats de ces différentes études suggèrent que l'éducation peut impacter les processus biologiques qui vont amener au développement du cancer, justifiant les analyses menées sur le cancer du sein.

A. DANS L'ENSEMBLE DES TRAVAUX RELATIFS AU CANCER DU SEIN

Ces travaux ont été menés dans une large cohorte prospective de femmes françaises dans laquelle un très grand nombre de médiateurs potentiels et de biomarqueurs étaient disponibles, ce qui constitue une forte originalité par rapport à la littérature existante. La variété des données disponibles permet en particulier de prendre en compte plusieurs mécanismes par lesquels l'environnement social tout au long de la vie peut agir sur le risque et la progression des cancers du sein. Le design prospectif limite à la fois les biais de rappel et les biais de causalité inverse. Les femmes ont été suivies depuis les années 1990 jusqu'à aujourd'hui, ce qui nous permet d'avoir une vision à long terme de l'évolution de la maladie et ce même en restreignant les périodes d'études pour s'adapter aux différents objectifs. Le processus d'identification des cas de cancer se déroule en deux étapes : dans un premier temps les femmes déclarent l'évolution de leur état de santé au moment où elles répondent au questionnaire et dans un second temps les médecins de E3N prennent contact avec le médecin des femmes afin d'avoir des informations complémentaires, ce qui a permis de valider les cas de cancer auto-déclarés et d'éviter les erreurs de classification.

La principale limite de la cohorte E3N réside dans son recrutement, en particulier lorsque nous travaillons sur les inégalités sociales. Les femmes volontaires incluses dans cette étude étaient

toutes affiliées à un système national de santé (MGEN) qui assure principalement les personnes travaillant dans le système éducatif français et leur conjoint. Bien que la cohorte comprenne également du personnel administratif et de nettoyage, le niveau d'éducation moyen de la cohorte est plus élevé que celui de la population générale. Cependant, en considérant la PSE à différentes périodes de la vie depuis l'enfance, nous avons pu observer une certaine variabilité sociale. Mais la variabilité sociale de nos échantillons, et en particulier de ceux disposant de données biologiques, est moins grande qu'en population générale, ce qui laisse envisager que certains de nos résultats sont probablement sous-estimés par rapport à ce que nous observerions dans une population moins sélectionnée socialement, en particulier l'effet de l'éducation sur le risque de cancer. D'autres associations pourraient être différentes en population générale par rapport à ce qui a été observé dans notre échantillon comme par exemple le rôle du dépistage. A cette sélection initiale liée aux critères d'inclusion dans E3N, nous ne pouvons pas exclure un biais de sélection dû à l'attrition et à la perte de suivi. Pour tenir compte de l'incertitude quant aux données manquantes, nous avons utilisé des méthodes d'imputations multiples et conduit des analyses de sensibilité sur les cas complets. Même en ayant considéré un grand nombre de covariables et de biomarqueurs, il est toujours possible que d'autres facteurs contribuent à la relation entre la PSE et le développement du cancer du sein, comme notamment la densité mammaire. En effet, plusieurs études ont rapporté une association indépendante entre la PSE, y compris l'éducation, et la densité mammaire^{215,216} qui est elle-même associée au risque de cancer du sein¹¹¹.

B. SPECIFICITES DES TRAVAUX 4 ET 5 RELATIFS AU CANCER DU SEIN

Les biomarqueurs utilisés dans le cadre du quatrième travail ont été dosés à partir d'échantillons sanguins prélevés sur des femmes volontaires dans le contexte d'études avec des objectifs spécifiques, ce qui a mené à une sélection de la population que nous avons utilisé par rapport à la population sur laquelle nous avons travaillé dans le premier travail relatif au cancer du sein. Dans cette population, les femmes chez qui les biomarqueurs ont été dosés étaient plus instruites, fumaient moins, étaient légèrement plus actives physiquement, participaient plus au dépistage et avaient plus souvent des antécédents de cancer par rapport

aux femmes pour lesquelles nous n'avons pas d'information sur les biomarqueurs (annexe 7.8). Les femmes sur lesquelles nous avons mené nos analyses ont des comportements globalement plus sains et prêtent plus attention à leur santé que ce qui peut être observé en population générale ce qui a également pu amener à une sous-estimation des effets que nous observons. Les mêmes observations ont été faites en comparant les caractéristiques des femmes qui se sont portées volontaires pour participer à la biobanque et celles qui ont décliné l'invitation²¹⁷. Malgré le grand ensemble de variables considérées pour l'ajustement, il peut y avoir quelques facteurs de confusion résiduels (par exemple, la prise d'un traitement) non pris en compte. De plus, bien que la cohorte E3N inclut un grand nombre de femmes, l'effectif se restreint quand nous considérons les participantes pour lesquelles les biomarqueurs sont disponibles. Ainsi il se peut que nous n'ayons pas la puissance suffisante pour détecter certains effets de petite taille entre les biomarqueurs et le risque de cancer du sein. De plus, dans le travail 4 pour lequel nous avons travaillé sur 3% de la population issue du troisième travail, nous n'avons pas retrouvé d'association significative entre l'éducation et le risque de cancer du sein que nous observions dans le travail 3. Nous avons donc décidé de ne pas faire un modèle global pour regarder le lien entre l'éducation et le risque de cancer du sein en incluant les marqueurs biologiques en plus des comportements de santé, des mesures anthropométriques et des facteurs reproductifs. Cela a également été le cas dans le travail 5 pour lequel nous n'avons pas regardé l'impact des marqueurs biologiques sur le lien entre les PSE et les caractéristiques des cancers du fait du faible effectif de cas de cancer du sein pour lequel nous disposons de l'ensemble des informations. Pour finir nous n'avons également pas pu étudier l'impact de la PSE sur les tumeurs triples négatives car un des marqueurs, la protéine HER2, n'était pas recherché systématiquement au moment du début du suivi des femmes de la cohorte E3N (annexe 8.8). Enfin, les différents travaux menés dans le cadre de cette thèse étaient à but exploratoire, afin d'identifier les médiateurs potentiels parmi un large ensemble de facteurs. Des analyses de médiation devront faire suite à cette thèse afin d'avoir une meilleure compréhension des voies causales par lesquelles les processus éducatifs opèrent.

Malgré la diversité et la qualité des données disponibles, et la taille importante des participantes à la cohorte E3N, nos travaux présentent certaines limites notamment en ce qui concerne la sélection des différentes populations utilisées et des tailles d'échantillons parfois

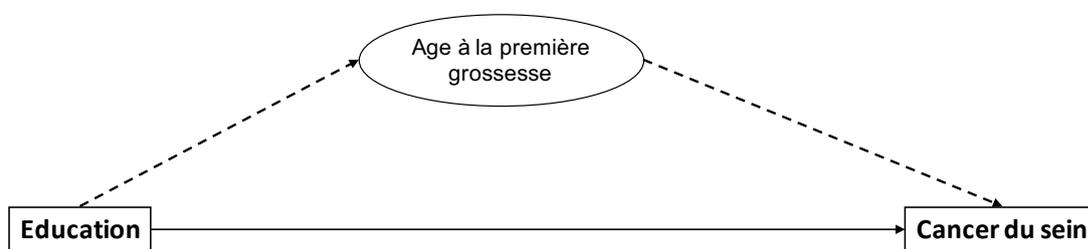
réduites. Il est néanmoins rare de pouvoir combiner autant d'informations sociales, biologiques, comportementales et reproductives en lien avec le cancer et ses caractéristiques, ce qui souligne l'originalité de ce travail de thèse et ouvre sur un ensemble de perspectives dont nous discuterons dans la Partie 5.

CHAPITRE II. PSE ET CANCER DU SEIN : MECANISMES POSSIBLES

1. POSITIONS SOCIO-ECONOMIQUES ET INCIDENCE DU CANCER DU SEIN

Le troisième travail de cette thèse a permis de mettre en évidence que l'âge au moment de la première grossesse était le facteur principal, parmi tous ceux que nous avons considéré, qui expliquait le sur-risque de cancer du sein chez les femmes favorisées (figure 11).

FIGURE 11 : SCHEMA RECAPITULATIF DES RESULTATS OBTENUS DANS LE CADRE DU TROISIEME TRAVAIL. LES FLECHES EN POINTILLE INDIQUENT UN LIEN INDIRECT ENTRE L'EDUCATION ET LE RISQUE DE CANCER DU SEIN.

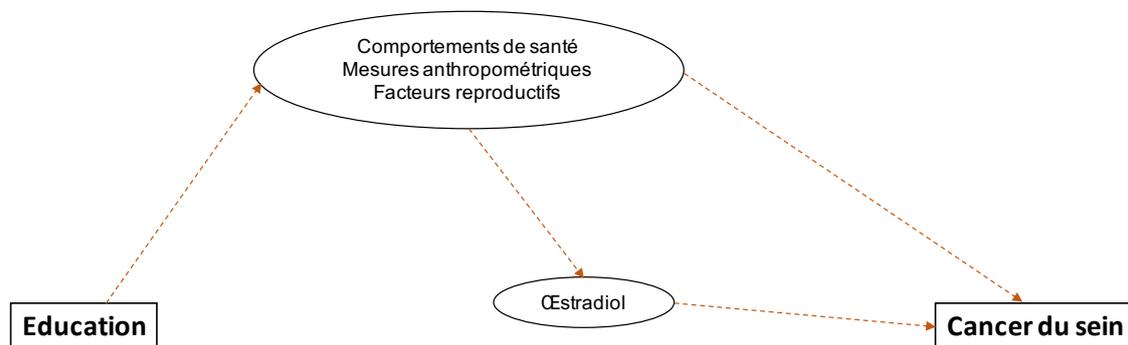


L'âge au moment de la première grossesse va pouvoir intervenir de différentes manières sur le risque de cancer du sein. Dans un premier temps, c'est durant cette période que les cellules des glandes mammaires vont se différencier totalement. José Russo a résumé les différentes étapes qui amènent à la différenciation cellulaire et l'impact que cela peut avoir dans le développement du cancer du sein²¹⁸. Les glandes mammaires qui se sont développées depuis la naissance jusqu'à la puberté, se trouvent dans un état peu différencié. Au moment de la première grossesse, la différenciation reprend pour atteindre son état de différenciation terminale au cours du dernier trimestre de la grossesse. Il souligne dans ces travaux que le pourcentage de cellules en prolifération diminue avec la différenciation des structures mammaires²¹⁸. De plus, il indique également que le pourcentage de cellules positives aux récepteurs aux œstrogènes et à la progestérone est plus élevé dans les structures les moins différenciées et qu'il diminue progressivement avec le degré de différenciation des structures²¹⁸. Cela implique une plus grande susceptibilité des structures peu différenciées à la transformation par des carcinogènes²¹⁸. La période entre la puberté et l'âge à la première grossesse semble être alors une période de vulnérabilité pour l'initiation du cancer du sein,

qui dure d'autant plus longtemps que l'âge à la première grossesse est tardif. Par ailleurs, il a également été mis en évidence que la première grossesse induit des changements hormonaux à long terme, notamment une réduction des niveaux de prolactine et d'œstrogènes et une augmentation des niveaux de SHBG, ce qui modifie également le risque de cancer du sein ²¹⁹.

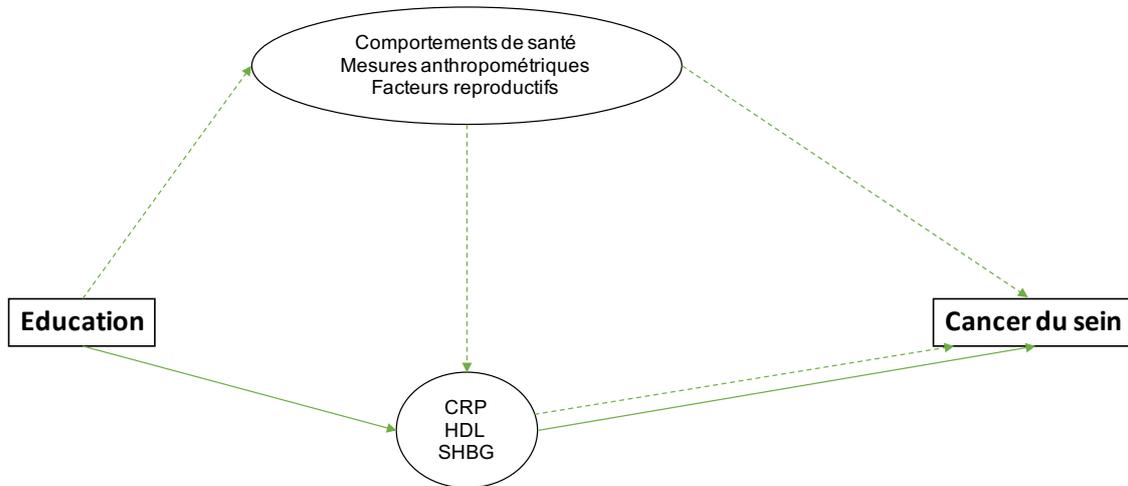
Dans le quatrième travail de cette thèse, nous avons mis en évidence un fonctionnement biologique socialement différencié qui peut s'accompagner d'un risque différent vis-à-vis du cancer du sein. Nos résultats montrent un profil biologique différent en fonction du niveau d'études : les femmes plus éduquées se caractérisent au niveau biologique par des niveaux d'inflammation et de testostérone plus faibles ainsi que par des taux d'œstradiol, de HDL et de SHBG plus élevés par rapport aux femmes moins éduquées. Différents mécanismes semblent expliquer ces différences comme nous pouvons le voir à travers les quatre figures suivantes :

FIGURE 12 : MECANISMES IMPLIQUES DANS LE LIEN ENTRE L'EDUCATION ET LE TAUX D'ŒSTRADIOL



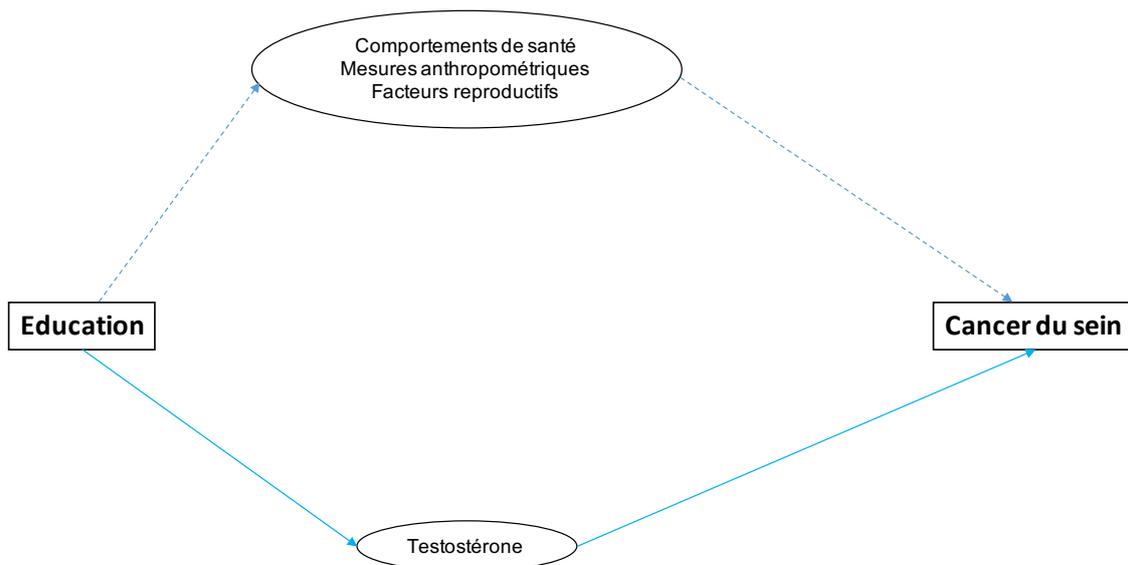
- l'association entre l'éducation et l'œstradiol semble totalement expliquée par les différents ajustements, à la fois sur les comportements de santé et les mesures anthropométriques mais également sur les facteurs reproductifs (figure 12). Ainsi, les femmes favorisées ont des comportements et des caractéristiques reproductives qui modulent leurs expositions aux œstrogènes et qui potentialisent leur risque de cancer du sein par rapport aux femmes moins favorisées. Ce mécanisme va dans le sens de la première hypothèse de ce travail de thèse.

FIGURE 13 : MECANISMES IMPLIQUES DANS LE LIEN ENTRE L'EDUCATION ET LES TAUX DE CRP, D'HDL ET DE SHBG



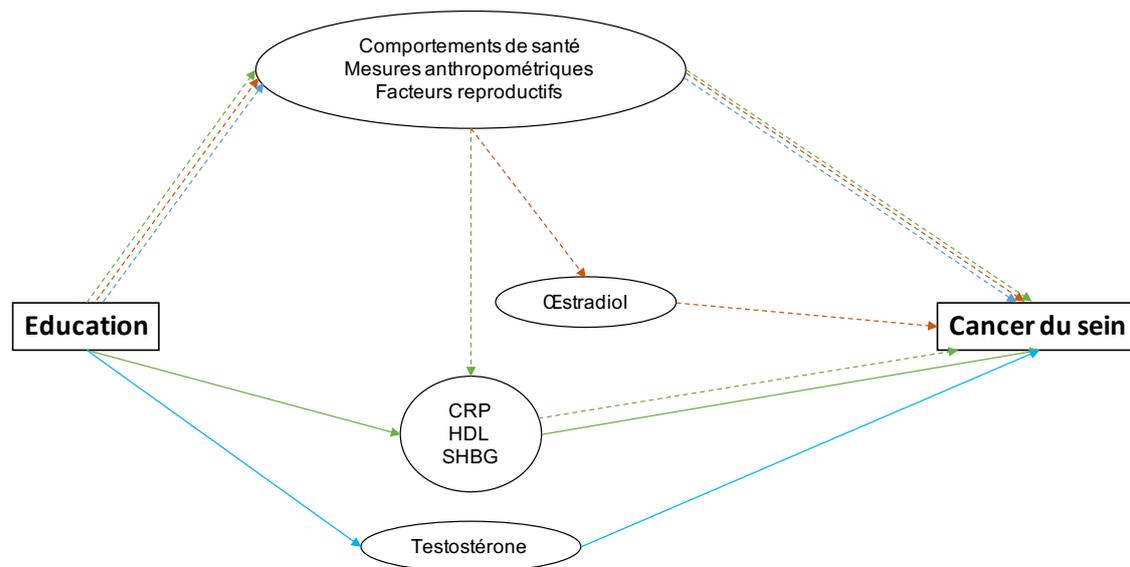
- les associations entre l'éducation et la CRP, le HDL et la SHBG sont en partie expliquées par les variables d'ajustement, en particulier sur les comportements de santé et les mesures anthropométriques (figure 13, trait pointillé). L'obésité pourrait jouer un rôle important dans l'association entre la PSE et ces marqueurs biologiques comme nous l'avons identifié dans le premier travail de cette thèse. Malgré l'impact des facteurs d'ajustement, une partie de l'association semble persister (figure 13, trait plein)

FIGURE 14 : MECANISMES IMPLIQUES DANS LE LIEN ENTRE L'EDUCATION ET LE TAUX DE TESTOSTERONE



- l'association entre l'éducation et la testostérone n'est pas expliquée par l'ajustement ni sur les comportements de santé et les mesures anthropométriques, ni sur les facteurs reproductifs (figure 14).

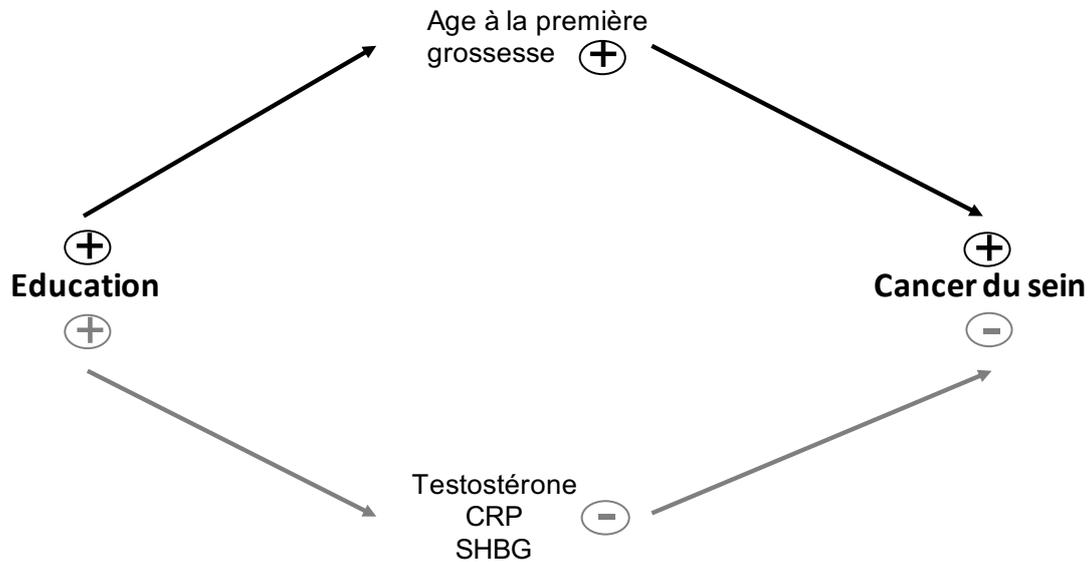
FIGURE 15 : SCHEMA RECAPITULATIF DES RESULTATS OBTENUS DANS LE CADRE DE MON TRAVAIL. LES FLECHES EN POINTILLE INDIQUENT UN LIEN INDIRECT ENTRE L'EDUCATION ET LE RISQUE DE CANCER DU SEIN.



La figure 15 résume les différents mécanismes qui peuvent lier l'éducation au fonctionnement biologique. Les associations qui persistent entre l'éducation et la CRP, le HDL, la SHBG et la testostérone, indiquées par les traits pleins, suggèrent que des voies alternatives méritent d'être étudiées davantage. Une hypothèse à creuser pourrait être celle du stress psychosocial dont nous discuterons davantage à la fin de ce chapitre et dans le chapitre III de la discussion, mais cela peut également être le résultat d'expositions chimiques et de maladies infectieuses.

Les mécanismes mis en évidence dans les travaux 3 et 4 suggèrent qu'un niveau élevé d'éducation peut avoir à la fois un effet délétère et un effet protecteur face au cancer du sein. Comme nous pouvons le voir sur la figure 16, un niveau élevé d'éducation est associé à un âge plus tardif au moment de la première grossesse ce qui augmente le risque de cancer du sein. Par ailleurs, les femmes qui ont un niveau d'éducation plus élevé ont un profil biologique (en particulier vis-à-vis de la testostérone, de la CRP et de la SHBG) moins à risque de cancer du sein.

FIGURE 16 : EFFETS DELETERE ET PROTECTEUR DU NIVEAU D'EDUCATION SUR LE RISQUE DE CANCER DU SEIN MIS EN EVIDENCE DANS LES TRAVAUX 3 ET 4.



Il est possible que le sur-risque que nous observons chez les femmes éduquées soit plus fort après ajustement sur la testostérone par exemple car le faible niveau de testostérone chez les femmes éduquées les « protège » du cancer du sein (contrairement aux femmes moins éduquées pour qui le niveau élevé constitue un facteur de risque). Ces différents liens entre éducation et cancer du sein nécessitent d'étudier ces mécanismes à l'aide d'analyses de médiation, comme je le développerai dans les perspectives.

2. POSITIONS SOCIO-ECONOMIQUES, CARACTERISTIQUES INITIALES ET SOUS TYPES DE CANCER DU SEIN

Dans le travail 5 de cette thèse, nous avons mis en évidence que les caractéristiques initiales et le développement des différents sous types de cancer du sein peuvent être socialement différents : les femmes ne développent pas les mêmes cancers en fonction de leur niveau d'éducation.

Nos analyses ont montré que les femmes avec un niveau plus faible d'éducation tendent à avoir des cancers n'exprimant pas de récepteurs hormonaux (RH-), ce qui s'explique largement par l'âge à la première grossesse comme nous pouvons le voir sur la figure 17.

L'ajustement sur le nombre de grossesses, montre également, dans une moindre mesure, une atténuation de l'effet.

FIGURE 17 : MECANISMES IMPLIQUES DANS LE LIEN ENTRE L'EDUCATION ET LES SOUS TYPES MOLECULAIRES DES CANCERS DU SEIN



Il a précédemment été suggéré que certains facteurs de risque peuvent affecter le microenvironnement inflammatoire du sein, entraînant une inflammation délétère, qui peut impacter les caractéristiques des cancers du sein. L'inflammation peut jouer un rôle dans pratiquement tous les aspects du développement tumoral, depuis l'initiation, la promotion, l'angiogenèse, jusqu'à la dissémination de métastases. Parmi les facteurs de risque identifiés, l'obésité et le fait d'avoir eu des enfants semblent particulièrement important dans la mise en place d'un environnement pro-inflammatoire. Il est possible que les mécanismes qui expliquent l'impact du nombre d'enfants reflètent les processus d'involution postnatale. L'involution correspond au processus qui ramène le sein à son état de base après la grossesse et l'allaitement²²⁰ durant une phase d'apoptose contrôlée et de remodelage des tissus²²¹. Elle se caractérise par la présence de nombreux médiateurs inflammatoires, d'un influx de macrophages et d'une infiltration importante de cellules immunitaires²²¹. L'allaitement semble être un facteur important pour atténuer les effets de l'involution en fin de grossesse.

Cette inflammation peut favoriser le développement de mutations qui vont pouvoir amener au développement d'un cancer ou provoquer une incapacité à éliminer les cellules mutées ce qui peut aboutir à des caractéristiques et sous types tumoraux de mauvais pronostic^{82,220}. Une précédente étude suggère que l'environnement pro-inflammatoire qui se produit au cours de l'involution mammaire initiée par l'absence d'allaitement ou de courtes périodes d'allaitement, peut favoriser la survie et l'expansion des cellules progénitrices basales, qui

n'expriment ni les récepteurs hormonaux ni une surexpression de la protéine HER2²²². Il a été mis en évidence que les cytokines inflammatoires favorisent l'inflammation des tissus et activent des voies de signalisation qui favorisent la biologie agressive notamment des tumeurs triples négatives⁸¹. Il a également été mis en évidence que l'inflammation pourrait influencer l'agressivité des tumeurs RH+²²¹. Il a par ailleurs été proposé dans les recherches mises en place pour comprendre les mécanismes impliqués dans la résistance aux thérapies endocriniennes que l'inflammation pourrait supprimer les récepteurs aux œstrogènes et diminuer leur expression²²³.

Dans mon travail, l'association observée entre un faible niveau d'éducation et un grade élevé ne semble pas expliquée par les comportements de santé, les mesures anthropométriques et les facteurs reproductifs que nous avons considérés dans notre étude. L'ajustement sur l'IMC comme marqueur de l'obésité, plutôt que distinctement sur le poids et la taille n'a pas modifié davantage l'association observée.

Les associations qui persistent malgré les ajustements réalisés dans les travaux 4 et 5 suggèrent que d'autres facteurs comportementaux et environnementaux restent à identifier. Le premier travail que j'ai mené dans le cadre de ma thèse a mis en évidence un lien entre l'éducation et l'inflammation dans différents contextes européens, non expliqué par les facteurs de risque classiques. Cette association est également observée sur une sous population de E3N dans le travail 4. Des travaux précédents menés dans l'équipe ont mis en évidence que le stress psychosocial pendant l'enfance était associé au fonctionnement biologique à l'âge adulte à travers une mesure de la charge allostatique incluant notamment des marqueurs de l'inflammation. L'impact du stress psychosocial pourrait alors être un mécanisme intéressant à étudier. Cependant, peu de cohortes intègrent ce type d'expositions dans les variables collectées. Il sera toutefois possible d'intégrer cette exposition dans de futurs travaux car une question relative au stress vécu a été intégrée au dernier questionnaire de E3N comme cela est détaillé en perspective à la fin de cette partie.

CHAPITRE III. PARTICULARITES DE L'INCORPORATION BIOLOGIQUE DE L'ENVIRONNEMENT SOCIAL SOULEVEES PAR LES DIFFERENTS TRAVAUX DE CETTE THESE

1. STRESS PSYCHOSOCIAL ET MARQUEURS BIOLOGIQUES : CAS DE LA TESTOSTERONE

Nous avons discuté dans la Partie 1, chapitre I, section 5 de cette thèse de l'impact des expositions psychosociales sur la mise en place des mécanismes de réponses au stress impliquant l'axe HPA et des répercussions qu'une exposition chronique peut engendrer sur les processus inflammatoires, métaboliques, cardiovasculaires et neuroendocriniens. Cependant l'impact de ces expositions chroniques sur les hormones sexuelles est peu connu et équivoque. Dans nos analyses, la principale hormone qui semble associée à l'éducation indépendamment des facteurs comportementaux, anthropométriques et reproductifs est la testostérone. L'association entre le stress et le niveau de testostérone est complexe et constitue une thématique de recherche à part entière. Il semblerait que la modification des taux de testostérone suite à un stress dépende du type de stress ainsi que de la durée et la fréquence d'expositions à ce stress.

L'hypothèse principale qui est retrouvée dans la littérature est que l'exposition à un stress chronique aboutirait à une diminution du taux de testostérone. La testostérone est le produit final de l'axe hypothalamo-pituitaire-gonadique (HPG). D'un point de vue biologique, il a été suggéré que la sécrétion d'androgènes comme finalité de l'axe HPG pourrait être inhibée lors de l'activation de l'axe HPA par la production de glucocorticoïdes qui en résulte durant un stress²²⁴. Cependant, des études ont également trouvé que l'exposition à un stress chronique peut aboutir à une augmentation des taux de testostérone à la fois chez l'animal²²⁵ et chez l'humain²²⁶. Il a récemment été proposé que les axes HPA et HPG sont des processus parallèles qui s'influencent continuellement, et pas nécessairement des axes mutuellement antagonistes. Il est possible que les deux axes, HPA et HPG, puissent agir distinctement, en compétition ou en coopération, en fonction de la façon dont les individus perçoivent leur contexte social et que cela puisse s'ajuster en fonction de la demande de l'instant²²⁷. La communication inhibitrice ou activatrice de ces deux axes pourrait dépendre du type de stress

considéré. D'après Shirtcliff & al, le phénomène de co-activation est décrit comme un "couplage", où le cortisol et la testostérone (et la déhydroépiandrostérone) sont positivement liés chez un individu²²⁸. Ce couplage positif entre le cortisol et la testostérone a été retrouvé dans de précédentes études^{229,230}. Les contextes à l'origine d'une activation du double axe pourraient être des contextes dans lesquels le stress stimule l'axe HPA par des mécanismes psychosociaux ou évaluatifs et où le défi stimule l'axe HPG par la motivation à atteindre des objectifs ou la menace de perte de statut²²⁸.

Dans certaines études, une association a également été identifiée entre un stress psychosocial, notamment suite à des expérimentations utilisant le *trier social stress test (TSST)*, et une augmentation du taux de testostérone^{231,232}. Le TSST est issu d'un protocole validé et bien établi pour induire un stress social aigu en laboratoire, qui amène notamment une dimension de stress socio-évaluatif. Une récente étude a suggéré que « l'augmentation des niveaux de testostérone après le TSST correspond à des niveaux élevés de testostérone pendant la compétition et pourrait être une indication que les participants ont considéré la situation de stress psychosocial comme un défi personnel. Le TSST avec un jury et une menace pour le moi social des participants correspond à une situation dans laquelle la domination et la soumission sont négociées, et qui tend à augmenter la sécrétion de testostérone »²³¹. L'augmentation du taux de testostérone est notamment retrouvée dans le contexte précédant une compétition.

D'autres mécanismes, principalement observés dans le cadre d'un stress aigu, ont été suggérés : la testostérone est principalement libérée par les gonades mais également en plus petite quantité par les glandes surrénales. La sécrétion surrénalienne est sous le contrôle de l'ACTH (hormone adrénocorticotrope), qui est générée par l'axe HPA suite à un stress. Il est possible que l'augmentation de la testostérone induite par le stress puisse être partiellement synthétisée par les surrénales²³². Il a également été suggéré que le cortisol, qui est le produit final de l'axe HPA, pourrait influencer la production de testostérone plus fortement dans les glandes surrénales que la production dans les gonades²³³.

Bien que les résultats des différentes études qui se sont intéressées au lien entre le stress et le taux de testostérone divergent, plusieurs pistes ont été avancées pour expliquer les mécanismes qui sont à l'œuvre. Il serait intéressant d'approfondir ces recherches pour affiner

les connaissances sur ce sujet, tant au niveau des durées que des types d'expositions, afin de voir si la testostérone pourrait être considérée comme un marqueur de stress psychosocial.

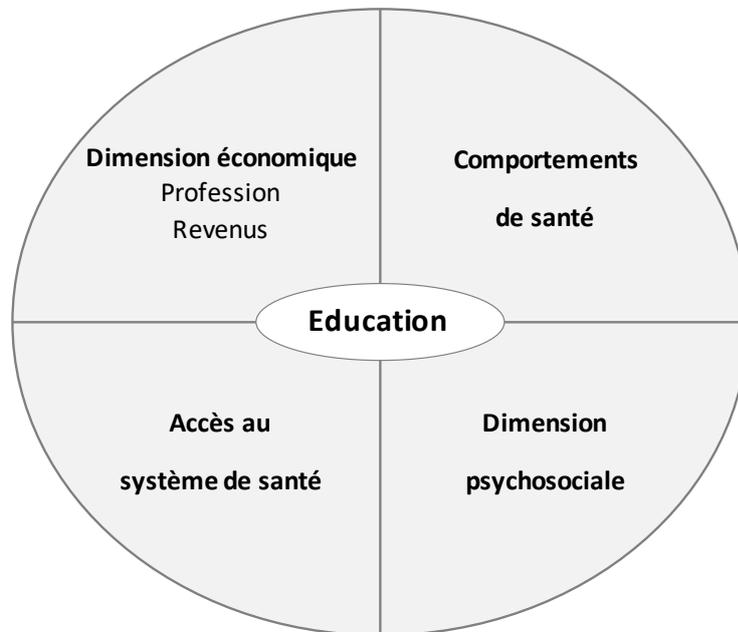
2. L'ÉDUCATION : DETERMINANT MAJEUR DE L'INCORPORATION BIOLOGIQUE DE L'ENVIRONNEMENT SOCIAL ET DU DEVELOPPEMENT DES MALADIES

Dans les différents travaux menés dans cette thèse, nous avons considéré trois PSE à différents moments de la vie dans un même modèle dans l'idée de prendre en compte l'historique des expositions socioéconomiques dans le cadre du parcours de vie, sur la biologie et la santé. L'éducation est l'indicateur de PSE qui ressort dans tous nos travaux, que ce soit quand nous nous intéressons au cancer du sein et à ses caractéristiques ou à l'incorporation biologique de l'environnement social.

Nous avons brièvement abordé l'utilisation du niveau d'éducation en tant qu'indicateur de PSE dans la Partie 1, chapitre I de cette thèse. Il a précédemment été identifié que les mécanismes capables de médier l'association entre l'éducation et la santé peuvent être regroupés suivant quatre dimensions (figure 18) : *économique, comportements de santé, socio-psychologique* et *accès au système de santé*²³⁴. Ces différentes dimensions ainsi que l'interprétation du niveau d'éducation dans le contexte global du parcours de vie ont été abordées dans les travaux de Galobardes & al¹⁰. A l'inverse des autres indicateurs de PSE qui vont refléter une partie de la vie relativement bien définie, l'éducation se rapporte à une période de vie qui fait la transition entre l'enfance et l'âge adulte. D'un côté, le niveau d'éducation atteint au début de l'âge adulte est le résultat de l'accessibilité et des performances au début du cursus en école primaire et reflète notamment les ressources matérielles et culturelles de l'environnement familial dans lequel l'enfant évolue. D'un autre côté, le niveau d'éducation atteint finalement va avoir un impact sur l'emploi exercé et les revenus (*dimension économique*). En ce sens, l'éducation capture à la fois l'impact de l'environnement vécu pendant l'enfance ainsi que les ressources dont bénéficient les personnes à l'âge adulte. L'éducation peut être vue à la fois comme un moteur d'opportunités mais aussi en tant que reproducteur d'inégalités²³⁴. De plus, les connaissances et les compétences acquises par les années d'études vont permettre aux personnes plus éduquées d'être à la fois plus à même d'interpréter les messages de prévention, de les appliquer

(*comportements de santé*), ainsi que d'utiliser les ressources disponibles comme notamment les services de santé (*accès au système de santé*) mais cela va également leurs permettre de faire face à différents stress et soucis du quotidien (*dimension socio-psychologique*).

FIGURE 18 : L'EDUCATION EST LIEE A QUATRE DIMENSIONS.



Parmi les différentes dimensions auxquelles sont liées l'éducation et qui sont susceptibles de jouer un rôle dans les modifications biologiques, nos résultats ont montré que les comportements de santé et la profession à l'âge adulte n'expliquaient pas complètement les associations observées. Cela souligne le fait que la dimension psychosociale et les voies de réponses au stress pourraient être des mécanismes intéressants à explorer.

Il est suspecté que l'éducation pourrait jouer un rôle de régulateur du stress physiologique en favorisant les stratégies d'adaptation et en procurant un sentiment de contrôle accru. Chandola et al. ont mis en évidence que le sentiment de contrôle pouvait médier une part de l'association entre l'éducation et la santé. Ils résumant ainsi que « l'éducation accroît le sentiment de contrôle d'une personne sur sa vie en développant les capacités d'analyse et de communication. Le sentiment de contrôle personnel peut également améliorer la santé en favorisant les comportements sains, en contrôlant la dépendance pour des bénéfices futurs à long terme pour la santé, alors que le manque de contrôle personnel peut être un facteur de stress avec des conséquences physiologiques négatives» ²³⁵.

PARTIE 5 - CONCLUSION ET PERSPECTIVES

CHAPITRE I. IMPACTS POTENTIELS DES RESULTATS

En France, la lutte contre les inégalités a pris une place importante à partir du deuxième Plan Cancer (2009-2013). Toutefois, les constats établis à la fin de ce plan et du suivant, le troisième Plan Cancer (2014-2019), font état d'une progression face à la réduction des inégalités, « mais de façon dispersée et insuffisamment structurée tant sur les inégalités sociales que territoriales »^{236,237}. Une des recommandations alors émises en vue du prochain plan était de « développer les approches globales prenant en compte le cumul de facteurs de risques à l'échelle d'un individu et combinant les différents comportements identifiés comme sains »^{236,237}. Les différents travaux menés dans le cadre de ma thèse s'inscrivent dans cette recommandation en apportant une vision générale des mécanismes qui peuvent lier l'environnement social au développement du cancer du sein et de ses différents sous types par la prise en compte des comportements de santé, des mesures anthropométriques, des facteurs relatifs à la vie reproductive et le fonctionnement biologique des femmes.

Il était également souligné que « les mesures nationales (tabac, alcool, dépistages), même globalement efficaces, ont pu finalement renforcer les inégalités du fait de leur moindre adaptation aux populations les plus à risque ou les moins favorisées. »^{236,237}. Une des recommandations était alors de « fonder les actions de réduction des inégalités sur des approches médicales et populationnelles permettant de cibler précisément les publics à prioriser »^{236,237}. A travers mes travaux de thèse, nous avons montré que les femmes qui ont un faible niveau d'éducation ont un profil biologique différent des femmes plus éduquées, et développent des cancers plus agressifs et de moins bon pronostic. Une plus grande attention devrait alors être portée aux spécificités y compris biologiques de cette population, au niveau des prochaines mesures de prévention, qui pourraient par exemple renforcer l'aide à la mise en place de comportements visant à limiter les phénomènes d'inflammation du tissu mammaire, notamment la pratique d'une activité physique pour lutter contre l'obésité et souligner les bienfaits de l'allaitement face au développement de cette maladie. Nos résultats font également état d'un risque global de cancer du sein, sans tenir compte du statut de gravité, plus important chez les femmes éduquées. Les femmes qui poursuivent de longues études tendent à retarder le moment auquel elles auront des enfants, ce qui les place à un risque plus important de développer ce type de cancer. Compte tenu de l'augmentation de l'âge à la première grossesse dans toutes les classes sociales, il est probable que cette inégalité

d'incidence, en faveur des femmes moins éduquées, ne soit plus retrouvée dans des populations de femmes avec un suivi plus récent, alors même que les femmes moins éduquées ont plus de risque de faire des cancers plus graves.

Ce travail met également en évidence une biologie socialement stratifiée, avec des taux de biomarqueurs potentiellement impliqués dans le développement de maladies différents en fonction de la PSE. Ainsi les travaux de cette thèse montrent une susceptibilité biologique vis-à-vis du cancer du sein différente en fonction du niveau d'études, que ce soit en terme d'incidence globale ou en terme de profil de cancer. La recherche et la surveillance de marqueurs biologiques apparaissent alors importantes à prendre en compte, ce qui suggère en amont une meilleure sensibilisation des médecins aux inégalités sociales de santé. Nos résultats montrent que l'éducation est notamment associée à l'inflammation ainsi qu'à la testostérone. Les marqueurs inflammatoires, et notamment la CRP, sont des marqueurs biologiques qui sont couramment dosés ce qui n'est pas le cas pour la testostérone. Nos résultats suggèrent que la testostérone pourrait être un biomarqueur intéressant à recueillir de manière plus systématique.

CHAPITRE II. PERSPECTIVES

Les travaux menés dans le cadre de cette thèse ont permis de mieux comprendre le lien entre PSE et fonctionnement biologique et d'affiner les connaissances sur les différents mécanismes impliqués dans les inégalités sociales face au cancer du sein notamment face à l'incidence de ce cancer, à ses différents sous types et plus largement à ses caractéristiques initiales. Cependant, d'autres travaux sont nécessaires pour mieux comprendre l'ensemble des mécanismes impliqués dans le continuum du cancer du sein.

D'une part, au vu des limites soulevées pour les travaux 4 et 5, d'autres études incluant une population plus large de participants pour lesquels les marqueurs biologiques ont été dosés seraient nécessaires pour approfondir les mécanismes de l'incorporation biologique de l'éducation face au risque de cancer du sein et l'impact que cela pourrait avoir sur le développement des différents sous types et du grade au moment du diagnostic. Il serait aussi important d'étudier ces mécanismes dans une population plus représentative de la population générale avec notamment une plus grande variabilité sociale. La cohorte E3N est incluse dans une étude de plus grande envergure, l'étude EPIC, qui regroupe 521 000 hommes et femmes issus de 10 pays européens. Il serait intéressant de répliquer nos analyses dans cette population.

D'autre part, malgré le large ensemble de facteurs que nous avons pris en compte dans nos analyses, d'autres facteurs seraient important à considérer dans les prochaines études, notamment la densité mammaire. Il serait également intéressant d'inclure dans les analyses une mesure d'exposition au stress chronique. Une question relative aux violences sexuelles et physiques vécues avant et après 15 ans a d'ailleurs été ajoutée dans le dernier questionnaire envoyé aux femmes E3N en 2018. La mesure de biomarqueurs répétée dans le temps serait aussi importante à prendre en compte car cela permettrait de tenir compte de la variabilité intra-individuelle des concentrations.

D'un point de vue méthodologique, nos travaux ont permis d'identifier des médiateurs potentiels du lien entre l'éducation et le risque de cancer du sein parmi un large panel de facteurs. Il serait par le suite important d'utiliser des approches causales afin de mieux comprendre les voies par lesquelles les processus éducatifs opèrent.

Des analyses de survie devraient être réalisées en combinant l'ensemble des mécanismes que nous avons mis en évidence afin d'approfondir l'exploration de l'hypothèse générale de ce travail de thèse, qui était que les inégalités sociales observées dans la mortalité par cancer du sein sont la conséquence de plusieurs mécanismes combinés, notamment une distribution socialement différenciée de l'incidence de la maladie et ses différents sous-types moléculaires mais aussi de ses caractéristiques initiales, et de l'accès au système de soin.

Ces travaux ont soulevé de nouvelles questions concernant les mécanismes, notamment le lien entre la PSE et les hormones sexuelles, qui mériteraient d'être explorées dans de futurs travaux. Les femmes qui ont un faible niveau d'éducation ont un taux de testostérone plus élevé, ce qui pourrait entraîner une augmentation du risque de cancer du sein dans cette population. Cette association n'est pas expliquée par les facteurs comportementaux, anthropométriques et reproductifs que nous avons considérés ici. Les mécanismes impliqués dans cette association, et notamment le rôle d'un stress psychosocial, devraient être investigués.

Enfin, il pourrait être intéressant de répliquer nos analyses à d'autres pathologies cancéreuses, et d'inclure alors une population d'hommes.

REFERENCES BIBLIOGRAPHIQUES

1. NHS England » Definitions for Health Inequalities [Internet]. [cited 2021 Apr 7]. Available from: <https://www.england.nhs.uk/ltphimenu/definitions-for-health-inequalities/>
2. Whitehead M, Dahlgren G. Concepts and principles for tackling social inequities in health: :45.
3. Les hommes cadres vivent toujours 6 ans de plus que les hommes ouvriers - Insee Première - 1584 [Internet]. [cited 2021 Jan 7]. Available from: <https://www.insee.fr/fr/statistiques/1908110>
4. McNamara CL, Balaj M, Thomson KH, Eikemo TA, Solheim EF, Bambra C. The socioeconomic distribution of non-communicable diseases in Europe: findings from the European Social Survey (2014) special module on the social determinants of health. *Eur J Public Health*. 2017 Feb 1;**27**(suppl_1):22–26.
5. Mackenbach JP, Stirbu I, Roskam A-JR, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. 2008 Jun 5;**358**(23):2468–2481.
6. Kunst AE, Groenhouf F, Mackenbach JP, Health EW. Occupational class and cause specific mortality in middle aged men in 11 European countries: comparison of population based studies. EU Working Group on Socioeconomic Inequalities in Health. *BMJ*. 1998 May 30;**316**(7145):1636–1642.
7. Krieger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;**18**:341–378.
8. Bartley M. Health inequality, an introduction to concepts, theories and methods. Cambridge, UK: Polity; 2016.
9. Haut Conseil de la santé publique. Indicateurs de suivi de l'évolution des inégalités sociales de santé dans les systèmes d'information en santé. 2013.
10. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health*. 2006 Jan;**60**(1):7–12.
11. Reducing social inequalities in cancer: evidence and priorities for research [Internet]. International Agency for Research on Cancer; 2019 [cited 2021 Apr 5]. Available from: <http://search.ebscohost.com/login.aspx?direct=true&site=edspub-live&scope=site&type=44&db=edspub&authtype=ip,guest&custid=ns011247&groupid=main&profile=eds&bquery=AN%2020337414>
12. Singh-Manoux A, Adler NE, Marmot MG. Subjective social status: its determinants and its association with measures of ill-health in the Whitehall II study. *Soc Sci Med*. 2003 Mar;**56**(6):1321–1333.
13. Cialani C, Mortazavi R. The effect of objective income and perceived economic resources on self-rated health. *Int J Equity Health*. 2020 Nov 4;**19**(1):196.
14. Galobardes B, Shaw M, Lawlor DA, Lynch JW. Indicators of socioeconomic position

(part 2). *J Epidemiol Community Health*. 2006 Feb;**60**(2):95–101.

15. OMS - Principaux concepts relatifs aux déterminants sociaux de la santé - https://www.who.int/social_determinants/final_report/key_concepts_fr.pdf?ua=1.
16. Anctil H, Jobin L, Pigeon M, et al. La santé et ses déterminants: mieux comprendre pour mieux agir. Québec: Ministère de la santé et des services sociaux, Direction des communications; 2012.
17. Hertzman C, Boyce T. How experience gets under the skin to create gradients in developmental health. *Annu Rev Public Health*. 2010;**31**:329-347 3p following 347.
18. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 2002 Apr;**31**(2):285–293.
19. Lynch J, Smith GD. A life course approach to chronic disease epidemiology. *Annu Rev Public Health*. 2005;**26**:1–35.
20. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health*. 2003 Oct;**57**(10):778–783.
21. Wiesel TN, Hubel DH. SINGLE-CELL RESPONSES IN STRIATE CORTEX OF KITTENS DEPRIVED OF VISION IN ONE EYE. *J Neurophysiol*. 1963 Nov;**26**:1003–1017.
22. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med*. 1994 Oct;**39**(7):887–903.
23. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int J Epidemiol*. 2001 Aug;**30**(4):668–677.
24. Krieger N. Embodiment: a conceptual glossary for epidemiology. *J Epidemiol Community Health*. 2005 May;**59**(5):350–355.
25. Hertzman C. The biological embedding of early experience and its effects on health in adulthood. *Ann N Y Acad Sci*. 1999;**896**:85–95.
26. Adler NE, Stewart J. Health disparities across the lifespan: meaning, methods, and mechanisms. *Ann N Y Acad Sci*. 2010 Feb;**1186**:5–23.
27. Kelly-Irving M, Delpierre C. The Embodiment Dynamic over the Life Course: A Case for Examining Cancer Aetiology. In: Meloni M, Cromby J, Fitzgerald D, Lloyd S, editors. *The Palgrave Handbook of Biology and Society* [Internet]. London: Palgrave Macmillan UK; 2018. p. 519–540. Available from: https://doi.org/10.1057/978-1-137-52879-7_22
28. Kelly-Irving M. Allostatic load: how stress in childhood affects life-course health outcomes. :22.
29. Blane D, Kelly-Irving M, Errico A d', Bartley M, Montgomery S. Social-biological transitions: how does the social become biological? *Longitudinal and Life Course Studies*. 2013

May 22;**4**(2):136–146.

30. Selye H. A syndrome produced by diverse noxious agents. 1936. *J Neuropsychiatry Clin Neurosci*. 1998;**10**(2):230–231.
31. Cannon, W. (1932). *Wisdom of the Body*: Norton & Company.
32. Seeman TE. Price of Adaptation—Allostatic Load and Its Health Consequences: MacArthur Studies of Successful Aging. *Archives of Internal Medicine*. 1997 Oct 27;**157**(19):2259.
33. Barboza Solís C, Kelly-Irving M, Fantin R, et al. Adverse childhood experiences and physiological wear-and-tear in midlife: Findings from the 1958 British birth cohort. *Proc Natl Acad Sci USA*. 2015 Feb 17;**112**(7):E738-746.
34. Muennig P, Sohler N, Mahato B. Socioeconomic status as an independent predictor of physiological biomarkers of cardiovascular disease: evidence from NHANES. *Prev Med*. 2007 Jul;**45**(1):35–40.
35. Robinson O, Carter AR, Ala-Korpela M, et al. Metabolic profiles of socio-economic position: a multi-cohort analysis. *Int J Epidemiol*. 2020 Nov 21;
36. Karimi M, Castagné R, Delpierre C, et al. Early-life inequalities and biological ageing: a multisystem Biological Health Score approach in UnderstandingSociety. *J Epidemiol Community Health*. 2019 Apr 3;jech-2018-212010.
37. Chadeau-Hyam M, Bodinier B, Vermeulen R, et al. Education, biological ageing, all-cause and cause-specific mortality and morbidity: UK biobank cohort study. *EClinicalMedicine* [Internet]. 2020 Nov 19 [cited 2021 Jan 12];**29–30**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7788440/>
38. Maurel M, Castagné R, Berger E, et al. Patterning of educational attainment across inflammatory markers: Findings from a multi-cohort study. *Brain Behav Immun*. 2020 Nov;**90**:303–310.
39. Castagné R, Delpierre C, Kelly-Irving M, et al. A life course approach to explore the biological embedding of socioeconomic position and social mobility through circulating inflammatory markers. *Scientific Reports*. 2016 Apr 27;**6**:25170.
40. Berger E, Castagné R, Chadeau-Hyam M, et al. Multi-cohort study identifies social determinants of systemic inflammation over the life course. *Nat Commun*. 2019 15;**10**(1):773.
41. Scrivo R, Vasile M, Bartosiewicz I, Valesini G. Inflammation as “common soil” of the multifactorial diseases. *Autoimmunity Reviews*. 2011 May;**10**(7):369–374.
42. Berger E, Delpierre C, Hosnijeh FS, et al. Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts. *Sci Rep*. 2018 Jul 17;**8**(1):10805.
43. Guo L, Liu S, Zhang S, et al. C-reactive protein and risk of breast cancer: A systematic

review and meta-analysis. *Sci Rep*. 2015 May 22;**5**:10508.

44. Chan DSM, Bandera EV, Greenwood DC, Norat T. Circulating C-Reactive Protein and Breast Cancer Risk-Systematic Literature Review and Meta-analysis of Prospective Cohort Studies. *Cancer Epidemiol Biomarkers Prev*. 2015 Oct;**24**(10):1439–1449.

45. Gordon-Dseagu VLZ. Cancer and health inequalities: an introduction to current evidence. Unpublished; 2009 [cited 2021 Mar 11]; Available from: <http://rgdoi.net/10.13140/RG.2.2.11025.17766>

46. AIACH (P.), AIACH (P.), BON (N.), DESCHAMPS (J.P.). Comportements et santé. Questions pour la prévention. Nancy: Presses Universitaires; 1992.

47. Global status report on noncommunicable diseases 2010.

48. Hecht SS. Cigarette smoking: cancer risks, carcinogens, and mechanisms. *Langenbecks Arch Surg*. 2006 Nov;**391**(6):603–613.

49. Andler R. NOUVEAU REPÈRE DE CONSOMMATION D'ALCOOL ET USAGE : RÉSULTATS DU BAROMÈTRE DE SANTÉ PUBLIQUE FRANCE 2017 / NEW ALCOHOL CONSUMPTION GUIDELINES AND USE IN FRANCE: RESULTS FROM THE SANTÉ PUBLIQUE FRANCE 2017 HEALTH BAROMETER. :8.

50. Grittner U, Kuntsche S, Gmel G, Bloomfield K. Alcohol consumption and social inequality at the individual and country levels--results from an international study. *Eur J Public Health*. 2013 Apr;**23**(2):332–339.

51. Boffetta P, Hashibe M. Alcohol and cancer. *The Lancet Oncology*. 2006 Feb 1;**7**(2):149–156.

52. Verdot C. CORPULENCE DES ENFANTS ET DES ADULTES EN FRANCE MÉTROPOLITAINE EN 2015. RÉSULTATS DE L'ÉTUDE ESTEBAN ET ÉVOLUTION DEPUIS 2006 / CHILDREN AND ADULTS BODY MASS INDEX IN FRANCE IN 2015. RESULTS OF THE ESTEBAN STUDY AND TRENDS. :8.

53. Percik R, Stumvoll M. Obesity and cancer. *Exp Clin Endocrinol Diabetes*. 2009 Nov;**117**(10):563–566.

54. Boffetta P, Kogevinas M, Westerholm P, Saracci R. Exposure to occupational carcinogens and social class differences in cancer occurrence. *IARC Sci Publ*. 1997;(138):331–341.

55. Loomis D, Guha N, Hall AL, Straif K. Identifying occupational carcinogens: an update from the IARC Monographs. *Occup Environ Med*. 2018 Aug;**75**(8):593–603.

56. Menvielle G, Franck J-E, Radoï L, et al. Quantifying the mediating effects of smoking and occupational exposures in the relation between education and lung cancer: the ICARE study. *Eur J Epidemiol*. 2016 Dec;**31**(12):1213–1221.

57. Menvielle G, Fayossé A, Radoï L, et al. The joint effect of asbestos exposure, tobacco

smoking and alcohol drinking on laryngeal cancer risk: evidence from the French population-based case-control study, ICARE. *Occup Environ Med*. 2016 Jan;**73**(1):28–33.

58. Loomis D, Grosse Y, Lauby-Secretan B, et al. The carcinogenicity of outdoor air pollution. *Lancet Oncol*. 2013 Dec;**14**(13):1262–1263.

59. Hajat A, Hsia C, O'Neill MS. Socioeconomic Disparities and Air Pollution Exposure: a Global Review. *Curr Environ Health Rep*. 2015 Dec;**2**(4):440–450.

60. Gares V, Panico L, Castagne R, Delpierre C, Kelly-Irving M. The role of the early social environment on Epstein Barr virus infection: a prospective observational design using the Millennium Cohort Study. *Epidemiol Infect*. 2017 Dec;**145**(16):3405–3412.

61. Power C, Hyppönen E, Smith GD. Socioeconomic position in childhood and early adult life and risk of mortality: a prospective study of the mothers of the 1958 British birth cohort. *Am J Public Health*. 2005 Aug;**95**(8):1396–1402.

62. Walsh D, McCartney G, Smith M, Armour G. Relationship between childhood socioeconomic position and adverse childhood experiences (ACEs): a systematic review. *J Epidemiol Community Health*. 2019 Dec;**73**(12):1087–1093.

63. Algren MH, Ekholm O, Nielsen L, Ersbøll AK, Bak CK, Andersen PT. Associations between perceived stress, socioeconomic status, and health-risk behaviour in deprived neighbourhoods in Denmark: a cross-sectional study. *BMC Public Health*. 2018 Feb 13;**18**(1):250.

64. Lam JR, Tyler J, Scurrah KJ, Reavley NJ, Dite GS. The Association between Socioeconomic Status and Psychological Distress: A Within and Between Twin Study. *Twin Res Hum Genet*. 2019 Oct;**22**(5):312–320.

65. Abate M, Citro M, Caputo M, Pisanti S, Martinelli R. Psychological Stress and Cancer: New Evidence of An Increasingly Strong Link. *Transl Med UniSa*. 2020 Dec;**23**:53–57.

66. Antoni MH, Lutgendorf SK, Cole SW, et al. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. *Nat Rev Cancer*. 2006;**6**(3):240–248.

67. Gidron Y, Ronson A. Psychosocial factors, biological mediators, and cancer prognosis: a new look at an old story. *Curr Opin Oncol*. 2008 Jul;**20**(4):386–392.

68. Lutgendorf SK, Sood AK, Antoni MH. Host factors and cancer progression: biobehavioral signaling pathways and interventions. *J Clin Oncol*. 2010 Sep 10;**28**(26):4094–4099.

69. Thaker PH, Han LY, Kamat AA, et al. Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma. *Nat Med*. 2006 Aug;**12**(8):939–944.

70. Hassan S, Karpova Y, Baiz D, et al. Behavioral stress accelerates prostate cancer development in mice. *J Clin Invest*. 2013 Feb;**123**(2):874–886.

71. Hermes GL, Delgado B, Tretiakova M, et al. Social isolation dysregulates endocrine and

REFERENCES BIBLIOGRAPHIQUES

behavioral stress while increasing malignant burden of spontaneous mammary tumors. *Proc Natl Acad Sci USA*. 2009 Dec 29;**106**(52):22393–22398.

72. Anda RF, Croft JB, Felitti VJ, et al. Adverse childhood experiences and smoking during adolescence and adulthood. *JAMA*. 1999 Nov 3;**282**(17):1652–1658.

73. Anda RF, Whitfield CL, Felitti VJ, et al. Adverse childhood experiences, alcoholic parents, and later risk of alcoholism and depression. *Psychiatr Serv*. 2002 Aug;**53**(8):1001–1009.

74. Dube SR, Cook ML, Edwards VJ. Health-related outcomes of adverse childhood experiences in Texas, 2002. *Prev Chronic Dis*. 2010 May;**7**(3):A52.

75. Clegg LX, Reichman ME, Miller BA, et al. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. *Cancer Causes Control* [Internet]. 2009 May [cited 2021 Mar 10];**20**(4). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711979/>

76. Kweon S-S, Kim M-G, Kang M-R, Shin M-H, Choi J-S. Difference of stage at cancer diagnosis by socioeconomic status for four target cancers of the National Cancer Screening Program in Korea: Results from the Gwangju and Jeonnam cancer registries. *J Epidemiol*. 2017 Jul;**27**(7):299–304.

77. Nazemi A, Ghodoussipour S, Pearce S, Bhanvadia S, Daneshmand S. Socioeconomic and insurance status are independent prognostic indicators of higher disease stage and worse prognosis in bladder cancer. *Urol Oncol*. 2019 Oct;**37**(10):784–790.

78. Adams J, White M, Forman D. Are there socioeconomic gradients in stage and grade of breast cancer at diagnosis? Cross sectional analysis of UK cancer registry data. *BMJ*. 2004 Jul 17;**329**(7458):142.

79. Percy-Laurry A, Altekruze SF, Hossain MB, O’Keefe AM, Johnson NJ, Kamangar F. Association Between Socioeconomic Status and Tumor Grade Among Black Men with Prostate Cancer. *J Natl Med Assoc*. 2018 Feb;**110**(1):53–57.

80. Li X, Yang J, Peng L, et al. Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Res Treat*. 2017 Jan 1;**161**(2):279–287.

81. Dietze EC, Chavez TA, Seewaldt VL. Obesity and Triple-Negative Breast Cancer: Disparities, Controversies, and Biology. *Am J Pathol*. 2018 Feb;**188**(2):280–290.

82. Blair CK, Wiggins CL, Nibbe AM, et al. Obesity and survival among a cohort of breast cancer patients is partially mediated by tumor characteristics. *NPJ Breast Cancer*. 2019;**5**:33.

83. Haut conseil de la santé publique. Les inégalités sociales de santé : sortir de la fatalité. 2009 Dec;101.

84. Revue de littérature sur les approches explicatives des inégalités de santé en France -

Catherine Dedourge, Christelle Rondeau.

85. Bryere J, Dejardin O, Bouvier V, et al. Socioeconomic environment and cancer incidence: a French population-based study in Normandy. *BMC Cancer*. 2014 Feb 13;**14**:87.
86. Hoebel J, Kroll LE, Fiebig J, et al. Socioeconomic Inequalities in Total and Site-Specific Cancer Incidence in Germany: A Population-Based Registry Study. *Front Oncol*. 2018;**8**:402.
87. Shack L, Jordan C, Thomson CS, Mak V, Møller H, UK Association of Cancer Registries. Variation in incidence of breast, lung and cervical cancer and malignant melanoma of skin by socioeconomic group in England. *BMC Cancer*. 2008 Sep 26;**8**:271.
88. Mihor A, Tomsic S, Zagar T, Lokar K, Zadnik V. Socioeconomic inequalities in cancer incidence in Europe: a comprehensive review of population-based epidemiological studies. *Radiol Oncol*. 2020 Feb 19;**54**(1):1–13.
89. Xie S-H, Lagergren J. Social group disparities in the incidence and prognosis of oesophageal cancer. *United European Gastroenterol J*. 2018 Apr;**6**(3):343–348.
90. Tron L, Belot A, Fauvernier M, et al. Socioeconomic environment and disparities in cancer survival for 19 solid tumor sites: An analysis of the French Network of Cancer Registries (FRANCIM) data. *Int J Cancer*. 2019 Mar 15;**144**(6):1262–1274.
91. Bryere J, Tron L, Menvielle G, Launoy G, French Network of Cancer Registries (FRANCIM). The respective parts of incidence and lethality in socioeconomic differences in cancer mortality. An analysis of the French network Cancer registries (FRANCIM) data. *Int J Equity Health*. 2019 Dec 3;**18**(1):189.
92. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021 Feb 4;
93. Defossez G, Le Guyader-Peyrou S, Uhry Z, Grosclaude P, Colonna M, Dantony E, et al. National estimates of cancer incidence and mortality in metropolitan France between 1990 and 2018. Overview. Saint-Maurice: Santé publique France, 2019. 20 p.
94. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global Cancer in Women: Burden and Trends. *Cancer Epidemiol Biomarkers Prev*. 2017 Apr;**26**(4):444–457.
95. Sobin LH, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours, 7th Edition. Wiley-Blackwell.
96. Stade au diagnostic des cancers du sein, du côlon et du rectum /Étude réalisée à partir des registres des cancers du réseau Francim, France 2018.
97. Perou CM, Sørli T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000 Aug 17;**406**(6797):747–752.
98. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J Natl Cancer Inst*. 2014 Apr 28;**106**(5).

REFERENCES BIBLIOGRAPHIQUES

99. Cortet M, Bertaut A, Molinié F, et al. Trends in molecular subtypes of breast cancer: description of incidence rates between 2007 and 2012 from three French registries. *BMC Cancer*. 2018 Feb 7;**18**(1):161.
100. Danforth DN. Disparities in breast cancer outcomes between Caucasian and African American women: a model for describing the relationship of biological and nonbiological factors. *Breast Cancer Res*. 2013;**15**(3):208.
101. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*. 2001 Oct 27;**358**(9291):1389–1399.
102. Clavel-Chapelon F. Differential effects of reproductive factors on the risk of pre- and postmenopausal breast cancer. Results from a large cohort of French women. *Br J Cancer*. 2002 Mar 4;**86**(5):723–727.
103. Dumitrescu RG, Cotarla I. Understanding breast cancer risk -- where do we stand in 2005? *J Cell Mol Med*. 2005 Mar;**9**(1):208–221.
104. Clavel-Chapelon F, Gerber M. Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? *Breast Cancer Res Treat*. 2002 Mar;**72**(2):107–115.
105. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol*. 2001 Mar;**2**(3):133–140.
106. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet*. 1996 Jun 22;**347**(9017):1713–1727.
107. Diet, nutrition, physical activity and breast cancer 2017. 2017;120.
108. Alcohol, tobacco and breast cancer – collaborative reanalysis of individual data from 53 epidemiological studies, including 58 515 women with breast cancer and 95 067 women without the disease. *Br J Cancer*. 2002 Nov 18;**87**(11):1234–1245.
109. Jones ME, Schoemaker MJ, Wright LB, Ashworth A, Swerdlow AJ. Smoking and risk of breast cancer in the Generations Study cohort. *Breast Cancer Res* [Internet]. 2017 [cited 2021 Apr 8];**19**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5698948/>
110. Dossus L, Boutron-Ruault M-C, Kaaks R, et al. Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. *Int J Cancer*. 2014 Apr 15;**134**(8):1871–1888.
111. McCormack VA, Santos Silva I dos. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2006 Jun;**15**(6):1159–1169.
112. McPherson K, Steel CM, Dixon JM. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *BMJ*. 2000 Sep 9;**321**(7261):624–628.

113. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. *Biochim Biophys Acta*. 2015 Aug;**1856**(1):73–85.
114. Yang XR, Chang-Claude J, Goode EL, et al. Associations of Breast Cancer Risk Factors With Tumor Subtypes: A Pooled Analysis From the Breast Cancer Association Consortium Studies. *JNCI Journal of the National Cancer Institute*. 2011 Feb 2;**103**(3):250.
115. Millikan RC, Newman B, Tse C-K, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat*. 2008 May;**109**(1):123–139.
116. Schneider BP, Winer EP, Foulkes WD, et al. Triple-negative breast cancer: risk factors to potential targets. *Clin Cancer Res*. 2008 Dec 15;**14**(24):8010–8018.
117. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev*. 2007 Mar;**16**(3):439–443.
118. Vona-Davis L, Rose DP. The influence of socioeconomic disparities on breast cancer tumor biology and prognosis: a review. *J Womens Health (Larchmt)*. 2009 Jun;**18**(6):883–893.
119. Hvidtfeldt UA. Mechanisms underlying social inequality in post- menopausal breast cancer. :26.
120. Lundqvist A, Andersson E, Ahlberg I, Nilbert M, Gerdtham U. Socioeconomic inequalities in breast cancer incidence and mortality in Europe-a systematic review and meta-analysis. *Eur J Public Health*. 2016 Oct;**26**(5):804–813.
121. Hussain SK, Altieri A, Sundquist J, Hemminki K. Influence of education level on breast cancer risk and survival in Sweden between 1990 and 2004. *Int J Cancer*. 2008 Jan 1;**122**(1):165–169.
122. Pudrovska T, Anikputa B. The Role of Early-Life Socioeconomic Status in Breast Cancer Incidence and Mortality: Unraveling Life Course Mechanisms. *J Aging Health*. 2012 Mar;**24**(2):323–344.
123. Villeneuve S, Févotte J, Anger A, et al. Breast cancer risk by occupation and industry: analysis of the CECILE study, a population-based case-control study in France. *Am J Ind Med*. 2011 Jul;**54**(7):499–509.
124. Danø H, Hansen KD, Jensen P, et al. Fertility pattern does not explain social gradient in breast cancer in denmark. *Int J Cancer*. 2004 Sep 1;**111**(3):451–456.
125. Carlsen K, Høybye MT, Dalton SO, Tjønneland A. Social inequality and incidence of and survival from breast cancer in a population-based study in Denmark, 1994-2003. *Eur J Cancer*. 2008 Sep;**44**(14):1996–2002.
126. Linden BWA van der, Courvoisier DS, Cheval B, et al. Effect of childhood socioeconomic conditions on cancer onset in later life: an ambidirectional cohort study. *Int J Public Health*. 2018;**63**(7):799–810.

REFERENCES BIBLIOGRAPHIQUES

127. Coughlin SS. Social determinants of breast cancer risk, stage, and survival. *Breast Cancer Res Treat.* 2019 Oct 1;**177**(3):537–548.
128. Antonova L, Aronson K, Mueller CR. Stress and breast cancer: from epidemiology to molecular biology. *Breast Cancer Res.* 2011;**13**(2):208.
129. Bahri N, Fathi Najafi T, Homaei Shandiz F, Tohidinik HR, Khajavi A. The relation between stressful life events and breast cancer: a systematic review and meta-analysis of cohort studies. *Breast Cancer Res Treat.* 2019 Jul 1;**176**(1):53–61.
130. Duijts SFA, Zeegers MPA, Borne BV. The association between stressful life events and breast cancer risk: a meta-analysis. *Int J Cancer.* 2003 Dec 20;**107**(6):1023–1029.
131. Touvier M, Fassier P, His M, et al. Cholesterol and breast cancer risk: a systematic review and meta-analysis of prospective studies. *Br J Nutr.* 2015 Aug 14;**114**(3):347–357.
132. Ni H, Liu H, Gao R. Serum Lipids and Breast Cancer Risk: A Meta-Analysis of Prospective Cohort Studies. *PLoS One* [Internet]. 2015 Nov 10 [cited 2021 Jan 12];**10**(11). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640529/>
133. Han H, Guo W, Shi W, et al. Hypertension and breast cancer risk: a systematic review and meta-analysis. *Sci Rep* [Internet]. 2017 Mar 20 [cited 2021 Jan 12];**7**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5357949/>
134. Seretis A, Cividini S, Markozannes G, et al. Association between blood pressure and risk of cancer development: a systematic review and meta-analysis of observational studies. *Sci Rep.* 2019 Jun 12;**9**(1):8565.
135. Cauley JA, Lucas FL, Kuller LH, Stone K, Browner W, Cummings SR. Elevated Serum Estradiol and Testosterone Concentrations Are Associated with a High Risk for Breast Cancer. *Ann Intern Med.* American College of Physicians; 1999 Feb 16;**130**(4_Part_1):270–277.
136. Hankinson SE, Eliassen AH. Endogenous estrogen, testosterone and progesterone levels in relation to breast cancer risk. *The Journal of Steroid Biochemistry and Molecular Biology.* 2007 Aug 1;**106**(1):24–30.
137. James RE, Lukanova A, Dossus L, et al. Postmenopausal serum sex steroids and risk of hormone receptor-positive and -negative breast cancer: a nested case-control study. *Cancer Prev Res (Phila).* 2011 Oct;**4**(10):1626–1635.
138. Key T, Appleby P, Barnes I, Reeves G, Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst.* 2002 Apr 17;**94**(8):606–616.
139. Sieri S, Krogh V, Bolelli G, et al. Sex Hormone Levels, Breast Cancer Risk, and Cancer Receptor Status in Postmenopausal Women: the ORDET Cohort. *Cancer Epidemiology Biomarkers & Prevention.* 2009 Jan 1;**18**(1):169–176.
140. Zeleniuch-Jacquotte A, Bruning PF, Bonfrer JM, et al. Relation of serum levels of testosterone and dehydroepiandrosterone sulfate to risk of breast cancer in postmenopausal

women. *Am J Epidemiol*. 1997 Jun 1;**145**(11):1030–1038.

141. Zhang X, Tworoger SS, Eliassen AH, Hankinson SE. Postmenopausal plasma sex hormone levels and breast cancer risk over 20 years of follow-up. *Breast Cancer Res Treat*. 2013 Feb;**137**(3):883–892.

142. He XY, Liao YD, Yu S, Zhang Y, Wang R. Sex hormone binding globulin and risk of breast cancer in postmenopausal women: a meta-analysis of prospective studies. *Horm Metab Res*. 2015 Jun;**47**(7):485–490.

143. Orsini M, Trétarre B, Daurès J-P, Bessaoud F. Individual socioeconomic status and breast cancer diagnostic stages: a French case-control study. *Eur J Public Health*. 2016;**26**(3):445–450.

144. Berger F, Doussau A, Gautier C, Gros F, Asselain B, Reyal F. [Impact of socioeconomic status on stage at diagnosis of breast cancer]. *Rev Epidemiol Sante Publique*. 2012 Feb;**60**(1):19–29.

145. Gentil-Brevet J, Colonna M, Danzon A, et al. The influence of socio-economic and surveillance characteristics on breast cancer survival: a French population-based study. *Br J Cancer*. 2008 Jan 15;**98**(1):217–224.

146. Feller A, Schmidlin K, Bordoni A, et al. Socioeconomic and demographic disparities in breast cancer stage at presentation and survival: A Swiss population-based study. *Int J Cancer*. 2017 Oct 15;**141**(8):1529–1539.

147. Riba LA, Gruner RA, Alapati A, James TA. Association between socioeconomic factors and outcomes in breast cancer. *The Breast Journal*. 2019;**25**(3):488–492.

148. Taylor A, Cheng KK. Social deprivation and breast cancer. *J Public Health Med*. 2003 Sep;**25**(3):228–233.

149. DeSantis C, Jemal A, Ward E. Disparities in breast cancer prognostic factors by race, insurance status, and education. *Cancer Causes Control*. 2010 Sep;**21**(9):1445–1450.

150. McKenzie F, Jeffreys M, Mannetje A 't, Pearce N. Prognostic factors in women with breast cancer: inequalities by ethnicity and socioeconomic position in New Zealand. *Cancer Causes Control*. 2008 May;**19**(4):403–411.

151. Ag C, A S, Dw L, et al. Relation between socioeconomic deprivation and pathological prognostic factors in women with breast cancer. *BMJ (Clinical research ed)* [Internet]. BMJ; 1994 Oct 22 [cited 2021 Apr 6];**309**(6961). Available from: <https://pubmed-ncbi-nlm-nih-gov.proxy.insermbiblio.inist.fr/7950739/>

152. Thomson CS, Hole DJ, Twelves CJ, Brewster DH, Black RJ, Scottish Cancer Therapy Network. Prognostic factors in women with breast cancer: distribution by socioeconomic status and effect on differences in survival. *J Epidemiol Community Health*. 2001 May;**55**(5):308–315.

REFERENCES BIBLIOGRAPHIQUES

153. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer*. 2007 May 1;**109**(9):1721–1728.
154. Auguste A, Cortet M, Dabakuyo-Yonli TS, et al. Breast cancer subtype of French women is not influenced by socioeconomic status: A population-based-study. *PLoS ONE*. 2017;**12**(2):e0170069.
155. Akinyemiju TF, Pisu M, Waterbor JW, Altekruse SF. Socioeconomic status and incidence of breast cancer by hormone receptor subtype. *Springerplus* [Internet]. 2015 Sep 17 [cited 2020 May 4];**4**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4573746/>
156. Prakash O, Hossain F, Danos D, Lassak A, Scribner R, Miele L. Racial Disparities in Triple Negative Breast Cancer: A Review of the Role of Biologic and Non-biologic Factors. *Front Public Health*. 2020;**8**:576964.
157. Siddharth S, Sharma D. Racial Disparity and Triple-Negative Breast Cancer in African-American Women: A Multifaceted Affair between Obesity, Biology, and Socioeconomic Determinants. *Cancers (Basel)*. 2018 Dec 14;**10**(12).
158. Rauscher GH, Campbell RT, Wiley EL, Hoskins K, Stolley MR, Warnecke RB. Mediation of Racial and Ethnic Disparities in Estrogen/Progesterone Receptor–Negative Breast Cancer by Socioeconomic Position and Reproductive Factors. *Am J Epidemiol*. 2016 May 15;**183**(10):884–893.
159. Dalton SO, Düring M, Ross L, et al. The relation between socioeconomic and demographic factors and tumour stage in women diagnosed with breast cancer in Denmark, 1983-1999. *Br J Cancer*. 2006 Sep 4;**95**(5):653–659.
160. Xing CY, Doose M, Qin B, et al. Prediagnostic Allostatic Load as a Predictor of Poorly Differentiated and Larger Sized Breast Cancers among Black Women in the Women’s Circle of Health Follow-Up Study. *Cancer Epidemiol Biomarkers Prev*. 2020;**29**(1):216–224.
161. Kaffashian F, Godward S, Davies T, Solomon L, McCann J, Duffy SW. Socioeconomic effects on breast cancer survival: proportion attributable to stage and morphology. *British Journal of Cancer*. Nature Publishing Group; 2003 Nov;**89**(9):1693–1696.
162. Yu XQ. Socioeconomic disparities in breast cancer survival: relation to stage at diagnosis, treatment and race. *BMC Cancer*. 2009 Oct 14;**9**:364.
163. Bastiaannet E, Craen AJM, Kuppen PJK, et al. Socioeconomic differences in survival among breast cancer patients in the Netherlands not explained by tumor size. *Breast Cancer Research and Treatment*. Springer Verlag; 2010 Nov;**127**(3):721–727.
164. Howlader N, Cronin KA, Kurian AW, Andridge R. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiol Biomarkers Prev*. American Association for Cancer Research; 2018 Jun 1;**27**(6):619–626.

165. Fisher ER, Redmond CK, Liu H, Rockette H, Fisher B. Correlation of estrogen receptor and pathologic characteristics of invasive breast cancer. *Cancer*. 1980 Jan 15;**45**(2):349–353.
166. Walsh PM, Byrne J, Kelly M, McDevitt J, Comber H. Socioeconomic disparity in survival after breast cancer in Ireland: observational study. *PLoS One*. 2014;**9**(11):e111729.
167. Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*. 1997;**26 Suppl 1**:S6-14.
168. Clavel-Chapelon F. Cohort Profile: The French E3N Cohort Study. *Int J Epidemiol*. 2015 Jun 1;**44**(3):801–809.
169. ICD-10: International statistical classification of diseases and related health problems. Geneva: World Health Organization; 2011.
170. Fritz AG, editor. International classification of diseases for oncology: ICD-O. Third edition, First revision. Geneva: World Health Organization; 2013.
171. Errico A d', Ricceri F, Stringhini S, et al. Socioeconomic indicators in epidemiologic research: A practical example from the LIFEPAATH study. *PLoS ONE*. 2017;**12**(5):e0178071.
172. Rose D, Harrison E. The European Socio-Economic Classification: A New Social Class Schema for Comparative European Research. *European Societies*. Routledge; 2007 Jul 1;**9**(3):459–490.
173. Dossus L, Jimenez-Corona A, Romieu I, et al. C-reactive protein and postmenopausal breast cancer risk: results from the E3N cohort study. *Cancer Causes Control*. 2014 Apr 1;**25**(4):533–539.
174. His M, Dartois L, Fagherazzi G, et al. Associations between serum lipids and breast cancer incidence and survival in the E3N prospective cohort study. *Cancer Causes Control*. 2017 Jan;**28**(1):77–88.
175. Engel P, Fagherazzi G, Boutten A, et al. Serum 25(OH) vitamin D and risk of breast cancer: a nested case-control study from the French E3N cohort. *Cancer Epidemiol Biomarkers Prev*. 2010 Sep;**19**(9):2341–2350.
176. Juster R-P, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*. 2010 Sep;**35**(1):2–16.
177. UK Chief Medical Officers' Low Risk Drinking Guidelines. :11.
178. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000 Sep;**32**(9 Suppl):S498-504.
179. Bousquet VH. Traitement des données manquantes en épidémiologie : application de l'imputation multiple à des données de surveillance et d'enquêtes. :339.
180. RUBIN DB. Inference and missing data. *Biometrika*. 1976 Dec 1;**63**(3):581–592.

REFERENCES BIBLIOGRAPHIQUES

181. Cox DR. Regression Models and Life-Tables. 1972;35.
182. Multinomial Logistic Regression - Starkweather J and Kay Moske A, 2011.
183. Santos Silva I dos, Beral V. Socioeconomic differences in reproductive behaviour. *IARC Sci Publ.* 1997;(138):285–308.
184. Cavelaars AE, Kunst AE, Geurts JJ, et al. Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Ann Hum Biol.* 2000 Aug;27(4):407–421.
185. Barriuso L, Miqueleiz E, Albaladejo R, Villanueva R, Santos JM, Regidor E. Socioeconomic position and childhood-adolescent weight status in rich countries: a systematic review, 1990-2013. *BMC Pediatr.* 2015 Sep 21;15:129.
186. Petrovic D, Mestral C de, Bochud M, et al. The contribution of health behaviors to socioeconomic inequalities in health: A systematic review. *Prev Med.* 2018;113:15–31.
187. Martikainen P, Brunner E, Marmot M. Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc Sci Med.* 2003 Apr;56(7):1397–1410.
188. Larsen SB, Olsen A, Lynch J, et al. Socioeconomic position and lifestyle in relation to breast cancer incidence among postmenopausal women: A prospective cohort study, Denmark, 1993–2006. *Cancer Epidemiology.* 2011 Oct;35(5):438–441.
189. Menvielle G, Kunst AE, Gils CH van, et al. The contribution of risk factors to the higher incidence of invasive and in situ breast cancers in women with higher levels of education in the European prospective investigation into cancer and nutrition. *Am J Epidemiol.* 2011 Jan 1;173(1):26–37.
190. Braaten T, Weiderpass E, Kumle M, Lund E. Explaining the socioeconomic variation in cancer risk in the Norwegian Women and Cancer Study. *Cancer Epidemiol Biomarkers Prev.* 2005 Nov;14(11 Pt 1):2591–2597.
191. Heck KE, Pamuk ER. Explaining the relation between education and postmenopausal breast cancer. *Am J Epidemiol.* 1997 Feb 15;145(4):366–372.
192. Beiki O, Hall P, Ekblom A, Moradi T. Breast cancer incidence and case fatality among 4.7 million women in relation to social and ethnic background: a population-based cohort study. *Breast Cancer Res.* 2012 Jan 6;14(1):R5.
193. Meijer M, Bloomfield K, Engholm G. Neighbourhoods matter too: the association between neighbourhood socioeconomic position, population density and breast, prostate and lung cancer incidence in Denmark between 2004 and 2008. *J Epidemiol Community Health.* 2013 Jan;67(1):6–13.
194. Rafique M, Zia S, Sultan MA. Impact of Socioeconomic Status on Birth Weight and Length of Newborns Delivered at Services Hospital Lahore. :9.
195. Khalid H, Khawar K, Fawad M, et al. Age of Menarche in Relation to Socioeconomic

Status, BMI, Physical Activity and Stress Among High School Girls. :7.

196. Schoenaker DA, Jackson CA, Rowlands JV, Mishra GD. Socioeconomic position, lifestyle factors and age at natural menopause: a systematic review and meta-analyses of studies across six continents. *Int J Epidemiol*. 2014 Oct;**43**(5):1542–1562.
197. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 2011 Mar 4;**144**(5):646–674.
198. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008 Jul 24;**454**(7203):436–444.
199. Fortunati N, Catalano MG, Boccuzzi G, Frairia R. Sex Hormone-Binding Globulin (SHBG), estradiol and breast cancer. *Mol Cell Endocrinol*. 2010 Mar 5;**316**(1):86–92.
200. Inoue M, Noda M, Kurahashi N, et al. Impact of metabolic factors on subsequent cancer risk: results from a large-scale population-based cohort study in Japan. *Eur J Cancer Prev*. 2009 Jun;**18**(3):240–247.
201. Martin LJ, Melnichouk O, Huszti E, et al. Serum lipids, lipoproteins, and risk of breast cancer: a nested case-control study using multiple time points. *J Natl Cancer Inst*. 2015 May;**107**(5).
202. Johnson KE, Siewert KM, Klarin D, et al. The relationship between circulating lipids and breast cancer risk: A Mendelian randomization study. *PLoS Med* [Internet]. 2020 Sep 11 [cited 2021 Jan 12];**17**(9). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7485834/>
203. Tjugen TB, Flaa A, Kjeldsen SE. The prognostic significance of heart rate for cardiovascular disease and hypertension. *Curr Hypertens Rep*. 2010 Jun;**12**(3):162–169.
204. Endogenous Hormones and Breast Cancer Collaborative Group, Key TJ, Appleby PN, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer*. 2011 Aug 23;**105**(5):709–722.
205. Nazmi A, Victora CG. Socioeconomic and racial/ethnic differentials of C-reactive protein levels: a systematic review of population-based studies. *BMC Public Health*. 2007 Aug 17;**7**:212.
206. Bann D, Hardy R, Cooper R, et al. Socioeconomic conditions across life related to multiple measures of the endocrine system in older adults: Longitudinal findings from a British birth cohort study. *Soc Sci Med*. 2015 Dec;**147**:190–199.
207. Cappola AR, Ratcliffe SJ, Bhasin S, et al. Determinants of serum total and free testosterone levels in women over the age of 65 years. *J Clin Endocrinol Metab*. 2007 Feb;**92**(2):509–516.
208. Kische H, Gross S, Wallaschofski H, et al. Clinical correlates of sex hormones in women: The study of health in Pomerania. *Metabolism*. 2016 Sep;**65**(9):1286–1296.

REFERENCES BIBLIOGRAPHIQUES

209. Bleil ME, Appelhans BM, Latham MD, et al. Neighborhood Socioeconomic Status During Childhood Versus Puberty in Relation to Endogenous Sex Hormone Levels in Adult Women. *Nurs Res*. 2015;**64**(3):211–220.
210. Heald A, Laing I, McLernon DJ, et al. Socioeconomic Deprivation as Measured by the Index of Multiple Deprivation and Its Association with Low Sex Hormone Binding Globulin in Women. *Open Biochem J*. 2017;**11**:1–7.
211. Houghton LC, Ganmaa D, Rosenberg PS, et al. Associations of Breast Cancer Risk Factors with Premenopausal Sex Hormones in Women with Very Low Breast Cancer Risk. *Int J Environ Res Public Health* [Internet]. 2016 Nov [cited 2020 Oct 5];**13**(11). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5129276/>
212. Lyratzopoulos G, Abel GA, Brown CH, et al. Socio-demographic inequalities in stage of cancer diagnosis: evidence from patients with female breast, lung, colon, rectal, prostate, renal, bladder, melanoma, ovarian and endometrial cancer. *Ann Oncol*. 2013 Mar;**24**(3):843–850.
213. Flamant C, Gauthier E, Clavel-Chapelon F. Determinants of non-compliance to recommendations on breast cancer screening among women participating in the French E3N cohort study. *Eur J Cancer Prev*. 2006 Feb;**15**(1):27–33.
214. Andaya AA, Enewold L, Horner M-J, Jatoi I, Shriver CD, Zhu K. Socioeconomic disparities and breast cancer hormone receptor status. *Cancer Causes Control*. 2012 Jun;**23**(6):951–958.
215. Samuels L, Harkness E, Astley SM, et al. The Relationship of Volumetric Breast Density to Socio-Economic Status in a Screening Population. In: Fujita H, Hara T, Muramatsu C, editors. *Breast Imaging*. Cham: Springer International Publishing; 2014. p. 273–281.
216. Aitken Z, Walker K, Stegeman BH, et al. Mammographic density and markers of socioeconomic status: a cross-sectional study. *BMC Cancer*. 2010 Feb 9;**10**:35.
217. Guibout C, Prisse N, Clavel-Chapelon F, E3N-EPIC. [Development of a bio-library in the cohort survey: E3N-EPIC]. *Rev Epidemiol Sante Publique*. 2003 Feb;**51**(1 Pt 2):137–141.
218. Russo J, Mailo D, Hu Y-F, Balogh G, Sheriff F, Russo IH. Breast differentiation and its implication in cancer prevention. *Clin Cancer Res*. 2005 Jan 15;**11**(2 Pt 2):931s–6s.
219. Colditz GA, Bohlke K, Berkey CS. Breast cancer risk accumulation starts early – Prevention must also. *Breast Cancer Res Treat*. 2014 Jun;**145**(3):567–579.
220. Ogony JW, Radisky DC, Ruddy KJ, et al. Immune Responses and Risk of Triple-negative Breast Cancer: Implications for Higher Rates among African American Women. *Cancer Prev Res (Phila)*. 2020 Nov;**13**(11):901–910.
221. Baumgarten SC, Frasar J. Minireview: Inflammation: an instigator of more aggressive estrogen receptor (ER) positive breast cancers. *Mol Endocrinol*. 2012 Mar;**26**(3):360–371.
222. Rojas-Lima E, Gamboa-Loira B, Cebrián ME, Rothenberg SJ, López-Carrillo L. A cumulative index of exposure to endogenous estrogens and breast cancer by molecular

- subtypes in northern Mexican women. *Breast Cancer Res Treat.* 2020 Apr;**180**(3):791–800.
223. Murray JI, West NR, Murphy LC, Watson PH. Intratumoural inflammation and endocrine resistance in breast cancer. *Endocr Relat Cancer.* 2015 Feb;**22**(1):R51-67.
224. Toufexis D, Rivarola MA, Lara H, Viau V. Stress and the reproductive axis. *J Neuroendocrinol.* 2014 Sep;**26**(9):573–586.
225. Tsuchimine S, Matsuno H, O’Hashi K, et al. Comparison of physiological and behavioral responses to chronic restraint stress between C57BL/6J and BALB/c mice. *Biochem Biophys Res Commun.* 2020 Feb 15;
226. Powell LH, Lovallo WR, Matthews KA, et al. Physiologic markers of chronic stress in premenopausal, middle-aged women. *Psychosom Med.* 2002 Jun;**64**(3):502–509.
227. Dismukes AR, Johnson MM, Vitacco MJ, Iturri F, Shirtcliff EA. Coupling of the HPA and HPG axes in the context of early life adversity in incarcerated male adolescents. *Dev Psychobiol.* 2015 Sep;**57**(6):705–718.
228. Shirtcliff EA, Dismukes AR, Marceau K, Ruttle PL, Simmons JG, Han G. A dual-axis approach to understanding neuroendocrine development. *Dev Psychobiol.* 2015 Sep;**57**(6):643–653.
229. Marceau K, Shirtcliff EA, Hastings PD, et al. Within-adolescent coupled changes in cortisol with DHEA and testosterone in response to three stressors during adolescence. *Psychoneuroendocrinology.* 2014 Mar;**41**:33–45.
230. Phan JM, Schneider E, Peres J, Miocevic O, Meyer V, Shirtcliff EA. Social Evaluative Threat with Verbal Performance Feedback Alters Neuroendocrine Response to Stress. *Horm Behav.* 2017 Nov;**96**:104–115.
231. Deuter CE, Duesenberg M, Hellmann-Regen J, et al. Psychosocial stress increases testosterone in patients with borderline personality disorder, post-traumatic stress disorder and healthy participants. *Borderline Personal Disord Emot Dysregul* [Internet]. 2021 Feb 1 [cited 2021 Mar 25];**8**. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7849084/>
232. Lennartsson A-K, Kushnir MM, Bergquist J, Billig H, Jonsdottir IH. Sex steroid levels temporarily increase in response to acute psychosocial stress in healthy men and women. *Int J Psychophysiol.* 2012 Jun;**84**(3):246–253.
233. Pletzer B, Poppelaars ES, Klackl J, Jonas E. The gonadal response to social stress and its relationship to cortisol. *Stress.* 2021 Mar 12;1–10.
234. Zajacova A, Lawrence EM. The Relationship Between Education and Health: Reducing Disparities Through a Contextual Approach. *Annu Rev Public Health.* 2018 Apr 1;**39**:273–289.
235. Chandola T, Clarke P, Morris JN, Blane D. Pathways between education and health: a causal modelling approach. *Journal of the Royal Statistical Society: Series A (Statistics in*

REFERENCES BIBLIOGRAPHIQUES

Society). 2006;**169**(2):337–359.

236. Recommandations 3ème plan cancer [Internet]. [cited 2021 Mar 25]. Available from: <https://www.asdugrandlyon.com/wp-content/uploads/2014/02/recommandations-3eme-plan-cancer.pdf>

237. Ott M-O, Quéré M, Romulus A-M. Evaluation du 3eme plan cancer - rapport de synthèse. :73.

ANNEXES

ANNEXE 1 - DETERMINANTS SOCIAUX ET INFLAMMATION: ETUDE
DANS SIX COHORTES EUROPEENNES

ARTICLE

<https://doi.org/10.1038/s41467-019-08732-x>

OPEN

Multi-cohort study identifies social determinants of systemic inflammation over the life course

Eloïse Berger¹, Raphaële Castagné¹, Marc Chadeau-Hyam², Murielle Bochud³, Angelo d'Errico⁴, Martina Gandini⁴, Maryam Karimi², Mika Kivimäki^{5,6}, Vittorio Krogh⁷, Michael Marmot⁵, Salvatore Panico⁸, Martin Preisig³, Fulvio Ricceri⁴, Carlotta Sacerdote⁹, Andrew Steptoe⁵, Silvia Stringhini¹⁰, Rosario Tumino¹¹, Paolo Vineis^{2,12}, Cyrille Delpierre¹ & Michelle Kelly-Irving¹

Chronic inflammation has been proposed as having a prominent role in the construction of social inequalities in health. Disentangling the effects of early life and adulthood social disadvantage on inflammation is key in elucidating biological mechanisms underlying socio-economic disparities. Here we explore the relationship between socioeconomic position (SEP) across the life course and inflammation (as measured by CRP levels) in up to 23,008 participants from six European cohort studies from three countries conducted between 1958 and 2013. We find a consistent inverse association between SEP and CRP across cohorts, where participants with a less advantaged SEP have higher levels of inflammation. Educational attainment is most strongly related to inflammation, after adjusting for health behaviours, body mass index and later-in-life SEP. These findings suggest socioeconomic disadvantage in young adulthood is independently associated with later life inflammation calling for further studies of the pathways operating through educational processes.

¹ LEASP, UMR 1027, Inserm-Université Toulouse III Paul Sabatier, Toulouse 31000, France. ² MRC-PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. ³ Psychiatric Epidemiology and Psychopathology Center, Department of Psychiatry, Lausanne University Hospital, Lausanne 1004, Switzerland. ⁴ Epidemiology Unit, ASL TO3 Piedmont Region, Grugliasco 10095, Italy. ⁵ Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK. ⁶ Clinicum, Faculty of Medicine, University of Helsinki, P. O. Box 20 Helsinki FI-00014, Finland. ⁷ Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan 20133, Italy. ⁸ Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples 80131, Italy. ⁹ Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin 10126, Italy. ¹⁰ Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne 1010, Switzerland. ¹¹ Cancer Registry and Histopathology Department, 'Civic - M. P. Arezzo' Hospital, ASP Ragusa, Ragusa 97100, Italy. ¹² Molecular and Genetic Epidemiology Unit, Italian Institute for Genomic Medicine (IIGM), Torino 10126, Italy. These authors contributed equally: Eloïse Berger, Raphaële Castagné. These authors jointly supervised: Paolo Vineis, Cyrille Delpierre, Michelle Kelly-Irving. Correspondence and requests for materials should be addressed to M. K.-I. (email: michelle.kelly@inserm.fr)

Heightened systemic inflammation has been linked to many chronic diseases¹. Inflammation is a set of responses that may be caused by a number of processes such as the presence of an infection, central-adipose tissue, or tumour cell development, but also occur as a consequence of the chronic solicitation of the stress response system, namely neuroinflammation or sterile inflammation^{2,3}. A higher level of basal inflammation has consequences for overall health and has been linked to mortality across various causes^{4,5}. CRP is an acute-phase protein synthesized by the liver in response to systemic effects of inflammation⁶ and is generally considered a marker of overall inflammatory response. CRP has been considered as a marker of atherosclerosis and coronary heart disease, playing an important role in the formation of plaques in arterial walls and triggering of a cardiac or cerebrovascular event⁷. Prospective epidemiologic studies suggest that elevated circulating levels of CRP are also associated with an increased risk of certain types (lung, colorectal), but not all cancers⁸.

Underlying the aetiological drivers of inflammation implicated in ageing and chronic diseases is the backdrop of the socio-economic environment. A number of studies highlight socio-economic disadvantage as an upstream determinant of increased basal inflammation. A systematic review of 25 population-based studies reported that low SEP mainly assessed by education was associated with elevated CRP level in adulthood across countries⁹. Elevated levels of others circulating inflammatory markers were also reported in disadvantaged socioeconomic groups in general^{10–13} and also regarding gender differences¹⁴. A recent meta-analysis of 15 studies focusing on SEP in childhood revealed an inverse association between early life SEP through parental education or occupation and adulthood CRP¹⁵. These relationships exemplified the biological embodiment, or the social-to-biological transition¹⁶ which could be an important set of processes and mechanisms involved in the construction of health inequalities over the life course. Several processes can explain the impact of the social environment on inflammation. First, socially disadvantaged populations are disproportionately exposed to environments that can be characterized as pro-inflammatory¹⁷. This includes exposure to infections due to overcrowded conditions, poor housing quality, or insufficient access to sanitation¹⁸. Second, adverse health behaviours, more prevalent among disadvantaged socioeconomic groups, may expose individuals to pro-inflammatory factors, such as tobacco smoking as well as nutritional behaviours leading towards central obesity^{19,20}. Finally, cumulative social disadvantage may also lead individuals towards experiencing adversities, or exacerbate such situations, and result in an over-solicited stress response system, which in turn, contribute to heightened basal inflammation^{21,22}.

Social exposures occur from childhood and across life course stages consisting of both biological and socially sensitive periods. They may lead to early wear-and-tear of physiological systems and ultimately to worsened health^{23–25}. In the available literature on the relationship between SEP and inflammation, the influence played by country and period-specific contexts on the social patterning of inflammatory response has been given limited attention due to the lack of available data and/or cross-country variable harmonisation. Investigating the temporal nature of social exposures over the life course and the inflammatory response later in life also needs to be better elucidated. Unpicking these questions may highlight mechanisms through which the socioeconomic gradient in health emerges over the life course, how to prevent it from emerging, and how to mitigate the effects of disadvantage on inflammatory processes underlying morbidity.

This study aims to examine several important aspects of the social-to-biological transition using harmonised data across six European cohort studies collected within the Lifepath consortium

(Supplementary Note 1)²⁶. Our approach consists of taking into account the chronology of exposures over the life course, to understand how they are associated with inflammation. First, we assessed the relationship between life course SEP (father's occupation, educational attainment, participant's last occupation) at three time points individually and systemic inflammation using the biomarker C reactive protein (CRP) measured in adulthood. We then investigated the potential impact of behavioural factors and body mass index (BMI) on this relationship. Second we investigated life course effects of SEP experiences by sequentially controlling for time-ordered SEP. To gain a better understanding of the gender and cohort effects, analyses were performed separately within each cohort for men and women. Cohort-specific estimates were combined using a random effect meta-analysis.

Results

Study population. The sample selection for each cohort is provided in Fig. 1 along with a description of the main participants characteristics by cohort in Table 1. Key characteristics of each study population showed small to moderate differences compared to our analytical sample (Supplementary Data 1). Complete data on CRP, two life course SEP variables (educational attainment and last occupation) as well as intermediate factors were available for 23,008 [55.7% of men] participants across the 6 cohorts and 13,078 [59.6% of men] participants when restricting to the 4 cohorts where father's occupation was available. Mean age ranged between 45.8 (SD 1) years of age in NCDS and 67.4 (SD 9.4) in ELSA. Between 46.8% (NCDS) and 51.7% (ELSA) of participants were women except for Whitehall II (29.2%). Mean serum CRP levels (SD) ranged from 1.9 (4.3) mg/L in Whitehall II to 4.0 (7.8) in ELSA. The proportion of participants with a low educational attainment ranged from 38.8% in Whitehall II to 76.3% in NCDS. The largest proportion of participants that reported drinking more than 21 (men) or 14 (women) alcoholic units per week was observed in NCDS (26.8%); additionally, EPIC-Italy had the largest proportion of smokers (29.6%), 28.1% of the participants from ELSA were obese and 40.6% of Skipogh participants were sedentary.

Descriptive characteristics for each study by SEP are shown in Supplementary Data 2. Variation was observed in participant's distribution in terms of their characteristics, by life course SEP and by study. Despite this, in general participants with a less advantaged SEP were more likely to smoke (except in EPIC-Italy) and systematically more likely to be overweight/obese and sedentary (except for EPIC-Italy) compared to the most advantaged. Alcohol consumption followed a heterogeneous pattern both by study and SEP.

Association between each respective SEP indicator and CRP.

Results for childhood SEP (measured by father's occupational position) are given in Table 2A and Fig. 2a. Despite the moderate to high degree of heterogeneity in study-specific estimates, results from meta-analysis revealed a significant association between less advantaged childhood SEP and higher CRP level in adulthood in the overall population (Model 1: Less advantaged vs More advantaged $\beta = 0.19$, $P < 0.001$) Controlling for either alcohol consumption, smoking or sedentary lifestyle had little effect on the observed association, whereas adjustment for BMI attenuated associations between SEP and CRP (Model 1 + BMI: Less advantaged vs More advantaged $\beta = 0.10$, $P = 0.005$) leading to a coefficient attenuation of 47.4% (Supplementary Note 2). Controlling for all intermediate factors in adulthood explained part but not all of the observed association between childhood SEP and CRP level in adulthood which remained significant (Model 2: Less advantaged vs More advantaged $\beta = 0.08$, $P = 0.021$).

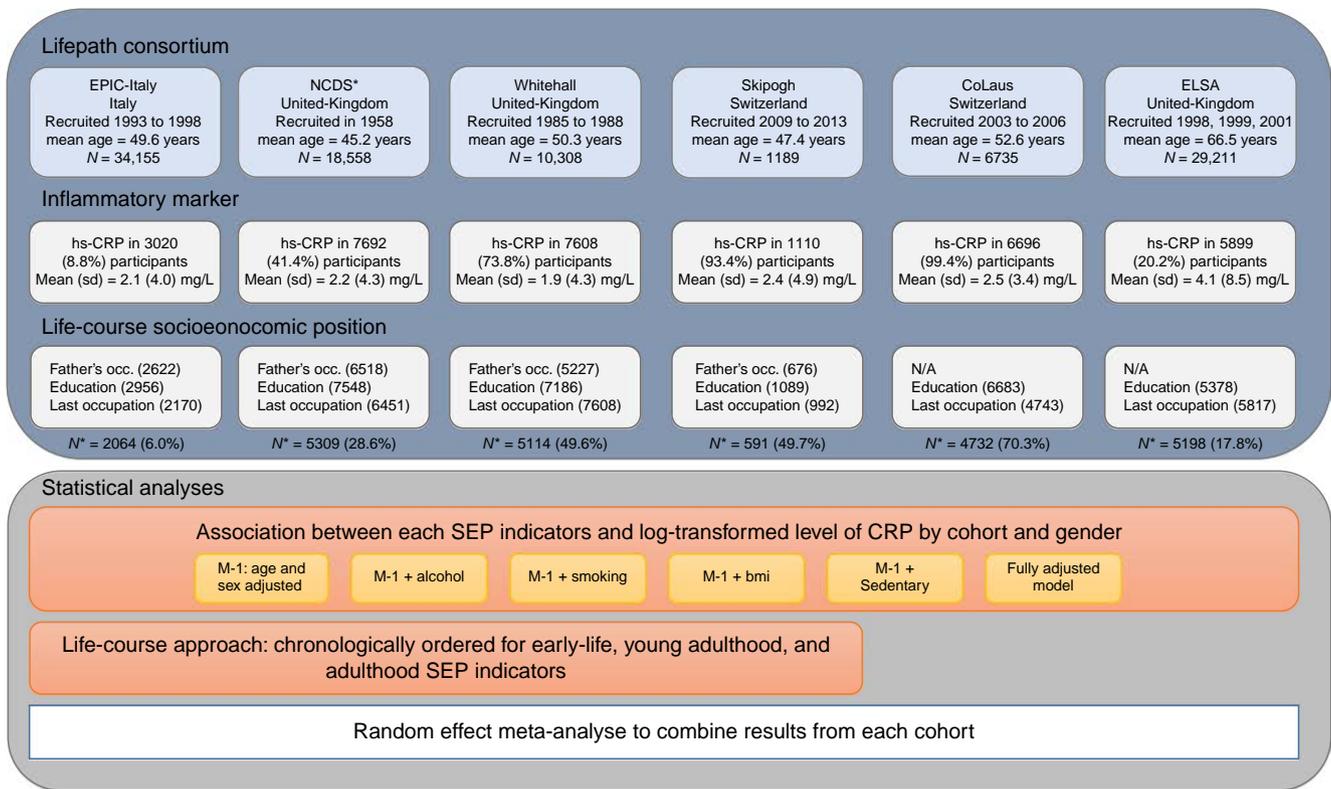


Fig. 1 Overview of the study workflow. *NCDS is the only birth cohort, therefore father’s occupation was collected prospectively and mean age corresponds to the age of participants at the time of the biomedical survey

A strong, graded association was observed between educational attainment and CRP (Table 2B and Fig. 2b). Results from the random effects meta-analyses revealed significant associations between low educational level and higher CRP level in the overall population (Model 1: Low vs High $\beta = 0.30, P < 0.001$). Observed associations remained markedly stable upon adjustment for alcohol consumption, smoking and sedentary lifestyle, but were weakened after adjustment for BMI (Model 1 + BMI: Low vs High: $\beta = 0.19, P < 0.001$) corresponding to a regression coefficient attenuation of 36.7% (Supplementary Note 2). The association between educational attainment and CRP, though attenuated, remained significant after adjustment for the four intermediate factors (Model 2: Low vs High $\beta = 0.14, P < 0.001$). In general, the strength of the associations between a medium level of education and CRP were weaker but significant compared to those with a low level of education, with a gradient persisting across educational groups in relation with CRP.

Having a less advantaged occupational position in adulthood was associated with having a higher CRP consistently in the random effect meta-analyses in the overall sample (Table 2C and Fig. 2c, Model 1 Less advantaged vs More advantaged: $\beta = 0.24, P < 0.001$). Among the four potential intermediate factors tested, the relationship between adulthood SEP and CRP was mostly affected by BMI in the random effect meta-analysis overall (Model 1 + BMI $\beta = 0.17, P < 0.001$) leading to a coefficient attenuation of 29.2% (Supplementary Note 2). The observed relationship between adulthood SEP and CRP was weakened when fully-adjusted (Model 2: $\beta = 0.11, P < 0.001$).

Contribution of SEP at each stage of the life course to adult CRP. We sequentially controlled for time-ordered life course SEP in the subset of four cohorts with complete data for SEP indicators over the life course (father’s occupation, education and occupation during adulthood), i.e. Skipogh, EPIC-Italy, Whitehall

and NCDS. This allowed us to evaluate the contribution of each specific SEP variable to examine their relative impact by life course stage (childhood, young adulthood and older adulthood), and by type of SEP measure (educational attainment or occupation). The association between less advantaged early-life SEP and CRP in adulthood (Model A $\beta = 0.19, P < 0.001$) was weakened when educational attainment (Model B $\beta = 0.11, P = 0.006$) or later occupation (Model C $\beta = 0.16, P = 0.001$) or both (Model D $\beta = 0.10, P = 0.011$) were included in the model and no longer significant in the fully adjusted model (Table 3).

An independent relationship between low educational attainment and high CRP was observed regardless of the model (Model B $\beta = 0.24, P < 0.001$, fully adjusted model $\beta = 0.12, P = 0.048$).

Participant’s last occupation was no longer significantly associated with CRP when controlling for education (Model D $\beta = 0.04, P = 0.693$), the association being almost null after adjustment for the four intermediate factors (Fully adjusted model $\beta = 0, P = 0.998$).

Gender effect. A significant association between less advantaged childhood SEP and higher CRP was observed in men (Model 1: Less advantaged vs More advantaged $\beta = 0.12, P = 0.011$, Supplementary Table 1A) and women (Model 1: Less advantaged vs More advantaged $\beta = 0.28, P < 0.001$, Supplementary Table 2A). The observed relationship between adulthood SEP and CRP was still significant albeit weakened after controlling for the four potential intermediate factors tested in women only (Model 2: Less advantaged vs More advantaged $\beta = 0.12, P = 0.046$, Supplementary Table 2A).

The strong, graded association was observed between educational attainment and CRP in men (Model 1: Low vs High $\beta = 0.28, P < 0.001$, Supplementary Table 1B) and women (Model 1: Low vs High $\beta = 0.33, P < 0.001$, Supplementary Table 2B). The association between educational attainment and CRP, though

Table 1 Descriptive statistics of the participants for the six cohort studies ($N = 23,008$ participants)

	Skipogh ($N = 591$)	EPIC-Italy ($N = 2064$)	CoLaus ($N = 4732$)	Whitehall ($N = 5114$)	ELSA ($N = 5198$)	NCDS ($N = 5309$)	All cohorts ($N = 23,008$)
	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)	Mean (SD) or N (%)
<i>Demographics</i>							
Age	48.8 (15.7)	52.6 (8.1)	49.2 (8.8)	49.9 (6)	67.4 (9.4)	45.8 (1)	53 (11)
Sex—Men	286 (48.39)	1071 (51.89)	2502 (52.87)	3618 (70.75)	2509 (48.27)	2823 (53.17)	12,809 (55.67)
Women	305 (51.61)	993 (48.11)	2230 (47.13)	1496 (29.25)	2689 (51.73)	2486 (46.83)	10,199 (44.33)
<i>Inflammation</i>							
CRP (mg/L)	2.6 (4.8)	2 (2.7)	2.2 (3.2)	1.9 (4.3)	4 (7.8)	2 (4)	2.5 (5)
<i>Socioeconomic position</i>							
<i>Father's occupation</i>							
Most advantaged	130 (22)	110 (5.33)	NA	487 (9.52)	NA	325 (6.12)	1052 (8.04)
Middle	212 (35.87)	838 (40.6)	NA	1561 (30.52)	NA	873 (16.44)	3484 (26.64)
Less advantaged	249 (42.13)	1116 (54.07)	NA	3066 (59.95)	NA	4111 (77.43)	8542 (65.32)
<i>Educational attainment</i>							
High	113 (19.12)	234 (11.34)	1665 (35.19)	1740 (34.02)	732 (14.08)	984 (18.53)	5468 (23.77)
Medium	188 (31.81)	476 (23.06)	668 (14.12)	1387 (27.12)	1130 (21.74)	273 (5.14)	4122 (17.92)
Low	290 (49.07)	1354 (65.6)	2399 (50.7)	1987 (38.85)	3336 (64.18)	4052 (76.32)	13,418 (58.32)
<i>Last occupation</i>							
Most advantaged	110 (18.61)	131 (6.35)	798 (16.86)	2619 (51.21)	1861 (35.8)	1752 (33)	7271 (31.6)
Middle	201 (34.01)	1012 (49.03)	1495 (31.59)	1626 (31.8)	1804 (34.71)	1690 (31.83)	7828 (34.02)
Less advantaged	280 (47.38)	921 (44.62)	2439 (51.54)	869 (16.99)	1533 (29.49)	1867 (35.17)	7909 (34.38)
<i>Intermediate factors</i>							
<i>Alcohol consumption</i>							
Abstainer	171 (28.93)	265 (12.84)	1215 (25.68)	964 (18.85)	2314 (44.52)	1043 (19.65)	5972 (25.96)
High	113 (19.12)	352 (17.05)	405 (8.56)	805 (15.74)	1303 (25.07)	1427 (26.88)	4405 (19.15)
Low	307 (51.95)	1447 (70.11)	3112 (65.77)	3345 (65.41)	1581 (30.42)	2839 (53.48)	12,631 (54.9)
<i>Smoking status</i>							
Current	149 (25.21)	611 (29.6)	1366 (28.87)	723 (14.14)	736 (14.16)	1272 (23.96)	4857 (21.11)
Former	209 (35.36)	619 (29.99)	1509 (31.89)	1909 (37.33)	2397 (46.11)	1473 (27.75)	8116 (35.27)
Never	233 (39.42)	834 (40.41)	1857 (39.24)	2482 (48.53)	2065 (39.73)	2564 (48.3)	10,035 (43.62)
<i>BMI</i>							
Underweight	12 (2.03)	6 (0.29)	46 (0.97)	29 (0.57)	28 (0.54)	16 (0.3)	137 (0.6)
Normal weight	294 (49.75)	763 (36.97)	2371 (50.11)	2648 (51.78)	1424 (27.4)	2557 (48.16)	10,057 (43.71)
Overweight	201 (34.01)	959 (46.46)	1702 (35.97)	1950 (38.13)	2287 (44)	2030 (38.24)	9129 (39.68)
Obese	84 (14.21)	336 (16.28)	613 (12.95)	487 (9.52)	1459 (28.07)	706 (13.3)	3685 (16.02)
<i>Sedentary</i>							
Yes	240 (40.61)	440 (21.32)	1666 (35.21)	991 (19.38)	1429 (27.49)	NA	4766 (26.93)
No	351 (59.39)	1624 (78.68)	3066 (64.79)	4123 (80.62)	3769 (72.51)	NA	12,933 (73.07)

attenuated, remained significant upon adjustment on the four intermediates factors in men (Model 2, Supplementary Table 1B) and women (Model 2, Supplementary Table 2B).

Men and women with a less advantaged SEP in adulthood had higher CRP levels compared to their more advantaged counterparts when analysed separately (Model 1 in men: Less advantaged vs more advantaged $\beta = 0.21$, $P = 0.001$, Model 1 in women: Less advantaged vs More advantaged $\beta = 0.31$, $P < 0.001$, Supplementary Table 1C and Table 2C, respectively). Regarding the life course SEP analyses, a similar pattern was observed in men as well as women (Supplementary Tables 3 and 4).

Heterogeneity between cohorts. Regarding cohort-specific findings, depicted in Supplementary Fig. 1, we observed that participants whose father had a middle or less advantaged occupational position had a significantly higher level of CRP in Skipogh, EPIC-Italy, Whitehall, and NCDS, but no longer significant after adjustment for all intermediate factors except in EPIC-Italy (Supplementary Table 5).

The association between low educational attainment and CRP was consistent across cohorts (Supplementary Fig. 2 and

Supplementary Table 6). This association was slightly attenuated upon adjustment for BMI but remained significant in CoLaus ($\beta = 0.14$, $p < 0.001$), Whitehall ($\beta = 0.11$, $p = 0.002$), ELSA ($\beta = 0.33$, $p < 0.001$), NCDS ($\beta = 0.28$, $p < 0.001$) and for CoLaus, ELSA and NCDS in the fully adjusted model ($\beta = 0.10$, $p = 0.003$; $\beta = 0.23$, $p < 0.001$; $\beta = 0.22$, $p < 0.001$, respectively, Supplementary Table 6).

Participants with less advantaged occupational position had a significantly higher level of CRP compared to those with a more advantaged occupation in CoLaus, Whitehall, ELSA and NCDS (Supplementary Fig. 3 and Supplementary Table 7). Among the four potential intermediate factors tested, the relationship between adulthood SEP and CRP was mostly affected by BMI remaining significant in CoLaus, Whitehall, ELSA, NCDS. The fully adjusted model followed the same pattern albeit weakened and the association in NCDS was no longer significant.

Regarding the life course SEP analysis, educational attainment was significantly associated with CRP in each cohort separately (Model B, Supplementary Table 8) but the adjustment for subsequent mediators played a different role according cohort: education was no longer significant in the fully adjusted model in Whitehall but still significant in NCDS.

Table 2 Multiple regression analyses of (A) father's occupational position in 4 cohorts (B) participant's educational attainment and (C) participant's last occupation with CRP at baseline in six cohorts from the Lifepath project. Meta-analyses results for the total population includes N = 13,078 for early life SEP and N = 23,008 for later in life SEP, except for Model 1 + sedentary where N = 7,769 and N = 17,699 respectively

	Category	(A) Father's occupational position ^a		(B) Participant's educational attainment ^b		(C) Participant's last occupation ^a	
		Middle (26.6%)	Less advantaged (65.3%)	Medium (17.9%)	Low (58.3%)	Middle (34.0%)	Less advantaged (34.4%)
Model 1 ^c	β (95% CI)	0.11 (-0.07; 0.28)	0.19 (0.11; 0.27)	0.15 (0.08; 0.21)	0.30 (0.22; 0.38)	0.09 (0.01; 0.16)	0.24 (0.14; 0.35)
	P-value	0.235	<0.001	<0.001	<0.001	0.023	<0.001
	I^2	74.8%	1.3%	43.4%	75.6%	67.9%	81.4%
	P_H	0.015	0.466	0.147	0.001	0.009	<0.001
Model 1 ^c + Alcohol	β (95% CI)	0.10 (-0.08; 0.28)	0.18 (0.11; 0.26)	0.14 (0.08; 0.20)	0.29 (0.21; 0.36)	0.08 (0.004; 0.15)	0.22 (0.12; 0.33)
	P-value	0.260	<0.001	<0.001	<0.001	0.038	<0.001
	I^2	76.2%	0.0%	32.7%	68.5%	67.0%	81.0%
	P_H	0.012	0.518	0.262	0.008	0.010	<0.001
Model 1 ^c + Smoking	β (95% CI)	0.11 (-0.07; 0.28)	0.18 (0.10; 0.25)	0.13 (0.07; 0.20)	0.27 (0.19; 0.35)	0.07 (0.01; 0.14)	0.21 (0.12; 0.31)
	P-value	0.233	<0.001	<0.001	<0.001	0.036	<0.001
	I^2	75.2%	0.0%	41.8%	74.6%	62.3%	79.1%
	P_H	0.014	0.452	0.167	0.001	0.026	<0.001
Model 1 ^c + BMI	β (95% CI)	0.06 (-0.10; 0.22)	0.10 (0.03; 0.17)	0.10 (0.03; 0.16)	0.19 (0.11; 0.28)	0.08 (0.02; 0.14)	0.17 (0.09; 0.25)
	P-value	0.493	0.005	0.003	<0.001	0.014	<0.001
	I^2	73.4%	0.0%	41.1%	79.3%	55.7%	73.3%
	P_H	0.019	0.451	0.171	<0.001	0.047	0.004
Model 1 ^c + Sedentary	β (95% CI)	0.16 (-0.002; 0.32)	0.17 (0.08; 0.26)	0.13 (0.06; 0.19)	0.25 (0.18; 0.32)	0.12 (0.07; 0.16)	0.26 (0.21; 0.32)
	P-value	0.053	0.000	<0.001	<0.001	<0.001	<0.001
	I^2	54.2%	0.0%	31.7%	60.4%	0.2%	18.5%
	P_H	0.109	0.381	0.296	0.048	0.488	0.099
Model 2 ^d	β (95% CI)	0.05 (-0.11; 0.21)	0.08 (0.01; 0.15)	0.07 (0.02; 0.13)	0.14 (0.07; 0.21)	0.05 (0; 0.10)	0.11 (0.05; 0.17)
	P-value	0.518	0.021	0.006	<0.001	0.051	<0.001
	I^2	74.1%	0.0%	21.4%	67.2%	35.4%	51.4%
	P_H	0.017	0.458	0.411	0.009	0.176	0.065

CI confidence interval I^2 percentage of between study heterogeneity, P_H P-value of heterogeneity test, BMI body mass index

^a Referent group: most advantaged

^b Referent group: high educational attainment

^c Model 1 adjusted for age and sex

^d Model 2 controlled for age, sex and additionally alcohol, smoking, BMI and sedentary

Sensitivity analyses. The exclusion of participants with a CRP ≥ 10 mg/L ($N = 870$, 3.8%) had little effect on the previous observed associations between life course SEP and CRP: a social gradient in adult CRP was still observed. Regarding educational attainment, participants with low educational level had a significant higher level of CRP (Model 1 meta-analysis $\beta = 0.26$, $P < 0.001$, Supplementary Table 9). Our results were slightly attenuated but still significant after adjustment for BMI (Model 1 + BMI meta-analysis $\beta = 0.17$, $P < 0.001$, Supplementary Table 9) and in the fully adjusted model (fully adjusted model meta-analysis $\beta = 0.12$, $P < 0.001$, Supplementary Table 9).

In Skipogh, CoLaus, Whitehall and ELSA, CRP measures were also collected after the first follow-up (Supplementary note 3): the strong, graded association between educational level and CRP was still observed few years later (fully adjusted meta-analysis: $\beta = 0.10$, $P = 0.012$ and $\beta = 0.15$, $P = 0.001$ for mid and low educational level, respectively, Supplementary Table 10).

Discussion

In this multi-cohort study using data from several European countries, we found that disadvantaged socioeconomic position at each life stage was associated with increased inflammation assessed using CRP. As a general pattern, accounting for behavioural factors (alcohol consumption, smoking status and sedentary) and BMI explained part, but not all of the observed SEP differences in inflammation. In general, associations between

life course SEP and inflammation were mainly attenuated by BMI in adulthood. A subsequent analysis carried out whereby life course SEP variables were added to the model in chronological order indicated that participants with a low educational attainment had a higher level of CRP independently of early life or later adulthood SEP as well as behavioural factors and BMI. The persistence of a significant association between low educational attainment and a high CRP level in adulthood indicates that educational attainment may be a consistent and important upstream risk factor for elevated inflammation. These findings suggest that in addition to health behaviours and BMI, alternative pathways by which SEP may affect inflammation deserve to be studied further. These include, for example, chemical exposures, infectious diseases (such as Herpes and Epstein Barr infections etc) and oral health conditions, as well as psychosocial stress.

Overall, our results are consistent with previous studies investigating the relationship between SEP at different life stages and chronic inflammation. A previous systematic review of published observational studies up to 2006 reported associations between adult SEP and CRP⁹. Since then the same relationship has also been demonstrated in various other observational studies from worldwide countries (Taiwan²⁷, Europe¹³, Brazil²⁸ and others^{17,29-31}). A recent meta-analysis of population-based and cross-sectional studies showed that low childhood SEP was associated with higher adulthood CRP¹⁵.

Though we observed significant associations in some of the fully adjusted models, notably for disadvantaged educational

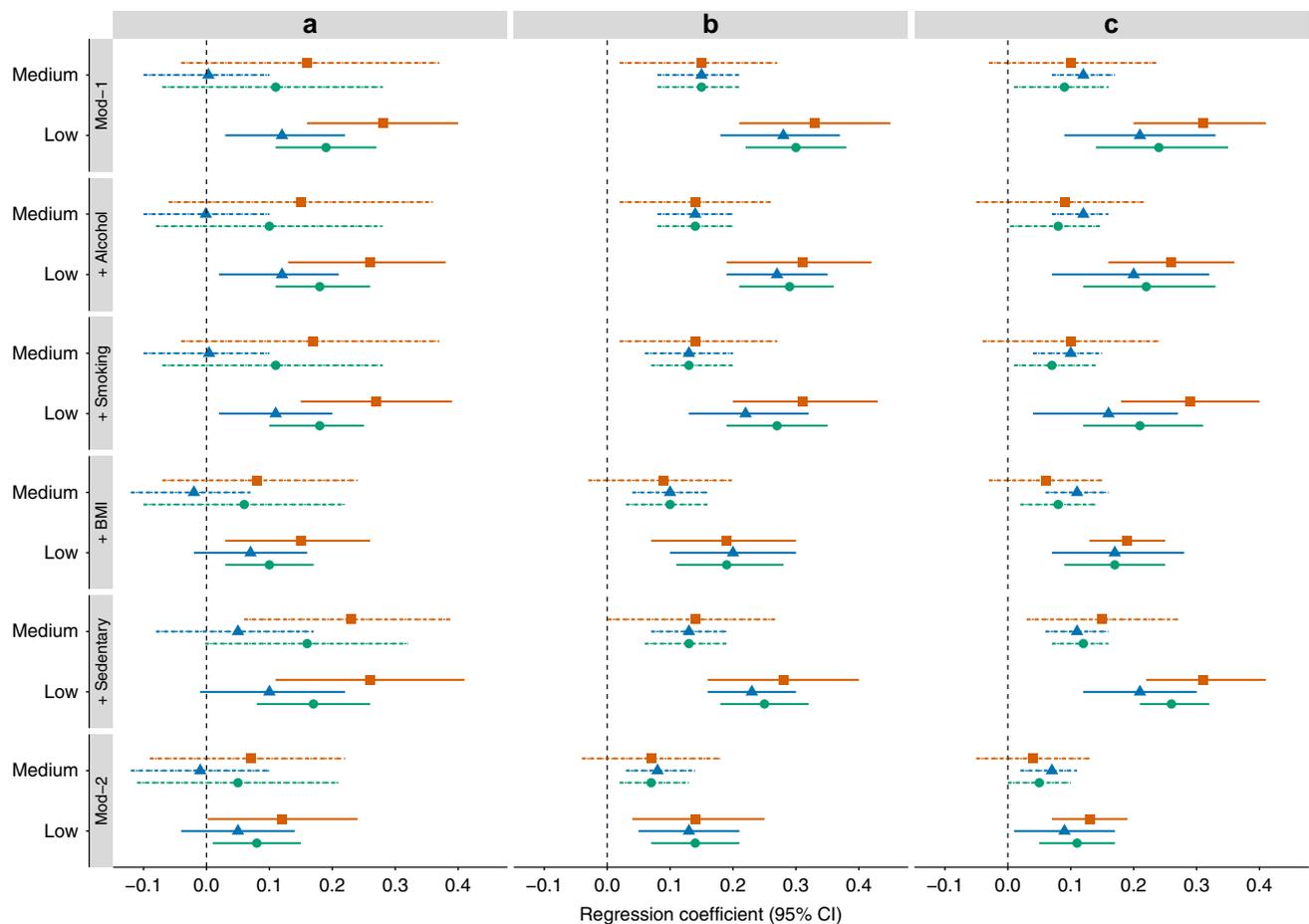


Fig. 2 Forest plot of regression coefficients [95% confidence interval] for the association between **a** father's occupational position, **b** participant's educational attainment, **c** participant's last occupation and CRP concentration at baseline in random effect meta-analysis framework for the total population and by gender for Model 1 (Mod-1), after adjustment for each intermediate factor (+Alcohol, +Smoking, +BMI, +Sedentary) and further adjusted for all intermediate factors together (Mod-2). The high SEP group was used as reference, solid lines represent the medium SEP group and dotted lines the low SEP group. Meta-analyses results for the total population (in orange) includes $N = 13,078$ for early life SEP and $N = 23,008$ for later in life SEP, respectively, $N = 7798$ and $N = 12,809$ for men (in blue) and $N = 5280$ and $N = 10,199$ for women (in green)

attainment and SEP, adjusting for intermediate factors weakened associations. This may reflect indirect pathways between SEP and inflammation. The intermediate factor that most influenced the association between SEP and CRP was BMI. Other studies have reported that BMI attenuates the relationships between SEP and CRP^{28,32}. Together these sources of evidence may point towards the accumulation of body fat among more socially disadvantaged populations as being a mechanism leading to higher levels of inflammation. Nutrient excess is related to dysregulation in the cellular and molecular intermediate factors of immunity and inflammation. With an estimated excess of 20–30 million macrophages accumulate with each kilogram of excess fat in humans, one could argue that increased adipose tissue mass is a state of increased inflammatory mass³³. Smoking is also a known inflammation-inducing behaviour³⁴. In the meta-analyses, alterations in the association between SEP and CRP when smoking was introduced were negligible. In the overall meta-analyses alcohol consumption did not affect the results once added to the models, nor did sedentary lifestyle. This phenomenon could be explained by the fact that smoking and alcohol do not always show a clear social gradient contrary to BMI which is consistently associated with SEP across all cohorts. As expected, the results indicate that the social distribution of intermediate factors is key to their ability to act as mediators between SEP and health.

Beyond behaviours and BMI, the association between less advantaged SEP and CRP remained, notably when educational attainment was examined. This may suggest other underlying social-to-biological mechanisms operating whereby the socio-economic environment confers inflammation-inducing or reducing mechanisms. One of the likely mechanisms involved is the stress response system, where higher educational attainment may act as a physiological stress-regulating buffer. In the case of education, a higher educational attainment may provide increased sense of control, which is a suggested pathway linking education to health³⁵. Though gaps in the literature remain as to whether psychological measures such as sense of control relate to inflammation^{36,37}, such relationships have recently been described for multiple stress biomarkers (known as sterile inflammation) and for real-life stress exposure^{38,39}. Emotional regulation has been associated with both academic success⁴⁰ in children and inflammation⁴¹. A more advantaged SEP may also favour access to better paid⁴² more stable occupations with access to work flexibility.

Overall, we found similar patterns in the relationship between SEP and CRP for men and women, though the effect size between father's occupation and educational attainment and CRP were higher in women. Some differences between men and women may come down to differences in immunological and inflammatory responses influenced by both sex and gender. Sex contributes to

Table 3 Life course multiple regression analyses of SEP with CRP at baseline in a random effect meta-analytical framework from 4 cohorts from the Lifepath (N = 13,078)

	Model A ^a					Model B ^b				
	β	95% CI	P-value	I ²	P _H	β	95% CI	P-value	I ²	P _H
<i>Father's occupation</i>										
Middle	0.11	(-0.07; 0.28)	0.235	74.8%	0.015	0.06	(-0.11; 0.23)	0.489	73.0%	0.020
Less advantaged	0.19	(0.11; 0.27)	<0.001	1.3%	0.466	0.11	(0.03; 0.19)	0.006	0.0%	0.585
<i>Educational level</i>										
Medium	—	—	—	—	—	0.10	(0.03; 0.16)	0.003	0.0%	0.640
Low	—	—	—	—	—	0.24	(0.14; 0.35)	<0.001	64.5%	0.026
<i>Last occupation</i>										
Middle	—	—	—	—	—	—	—	—	—	—
Less advantaged	—	—	—	—	—	—	—	—	—	—
Model C^c						Model D^d				
Father's occupation	0.09	(-0.09; 0.27)	0.306	76.2%	0.013	0.06	(-0.12; 0.24)	0.501	75.0%	0.015
Middle	0.16	(0.07; 0.26)	0.001	28.7%	0.270	0.10	(0.02; 0.18)	0.011	0.0%	0.480
Less advantaged	—	—	—	—	—	—	—	—	—	—
Educational level	—	—	—	—	—	0.10	(0.01; 0.18)	0.021	19.2%	0.433
Medium	—	—	—	—	—	0.23	(0.09; 0.38)	0.002	78.9%	0.000
Low	—	—	—	—	—	—	—	—	—	—
<i>Last occupation</i>										
Middle	0.02	(-0.10; 0.14)	0.758	73.9%	0.002	-0.03	(-0.17; 0.12)	0.706	79.3%	0.000
Less advantaged	0.14	(-0.02; 0.31)	0.084	82.6%	0.000	0.04	(-0.17; 0.25)	0.693	88.2%	<0.001
Fully adjusted^e										
Father's occupation										
Middle	0.04	(-0.13; 0.21)	0.658	76.0%	0.012					
Less advantaged	0.05	(-0.02; 0.12)	0.192	0.1%	0.359					
Educational level										
Medium	0.05	(-0.03; 0.12)	0.231	18.3%	0.467					
Low	0.12	(0.001; 0.23)	0.048	69.8%	0.008					
<i>Last occupation</i>										
Middle	-0.02	(-0.12; 0.08)	0.683	60.9%	0.038					
Less advantaged	0.00	(-0.13; 0.13)	0.998	70.8%	0.012					

CI confidence interval, I² heterogeneity, P_H P-value of heterogeneity test, BMI body mass index
^a Model A adjusted for age, sex, father's occupational position
^b Model B adjusted for age, sex, father's occupational position and participant's educational attainment
^c Model C adjusted for age, sex, father's occupational position and participant's last occupation
^d Model D adjusted for age, sex, father's occupational position, participant's educational attainment and participant's last occupation
^e Fully adjusted model controlled age, sex, father's occupational position, participant's educational attainment, participant's last occupation and additionally alcohol, smoking, BMI and sedentary

differences that influence the deposit of adipose tissue, its physiological interaction with the endocrine system, and associated metabolic risk over the life course⁴³. Gender may reflect behaviours that influence exposure to microorganisms, access to healthcare or health-related behaviours that affect health trajectories⁴⁴. We also observed that the attenuation of the relationship once BMI was introduced into the models appeared to be greater among women. This may suggest that inflammation induced through greater body mass and adipose tissue was a more significant pathway for women from less advantaged SEP groups⁴⁵.

Findings were directionally consistent across the cohort studies from the UK, Italy and Switzerland, representing a variety of contexts and period, although differences in effect size were observed. This suggests that the social-to-biological process we are observing is shared across European high-income country populations. Some differences between cohorts are likely down to their specific characteristics. The strongest association between father's occupation and CRP were observed for the NCDS and Skipogh even though the NCDS reported among the highest proportions of manual occupation among participant's fathers, while Skipogh had the lowest. Whitehall II, and occupational cohort of British civil-servants had one of the lowest proportions of respondents with a manual occupation, and the relationship between manual occupation and CRP was the strongest among the cohorts we studied. BMI has also a dominant role across cohorts and for all SEP variables (Supplementary Figs. 1–3; Supplementary Tables 5–7) even though the prevalence of obese individuals by SEP differed by cohort study: ELSA, a study of ageing in an elderly English population, was the cohort with the lowest prevalence of normal BMI, and with the highest proportion of obese individuals compared to the other cohorts. These results highlight the important to take into account the social distribution of risk factors in a specific population to understand which of them may contribute to the relationship between SEP and biology.

The main strength of the study was the use of harmonised data from six cohort studies on participants from three European countries. The longitudinal nature of the studies allowed us to examine how SEP at different stages of the life course preceding the measurement of CRP are related. This allows us to begin unpacking the complex relationships between social and biological variables by measuring them at different time points. Given the varied nature of the cohorts and their populations, we were able to maintain the assumption that the relationships being assessed in the models were cohort specific by using a random effects model when pooling the cohorts together. The breadth and depth of data available allowed us to perform multiple sensitivity analyses, which underlined further the stability of the results. Finally, the use of CRP as a reliable and reproducible biomarker measuring overall inflammatory response allows for comparisons, given its wide availability as a biomarker through high-sensitivity assays which are stable.

All cohort studies are susceptible to attrition, and biased samples. The cohorts we used may be not representative of the general population, in particular because of a potential underrepresentation of the more advantaged/less advantaged SEP categories. In these analyses, we used complete case data from each dataset to conduct cohort-specific and pooled meta-analyses using random effects modelling. Our models included a set of harmonised variables available across the cohorts allowing us to adjust for covariates, however there may be some residual confounding. This possibility of confounding means that we cannot fully ascertain causal directionality. Furthermore, the harmonisation process favours standardising variables across cohorts, which may mean that some cohort-specificities are smoothed out, or lost. This process also led to some variation in our models and results (only 4 cohorts contained father's occupation and a lack of equivalent harmonised variable for sedentary lifestyle for the NCDS). There is likely to be measurement error and heterogeneity across cohorts regarding socioeconomic variables. Father's occupation was collected in a variety of ways, referring to historical and

country specific occupations, and ultimately these variables were divided in three categories. This is likely to have led to some degree of misclassification. Additionally, the distribution of the SEP measures was also skewed. To assess the robustness of our results and since misclassification error may have occurred for the assessment of life course SEP, we repeated the main analysis (i) using the Less advantaged/Low group as a reference (Supplementary Table 11) and (ii) with a binary indicator of father's occupational position and participant's last occupation (Supplementary Table 12). Estimates of the associations were all consistent with those in the main analysis, suggesting that our findings are robust. The SEP indicators we used measure partially different and inter-dependent aspects of life experiences and may be related to different perceptions and belief about health-related behaviours. Education, for instance, reflects the transition from childhood/adolescence to adulthood together with the intellectual and socioeconomic resources of the family, but it is also an important determinant of future employment and income^{46,47}.

To meet the assumption that CRP follows a Gaussian distribution in the sample, values were transformed on the natural log scale to approximate the normal curve; alternative techniques to address the skewness of the CRP distribution include parametric modelling based on flexible size distributions⁴⁸. We only had one inflammation biomarker that was available across all these cohorts where life course SEP variables had been collected, therefore the generalisation to others inflammatory-related biomarkers remains to be shown. While CRP is a useful proxy for overall inflammatory status, we would ideally have used it alongside Tumour Necrosis Factor α and Interleukin 6 for example, in order to better capture the complex regulatory cascades involved in inflammation. The inflammatory response is coordinated by a number of cytokines as well as signalling proteins of the immune system⁴⁹. Future investigations on inflammatory-related biomarkers at wider and larger scale are needed to disentangle the complex inflammatory responses associated to pathological and immunological processes.

Our findings provide consistent evidence across cohorts that less advantaged socioeconomic conditions experienced at three different life stages are associated with increased overall inflammation in adulthood. The findings point towards educational attainment as the SEP variable most strongly associated with CRP, suggesting that further examination of social-to-biological pathways operating through educational processes are worthy of biomedical and public health attention. Questions remain as to whether educational attainment captures material and/or psychosocial exposures affecting biology through the maturation of the immune system, having lasting effects on overall inflammation.

In conclusion, our study highlights the important role that, from early life, social factors play in health beyond behaviours and lifestyle factors. We document the biological consequences – through inflammation – of social disadvantages, justifying the need to intervene from early life to reduce social disparities in health. Further work integrating biosocial perspectives in longitudinal settings is needed to better understand the mechanisms through which the social environment shapes physiological processes that are important for complex health outcomes.

Methods

Study population. Lifepath is a Horizon 2020 Research and Innovation Program European funded project which aims to investigate the biological pathways underlying social differences in healthy ageing. For this purpose, the project includes a consortium of eighteen cohort studies (child and adult) across different countries and time periods, with demographic, clinical, biological and socioeconomic data. Data harmonization were performed in order to merge and analyse the cohorts together⁵⁰. More details about the Lifepath project are available elsewhere (<http://www.lifepathproject.eu>).

We selected six cohorts for which data on CRP, at least two measures of socioeconomic position at two different life course stages, and behaviours/BMI were available. Detailed information on cohorts used in our study is provided in Supplementary Table 13. Briefly, our panel of cohorts includes a subset of the Italian component of the European prospective investigation on nutrition and cancer study (EPIC-Italy)⁵¹; two cohorts from Switzerland: CoLaus⁵² and Skipogh⁵³; and three British cohorts: Whitehall II⁵⁴, the English longitudinal study of ageing (ELSA⁵⁵), and the National Child Development Study (NCDS⁵⁶). All Lifepath cohorts have been described in detail elsewhere⁵⁰. Each study was approved by the relevant local or national ethics committees and all participants gave informed consent to participate.

Inflammatory markers. We used CRP to measure overall inflammation, selecting the first available measurement of CRP in each cohort for our analyses. Information about sampling methods and laboratory analyses for each cohort study are given in Supplementary Table 15. Baseline CRP was measured in mg/L using high-sensitivity assay in all studies. CRP values were natural log transformed in order to normalise their distributions given their skewed nature.

Life course socioeconomic positions. We used the harmonised definition of life course SEP across cohort study described in details by d'Errico et al.²⁶. Childhood SEP was ascertained using father's occupational position reported by the participants and recoded according to the European Socioeconomic Classification (E-SeC) where occupations are classified according to their employment relations and work conditions. We applied 3 E-SeC categories: less advantaged occupations [lower clerical, services, and sales workers, skilled workers, semi-skilled and unskilled workers, E-SeC class 7–9], middle occupations [small employers and self-employed, farmers, lower supervisors and technicians, E-SeC class 4–6], and more advantaged occupations [higher professionals and managers, higher clerical, services, and sales workers, E-SeC class 1–3].

SEP in young adulthood was measured using the participant's educational attainment categorised in three groups: primary or lower secondary school (from 7 to 9 years after kindergarten with a basic curriculum in languages, mathematics and other subjects); higher secondary school (around 4–5 years more, high school diploma level) and tertiary education (any degree after high school, such as BSc, MSc, and further education)²⁶ hereafter referred as low, medium and high educational level respectively. Adulthood SEP was measured by the participant's last occupational position in three groups: less advantaged occupations, middle occupations, and more advantaged occupations according to the same three occupational classes used for father's occupational position and based on self-reported information.

Health behaviours and lifestyle factors. The following factors have been found to be associated with both SEP and CRP in the literature and were therefore considered as intermediate variables that may mediate the relationship between SEP and CRP: BMI^{57,58}, categorical: <18.5 kg/m², underweight; 18.5–24.9 kg/m², normal; 25.0–29.9 kg/m², overweight; >18.5 kg/m², obese); smoking status^{34,59} (categorical: current, former, never); alcohol consumption^{60,61} (categorical: abstainers, moderate consumption – men: ≤ 21 alcohol units per week/women: ≤ 14 alcohol units per week but not abstainers, high consumption – men: > 21 alcoholic units per week/women: > 14 alcoholic units per week); sedentary^{62,63} (categorical: sedentary, no sedentary based on response to questions on leisure physical activity) except in NCDS in which this variable was not collected. We selected these intermediate variables from the closest data collection wave to that of the CRP measurement. If data were unavailable at the same wave, we imputed data from the next available data wave (for CoLaus: $n = 12$, 0.3%; Skipogh: $n = 159$, 26.9%; Whitehall: $n = 277$, 5.4%; ELSA: $n = 69$, 1.3%).

Statistical analysis. Means and frequencies were reported for all continuous and categorical baseline characteristics by cohort and by each SEP variable (father's occupation, educational attainment, participant's last occupation). Chi-squared test or Fisher exact test for the categorical variables and *T*-test or Wilcoxon rank test for continuous variables were used to estimate bivariate associations.

Linear regression models were used to investigate the relationship between SEP and CRP concentration at baseline. We defined a minimally adjusted model controlling for age and sex (sex only for the NCDS where participants are the same age) (Model 1), then we further adjusted for each potential intermediate factor independently. We finally defined a second model including all potential intermediate factors (Model 2). For all three SEP indicators, the highest socioeconomic group was used as reference (non-manual for father's job and last occupation and high for educational level). A positive regression coefficient therefore indicates an increased level of CRP in less advantaged socioeconomic groups.

To mimic life course experiences, we sequentially adjusted for the chronologically ordered SEP indicators; resulting in four time-sequenced models on the subset of the four cohorts (Skipogh, EPIC-Italy, Whitehall and NCDS) containing all three SEP indicators:

Model A: Father's occupation + Age + Sex

Model B: Father's occupation + Educational attainment + Age + Sex

Model C: Father's occupation + Last occupation + Age + Sex

Model D: Father's occupation + Educational attainment + Last occupation + Age + Sex

Fully adjusted: Model D + Intermediate factors

Model A allows us to identify the potential early life biological embedding of social conditions, while model B and C allows the identification of emerging social-biological signals specific to young adult (model B) and later adulthood (model C) experiences. The resulting regression coefficients measure mutually adjusted effects. Model D evaluates the relative contribution of each SEP measure across the life course. The fully adjusted model estimates the effect of each SEP indicator, controlled for the intermediate factors in adulthood.

None of the interaction tested between gender and SEP was found significant (Supplementary Table 14) within each cohort studies. But prior research on health and health behaviours have indicated that there may be important gender differences pattern by SEP⁶⁴. Evidence are also accumulating showing potential sex differences along the life course process of disease development and progression. Furthermore physiological responses to chronic stress could differently impact men and women. So far few studies have examined gender in relation to SEP and CRP in cohort studies setting^{14,65,66}. Therefore, the multivariate linear analysis was run separately by gender.

Random effects meta-analyses⁶⁷ were conducted using the metafor R package. Between-study heterogeneity was estimated through a restricted maximum-likelihood estimator. The combined effect represents the mean of the population of true effects. We reported the estimated average effect ($\hat{\mu}$) and the estimated percentage of the total amount of variability that can be attributed to heterogeneity (I^2) together with the Q-statistic associated p -value.

We repeated the analyses excluding individuals with CRP above 10 mg/L, as high CRP maybe more likely result from an acute infection than chronic inflammation (Supplementary Table 9). Analyses were repeated using CRP measured at the next follow-up when available (Skipogh, CoLaus, Whitehall and ELSA, Supplementary Table 10). We ran all the analyses using the less advantaged/low group as a reference for each life course SEP (Supplementary Table 11) and using dichotomized SEP indicators for father's occupational position and participant's last occupation (Supplementary Table 12)

Statistical analyses were performed using R. version 1.1.383.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data and computer code uses to support the findings of this study are available from the corresponding author upon reasonable request.

Received: 4 July 2018 Accepted: 28 January 2019

Published online: 15 February 2019

References

- Danesh, J. et al. Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ* **321**, 199–204 (2000).
- Black, P. H. Stress and the inflammatory response: a review of neurogenic inflammation. *Brain Behav. Immun.* **16**, 622–653 (2002).
- Kivimäki, M. & Steptoe, A. Effects of stress on the development and progression of cardiovascular disease. *Nat. Rev. Cardiol.* **15**, 215–229 (2018).
- Emerging risk factors collaboration. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet* **375**, 132–140 (2010).
- Li, Y. et al. Hs-CRP and all-cause, cardiovascular, and cancer mortality risk: a meta-analysis. *Atherosclerosis* **259**, 75–82 (2017).
- Pepys, M. B. & Baltz, M. L. Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv. Immunol.* **34**, 141–212 (1983).
- Ridker, P. M. A test in context: high-sensitivity C-reactive protein. *J. Am. Coll. Cardiol.* **67**, 712–723 (2016).
- Allin, K. H. & Nordestgaard, B. G. Elevated C-reactive protein in the diagnosis, prognosis, and cause of cancer. *Crit. Rev. Clin. Lab. Sci.* **48**, 155–170 (2011).
- Nazmi, A. & Victora, C. G. Socioeconomic and racial/ethnic differentials of C-reactive protein levels: a systematic review of population-based studies. *BMC Public Health* **7**, 212 (2007).
- Stepanikova, I., Bateman, L. B. & Oates, G. R. Systemic inflammation in midlife: race, socioeconomic status, and perceived discrimination. *Am. J. Prev. Med.* **52**, S63–S76 (2017).
- Davillas, A., Benzeval, M. & Kumari, M. Socio-economic inequalities in C-reactive protein and fibrinogen across the adult age span: findings from Understanding Society. *Sci. Rep.* **7**, 2641 (2017).
- West, D. A. et al. Life-course origins of social inequalities in adult immune cell markers of inflammation in a developing southern Chinese population: the Guangzhou Biobank Cohort Study. *BMC Public Health* **12**, 269 (2012).
- Fraga, S. et al. Association of socioeconomic status with inflammatory markers: a two cohort comparison. *Prev. Med.* **71**, 12–19 (2015).
- Gruenewald, T. L., Cohen, S., Matthews, K. A., Tracy, R. & Seeman, T. E. Association of socioeconomic status with inflammation markers in black and white men and women in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Soc. Sci. Med.* **69**, 451–459 (2009).
- Liu, R. S. et al. Socioeconomic status in childhood and C reactive protein in adulthood: a systematic review and meta-analysis. *J. Epidemiol. Community Health* **71**, 817–826 (2017).
- Kelly-Irving, M. & Delpierre, C. Embodiment over the lifecourse: a case for examining cancer aetiology. In *Handbook of biology and society* (2017).
- Pollitt, R. A. et al. Early-life and adult socioeconomic status and inflammatory risk markers in adulthood. *Eur. J. Epidemiol.* **22**, 55–66 (2007).
- Gares, V. et al. The role of the early social environment on Epstein Barr virus infection: a prospective observational design using the Millennium Cohort Study. *Epidemiol. Infect.* **145**, 3405–3412 (2017).
- Petrovic, D. et al. The contribution of health behaviors to socioeconomic inequalities in health: a systematic review. *Prev. Med.* **113**, 15–31 (2018).
- Martikainen, P., Brunner, E. & Marmot, M. Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc. Sci. Med.* **56**, 1397–1410 (2003).
- Danese, A. et al. Adverse childhood experiences and adult risk factors for age-related disease: depression, inflammation, and clustering of metabolic risk markers. *Arch. Pediatr. Adolesc. Med.* **163**, 1135–1143 (2009).
- Danese, A., Pariante, C. M., Caspi, A., Taylor, A. & Poulton, R. Childhood maltreatment predicts adult inflammation in a life-course study. *Proc. Natl Acad. Sci. USA* **104**, 1319–1324 (2007).
- Barboza Solís, C. et al. Mediating pathways between parental socio-economic position and allostatic load in mid-life: findings from the 1958 British birth cohort. *Soc. Sci. Med.* **165**, 19–27 (2016).
- Barboza Solís, C. et al. Adverse childhood experiences and physiological wear-and-tear in midlife: findings from the 1958 British birth cohort. *Proc. Natl Acad. Sci. USA* **112**, E738–E746 (2015).
- Barboza Solís, C., Fantin, R., Kelly-Irving, M. & Delpierre, C. Physiological wear-and-tear and later subjective health in mid-life: findings from the 1958 British birth cohort. *Psychoneuroendocrinology* **74**, 24–33 (2016).
- d'Errico, A. et al. Socioeconomic indicators in epidemiologic research: a practical example from the LIFEPAH study. *PLOS ONE* **12**, e0178071 (2017).
- Lin, Y.-H., Jen, M.-H. & Chien, K.-L. Association between life-course socioeconomic position and inflammatory biomarkers in older age: a nationally representative cohort study in Taiwan. *BMC Geriatr.* **17**, 201 (2017).
- Camelo, L. V. et al. Life course socioeconomic position and C-reactive protein: mediating role of health-risk behaviors and metabolic alterations. The Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *PLoS ONE* **9**, e108426 (2014).
- Koster, A. et al. Association of inflammatory markers with socioeconomic status. *J. Gerontol. A Biol. Sci. Med. Sci.* **61**, 284–290 (2006).
- Loucks, E. B. et al. Life course socioeconomic position is associated with inflammatory markers: the Framingham Offspring Study. *Soc. Sci. Med.* **71**, 187–195 (2010).
- Stringhini, S. et al. Association of lifecourse socioeconomic status with chronic inflammation and type 2 diabetes risk: the Whitehall II prospective cohort study. *PLoS Med.* **10**, e1001479 (2013).
- Jones, R. et al. Novel coronary heart disease risk factors at 60–64 years and life course socioeconomic position: the 1946 British birth cohort. *Atherosclerosis* **238**, 70–76 (2015).
- Lumeng, C. N. & Saltiel, A. R. Inflammatory links between obesity and metabolic disease. *J. Clin. Invest.* **121**, 2111–2117 (2011).
- Shiels, M. S. et al. Cigarette smoking and variations in systemic immune and inflammation markers. *J. Natl Cancer Inst.* **106**, dju294–dju294 (2014).
- Chandola, T., Clarke, P., Morris, J. N. & Blane, D. Pathways between education and health: a causal modelling approach. *J. R. Stat. Soc. Ser. -Stat. Soc.* **169**, 337–359 (2006).
- Boylan, J. M. & Ryff, C. D. Varieties of anger and the inverse link between education and inflammation: toward an integrative framework. *Psychosom. Med.* **75**, 566–574 (2013).
- Sin, N. L., Graham-Engeland, J. E. & Almeida, D. M. Daily positive events and inflammation: findings from the National Study of Daily Experiences. *Brain Behav. Immun.* **43**, 130–138 (2015).

38. Fleshner, M. & Crane, C. R. Exosomes, DAMPs and miRNA: features of stress physiology and immune homeostasis. *Trends Immunol.* **38**, 768–776 (2017).
39. Magnusson Hanson, L. L. et al. Work stress, anthropometry, lung function, blood pressure, and blood-based biomarkers: a cross-sectional study of 43,593 French men and women. *Sci. Rep.* **7**, 9282 (2017).
40. Graziano, P. A., Reavis, R. D., Keane, S. P. & Calkins, S. D. The role of emotion regulation in children's early academic success. *J. Sch. Psychol.* **45**, 3–19 (2007).
41. Appleton, A. A., Buka, S. L., Loucks, E. B., Gilman, S. E. & Kubzansky, L. D. Divergent associations of adaptive and maladaptive emotion regulation strategies with inflammation. *Health Psychol.* **32**, 748–756 (2013).
42. Friedman, E. M. & Herd, P. Income, education, and inflammation: differential Associations in a National Probability Sample (The MIDUS Study). *Psychosom. Med.* **72**, 290–300 (2010).
43. Palmer, B. F. & Clegg, D. J. The sexual dimorphism of obesity. *Mol. Cell Endocrinol.* **402**, 113–119 (2015).
44. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626 (2016).
45. Yang, Y. C., Johnson, M. P., Schorpp, K. M., Boen, C. E. & Harris, K. M. Young adult risk factors for cancer: obesity, inflammation, and sociobehavioral mechanisms. *Am. J. Prev. Med.* **53**, S21–S29 (2017).
46. Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W. & Davey Smith, G. Indicators of socioeconomic position. *J. Epidemiol. Community Health* **60** (part 1), 7–12 (2006).
47. Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W. & Davey Smith, G. Indicators of socioeconomic position. *J. Epidemiol. Community Health* **60** (part 2), 95–101 (2006).
48. Davillas, A. & Jones, A. M. Parametric models for biomarkers based on flexible size distributions. *Health Econ.* <https://doi.org/10.1002/hec.3787> (2018).
49. Ashley, N. T., Weil, Z. M. & Nelson, R. J. Inflammation: mechanisms, costs, and natural variation. *Annu. Rev. Ecol. Evol. Syst.* **43**, 385–406 (2012).
50. Vineis, P. et al. The biology of inequalities in health: the LIFEPAATH project. *Longitud. Life Course Stud.* **8**, 33 (2017).
51. Palli, D. et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori* **89**, 586–593 (2003).
52. Firmann, M. et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
53. Pruijm, M. et al. Heritability, determinants and reference values of renal length: a family-based population study. *Eur. Radiol.* **23**, 2899–2905 (2013).
54. Reid, D. D. et al. Cardiorespiratory disease and diabetes among middle-aged male Civil Servants. A study of screening and intervention. *Lancet* **1**, 469–473 (1974).
55. Steptoe, A., Breeze, E., Banks, J. & Nazroo, J. Cohort profile: the English longitudinal study of ageing. *Int. J. Epidemiol.* **42**, 1640–1648 (2013).
56. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* **35**, 34–41 (2006).
57. McLaren, L. Socioeconomic status and obesity. *Epidemiol. Rev.* **29**, 29–48 (2007).
58. Choi, J., Joseph, L. & Pilote, L. Obesity and C-reactive protein in various populations: a systematic review and meta-analysis. *Obes. Rev.* **14**, 232–244 (2013).
59. Giskes, K. et al. Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. *J. Epidemiol. Community Health* **59**, 395–401 (2005).
60. González-Reimers, E., Santolaria-Fernández, F., Martín-González, M. C., Fernández-Rodríguez, C. M. & Quintero-Platt, G. Alcoholism: a systemic proinflammatory condition. *World J. Gastroenterol.* **20**, 14660–14671 (2014).
61. Mackenbach, J. P. et al. Inequalities in alcohol-related mortality in 17 European Countries: a retrospective analysis of mortality registers. *PLoS Med.* **12**, e1001909 (2015).
62. Hamer, M. et al. Physical activity and inflammatory markers over 10 years: follow-up in men and women from the Whitehall II cohort study. *Circulation* **126**, 928–933 (2012).
63. O'Donoghue, G. et al. Socio-economic determinants of physical activity across the life course: a 'DEterminants of DIet and Physical ACTivity' (DEDIPAC) umbrella literature review. *PLoS ONE* **13**, e0190737 (2018).
64. Matthews, S., Manor, O. & Power, C. Social inequalities in health: are there gender differences? *Soc. Sci. Med.* **48**, 49–60 (1999).
65. Khera, A. et al. Race and gender differences in C-reactive protein levels. *J. Am. Coll. Cardiol.* **46**, 464–469 (2005).
66. Lakoski, S. G. et al. Gender and C-reactive protein: data from the Multiethnic Study of Atherosclerosis (MESA) cohort. *Am. Heart J.* **152**, 593–598 (2006).
67. Wolfgang, V. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Stat.* **30**, 261–293 (2005).

Acknowledgements

This study was supported by the European Commission Horizon 2020 grant number 633666 to Paolo Vineis. Thank you to the Lifepath Consortium (see Supplementary Note 1). We are also grateful to The Centre for Longitudinal Studies, UCL Institute of Education for the use of these data and to the UK Data Archive and UK Data Service for making them available. M.K. is supported by the MRC (K013351), the Academy of Finland (311492) and NordForsk. The Centre for Environment and Health is supported by the Medical Research Council and Public Health England (MR/L01341X/1).

Author contributions

C.D., P.V., and M.K.-I. designed research; M.B., A.E., M.G., M.Ka., M.Ki., V.K., M.M., S.P., M.P., F.R., C.S., A.S., S.S., R.T., P.V. collected the data; M.B., A.E., M.G., M.Ka., M.Ki., V.K., M.M., S.P., M.P., F.R., C.S., A.S., S.S., R.T., P.V. contributed reagents, materials or analysis tools; E.B., R.C., M.C.-H., C.D. and M.K.-I. analysed the data; E.B., R.C., P.V., C.D., and M.K.-I. discussed the results and wrote the paper. All authors reviewed and commented on the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-08732-x>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: Nature Communications thanks Apostolos Davillas and Sijmen Reijneveld for their contribution to the peer review of this work. Peer reviewer reports are available.

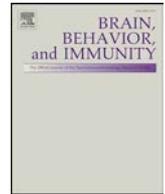
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

ANNEXE 2 - ASSOCIATION ENTRE L'EDUCATION ET L'INFLAMMATION
APPROCHEE PAR DIFFERENTS MARQUEURS INFLAMMATOIRES :
RESULTATS D'UNE ETUDE MULTI-COHORTES



Patterning of educational attainment across inflammatory markers: Findings from a multi-cohort study

Marine Maurel^{a,1}, Raphaële Castagné^{a,1}, Eloïse Berger^a, Murielle Bochud^b, Marc Chadeau-Hyam^c, Silvia Fraga^d, Martina Gandini^{e,g}, Nina Hutri-Kähönen^f, Sirpa Jalkanen^{h,i}, Mika Kivimäki^{j,k}, Michael Marmot^l, Cathal McCrory^m, Martin Preisig^b, Olli Raitakari^{n,o,p}, Fulvio Ricceri^{e,q}, Marko Salmi^{h,i}, Andrew Steptoeⁱ, Paolo Vineis^{c,r}, Cyrille Delpierre^{a,2}, Michelle Kelly-Irving^{a,*,2}

^a UMR1027, Université de Toulouse, UPS, Inserm, Toulouse, France

^b Psychiatric Epidemiology and Psychopathology Research Center, Department of Psychiatry, Lausanne University Hospital and University of Lausanne, Lausanne 1004, Switzerland

^c MRC-PHE Centre for Environment and Health, School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK

^d EPIUnit-Instituto de Saúde Pública da Universidade do Porto, Porto, Portugal

^e Epidemiology Unit, ASL TO3 Piedmont Region, Grugliasco 10095, Italy

^f Department of Pediatrics, Tampere University Hospital and Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland

^g Environmental Epidemiological Unit, Regional Environmental Protection Agency, Piedmont Region, Via Pio VII 9, 10135 Turin, Italy

^h MediCity Research Laboratory, University of Turku, Turku, Finland

ⁱ Institute of Biomedicine, University of Turku, Turku, Finland

^j Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK

^k Clinicum, Faculty of Medicine, University of Helsinki, P. O. Box 20, Helsinki FI-00014, Finland

^l UCL Institute of Health Equity, Department of Epidemiology and Public Health, UCL, 1-19 Torrington Place, WC1E 6BT, London

^m Department of Medical Gerontology, Trinity College Dublin, Ireland

ⁿ Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland

^o Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland

^p Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland

^q Department of Clinical and Biological Sciences, University of Turin, Italy

^r Italian Institute for Genomic Medicine, Torino, Italy

ARTICLE INFO

Keywords:

Inflammation
Cohort studies
Educational level
Social inequalities in health

ABSTRACT

Background: Evidence suggests that the inflammatory reaction, an adaptive response triggered by a variety of harmful stimuli and conditions involved in the risk and development of many chronic diseases, is a potential pathway through which the socioeconomic environment is biologically embedded. Difficulty in interpreting the role of the inflammatory system in the embodiment dynamic arises because of heterogeneity across studies that use a limited but varied number of inflammatory markers. There is no consensus in the literature as to which inflammatory markers beyond the C-reactive protein and to a lesser extent interleukin 6 are related to the social environment. Accordingly, we aimed to investigate the association between educational attainment, and several markers of inflammation – C-reactive protein, fibrinogen, interleukin 6, interleukin 1 β and tumor necrosis factor α – in 6 European cohort studies.

Methods: Up to 17,470 participants from six European cohort studies with data on educational attainment, health behaviors and lifestyle factors, and at least two different inflammatory markers. Four sub-datasets were drawn with varying numbers of participants to allow pairwise comparison of the social patterning of C-reactive protein and any other inflammatory markers. To evaluate within each sub-dataset the importance of the context and cohort specificities, linear regression-based analyses were performed separately for each cohort and combined in a random effect meta-analysis to determine the relationship between educational attainment and inflammation.

Results: We found that the magnitude of the relationship between educational attainment and five inflammatory biomarkers (C-reactive protein, fibrinogen, interleukin 6 and 1 β and tumor necrosis factor α) was variable. By

* Corresponding author.

E-mail address: michelle.kelly@inserm.fr (M. Kelly-Irving).

¹ These authors contributed equally.

² These authors jointly supervised this work.

far the most socially patterned biomarker was C-reactive protein, followed by fibrinogen and to lesser extent interleukin 6, where a low educational attainment was associated with higher inflammation even after adjusting for health behaviours and body mass index. No association was found with interleukin 1 β and tumor necrosis factor α .

Conclusions: Our study suggests different educational patterning of inflammatory biomarkers. Further large-scale research is needed to explore social differences in the inflammatory cascade in greater detail and the extent to which these differences contribute to social inequalities in health.

1. Introduction

Inflammation is an important and necessary component of the immune system broadly defined as the body's protective reaction in response to infection or injury. A successful acute inflammatory response results in the elimination of the inciting stimulus followed by a resolution and repair phase in order to restore homeostasis (Medzhitov, 2008). Much less is known about the causes and mechanisms of chronic systemic inflammation, it could occur when the organism has to adapt when confronted with homeostatic imbalances (Medzhitov, 2008). Chronic systemic inflammation in response to persistent threats could contribute to disease progression through dysregulation of homeostatic functions, control of the metabolism, tissue repair, and of the hypothalamus-pituitary axis (Straub, 2017; Bennett et al., 2018; Franceschi et al., 2018). As such, chronic inflammation provides a unifying pathophysiological mechanism involved in many chronic diseases, including diabetes and obesity, cardiovascular disease, certain cancers and inflammatory bowel diseases, arthritis, and osteoporosis (Bennett et al., 2018; Furman et al., 2019).

Because the inflammatory response may be influenced by a large range of environmental challenges, an overall increase in systemic inflammation has emerged as one possible physiological pathway through which the social environment 'gets under the skin' to affect biological processes and confers a risk for negative health outcomes (Miller et al., 2009). This refers to a dynamic process, namely the biological embodiment of the social environment, or the social-to-biological transition (Blane et al., 2013). First, socially disadvantaged groups are more likely to live in areas where they are disproportionately exposed to harm, such as air-pollution and damp housing, overcrowded conditions, poor housing quality, or insufficient access to sanitation, that are likely to increase systemic inflammation (Gares et al., 2017). Second, adverse health behaviours, including smoking, alcohol consumption and nutrition are more prevalent among socially disadvantaged groups (Petrovic et al., 2018) and are known to influence inflammation. Third, since socially disadvantaged groups are more likely to be exposed to a higher burden of exposures in different domains of everyday life including stressful life situations, negative life events, family stress or adversities (Fagundes et al., 2013), social disadvantage across the life course may result in an over-solicited stress response involving the inflammatory system. This nexus of dynamic processes may contribute to heightening the level of basal inflammation. Therefore chronic inflammation may be crucial in understanding social inequalities in health by affecting levels of inflammatory cytokines and acute-phase proteins (Kivimäki and Steptoe, 2018).

Numerous studies have examined the association between various indicators of socioeconomic position (SEP), such as education, occupation and income, and markers of systemic inflammation. Most research has focused on C reactive protein (CRP), an acute-phase protein released by the liver in response to interleukin 1, 6 and 7 (IL-1, IL-6 and IL-7) (Eklund, 2009). In 2007, a systematic review of 25 population-based studies reported an inverse relationship between adult SEP and CRP across countries independently of age, sex, body mass index (BMI), and smoking, that where controlled for in most studies (Nazmi and Victora, 2007). More recently, a meta-analysis of 43 North American studies including sub-sets of investigations on IL-6 (N = 18 studies) and CRP (N = 35 studies) found a higher level of CRP and IL-6 among

socially disadvantaged participants compared to their advantaged counterparts (measured by educational attainment or household income). The association with CRP was partly explained by obesity and smoking while the association with IL-6 weakened with age but was not affected by BMI and smoking (Muscatell et al., 2018). Among the other biomarkers of inflammation, fibrinogen, an acute-phase protein also released by the liver, is less frequently used. Panagiotakos et al. (2004), Pollitt et al. (2008), Steinvil et al. (2008) and Fraga et al. (2015) found a higher level of fibrinogen in participants with lower education. While this relationship was attenuated by health behaviours and cardiovascular risk factors, evidence of the association remained after adjustment (Pollitt et al., 2008; Steinvil et al., 2008). A limited number of studies have examined another inflammatory biomarker, tumour necrosis factor α , TNF- α (Steptoe et al., 2002; Koster et al., 2006; Fraga et al., 2015), an important immune mediator secreted by macrophages and monocytes released after a stimulus (i.e. in response to IL-1). For this inflammatory marker the results are not conclusive (Steptoe et al. (2002) and Koster et al. (2006) Fraga et al. (2015).

There is no consensus in the literature about the most relevant inflammatory markers in terms of social patterning nor standard biomarkers that would be relevant for indicating the presence of health-damaging chronic inflammation (Furman et al., 2019). A difficulty in interpreting the role of the inflammatory system in the embodiment of SEP arises because of the heterogeneity of studies that use a limited but varied number of inflammatory markers. Furthermore, cross-country comparison is limited by the variety of SEP indicators used, their meaning in different national contexts, and the lack of harmonisation of both SEP and inflammatory markers across studies. Additionally, evidence on the pathways relating SEP to inflammatory markers across countries is limited due to the fact that behavioural and personal characteristics that may mediate the association between SEP and inflammation are inconsistently taken into account. It is therefore unclear whether the association between lower SEP and chronic inflammation is a stable phenomenon across countries, persisting even after adjusting for health behaviours, and whether this relationship depends on the biomarkers used to measure inflammation.

We previously reported a consistent inverse association between SEP, (using father's occupation, education, and last occupation) and CRP using data from six European cohort studies from three countries (Berger et al., 2019). In that study, educational attainment was the SEP marker most strongly related to CRP level in adulthood, after adjusting for health behaviours, body mass index and later-in-life SEP. Educational attainment captures components of the social environment from earlier in life, and is associated with later occupational social class and income (d'Errico et al 2017). It is also a useful variable when comparing SEP across countries or cohorts because it is often relatively consistently collected, is usually available and can be harmonised across contexts. As with all SEP proxy variables, it must be used while remaining aware of local time, place and cohort differences. Based on these findings and to capture the complex regulatory cascades involved in inflammation, the present study aims to describe in detail the educational patterning of several different inflammatory markers. To achieve this, we assessed the relationship between educational attainment and CRP, fibrinogen, IL-6, IL-1 β and TNF- α – using harmonised data across six European cohort studies within the Lifepath consortium. We also described these associations adjusted for behavioural variables to examine the patterns

of adjusted associations between education and each inflammatory biomarker.

2. Methods

2.1. Study population

This study is part of a Horizon 2020 Research and Innovation Program European funded project, the Lifepath project (<http://www.lifepathproject.eu/>), which includes a consortium of eighteen cohort studies across different countries and time period, with demographic, clinical, biological and socioeconomic data, with the aim to investigate biological pathways underlying social differences in healthy ageing. Data harmonization were performed in order to pool and analyse the cohorts together (Vineis et al., 2017). We selected the six cohort studies for which data on at least two different measures of inflammatory markers including CRP, educational attainment, smoking, alcohol consumption, physical activity and BMI were available. This selection allowed for sets of comparative analyses wherein one consistent inflammatory biomarker (CRP) was common across them all. The selected cohorts were:

- one cohort from Finland: the Young Finns study: YFS (Raitakari et al., 2008)

- two cohorts from Switzerland: SKIPOGH (Alwan et al., 2014) and CoLaus (Firmann et al., 2008)

- three British cohorts: Whitehall II (Marmot et al., 1991), the National Child Development Study, NCDS (Power and Elliott, 2006) and the English Longitudinal Study of Ageing, ELSA (Stephoe et al., 2013). A detailed description of all cohorts included in the Lifepath consortium has been given elsewhere (Vineis et al., 2017).

2.2. Educational attainment

We used the harmonised definition of educational attainment across cohort study described in details by d'Errico et al. (d'Errico et al., 2017). Briefly, participant's educational attainment was categorised in three groups (i) 'low' (no school, primary or lower secondary school: from 7 to 9 years after kindergarten with a basic curriculum in languages, mathematics and other subjects), (ii) 'medium' (higher secondary school: around 4–5 years more, high school diploma level) and (iii) 'high' educational level (tertiary education: any degree after high school, such as BSc, MSc, and further education).

2.3. Inflammatory markers

We used five biomarkers to measure inflammation: CRP, fibrinogen, IL-6, IL-1 β and TNF- α ; selecting the first available wave in which CRP and any other inflammatory markers had been measured. In CoLaus and Skipogh, CRP, IL-6, IL-1 β and TNF- α were collected at baseline while in ELSA, both CRP and fibrinogen were collected in the second wave and in Whitehall, CRP, fibrinogen and IL-6 were measured in the third data of collection wave. CRP and fibrinogen were measured when participants were 45 years old in NCDS, with the exception of YFS for which we selected measures at the last available wave in order to obtain an age range among participants comparable to the age of participants in the other cohort studies. For each cohort study, a detailed description of sampling methods and laboratory analyses is given in Supplementary Table 1. Given their skewed nature, CRP, fibrinogen, IL-6 and TNF- α values were natural log-transformed to normalise distributions and because of both the skewness of the distribution and the high number of 0 values (41.3%), a constant (1) was added to IL-1 β values prior to the log transformation.

While the association between social measures and CRP is widely described, examining other components of the inflammatory system using different inflammatory biomarkers would further our understanding of potential social-to-biological processes. The inflammatory

biomarkers we selected for this analysis were available across some of the cohorts, but they also represent different stages of inflammatory responses.

IL-6, IL-1 β and TNF- α are all pro-inflammatory cytokines produced within adipose tissue (Coppack, 2001), their synthesis being related to adiposity and fatty acid metabolism. TNF- α also down-regulates insulin-stimulated glucose uptake, all three appear to act upon adipose tissue regulation to different extents (Coppack 2001). These markers are upstream of CRP, namely, IL-6 is the primary cytokine leading to hepatic CRP production (Ridker 2016). These biomarkers are indeed related, however are also implicated in different inflammatory mechanisms. The importance of the IL-6, IL-1 β and TNF- α cytokines in adipose tissue metabolism may be an important consideration for social-to-biological questions, given the well-established socioeconomic patterns in overweight and obesity prevalence.

Fibrinogen however is produced in the liver, forming part of the clotting cascade. It represents a link between inflammation, thrombosis and cardiovascular risk, and is up-regulated during the acute phase inflammatory response driven upstream by cytokines such as IL-6 (Rooney et al 2011). While fibrinogen and CRP have similarities in terms of their synthesis and place in the cascade, fibrinogen also acts as a marker for haemostatic processes. The role of fibrinogen may be interesting as a potential social-to-biological mediator given the socioeconomic pattern of cardiovascular disease.

2.4. Covariates

Pre-defined, harmonised covariates included potential confounders, such as age, sex and potential mediators were alcohol consumption, physical activity, smoking and body mass index (BMI). These variables were selected from the closest data collection wave to that of the inflammatory marker measurements.

Alcohol consumption was harmonised in units of pure alcohol per week and we categorised participants as abstainers (0 units per week), moderate consumers (men \leq 21 alcohol units per week; women \leq 14 alcohol units per week) and high consumers (men $>$ 21 alcohol units per week; women $>$ 14 alcohol units per week). Because the physical activity was measured with different questions in each study, sedentariness was derived from leisure physical activity and dichotomised into sedentary yes/no, except in the NCDS cohort in which this variable was not available. Self-reported smoking was categorised into never, current and former smoker, except in YFS in which a binary smoking status indicator (never versus ever smoker) was available. BMI was calculated from height and weight and categorised into the WHO BMI groups using standard cut-offs: underweight ($<$ 18.5 kg/m²), normal (18.5–24.9 kg/m²), overweight (25.0–29.9 kg/m²) and obese ($>$ 30 kg/m²). If data were missing at the same wave, we used data from the next available data wave (for CoLaus: $n = 12$, 0.2%; ELSA: $n = 66$, 1.7%; Skipogh: $n = 266$, 28.4%; Whitehall: $n = 374$, 5.7%).

2.5. Ethics statement

Each study was approved by the relevant local or national ethics committees/board and all participants gave informed consent to participate.

2.6. Statistical analysis

In total, four sub-datasets (one for each CRP plus one other inflammatory marker-pair[either fibrinogen, IL-6 TNF- α or IL-1 β]) with varying numbers of participants were created: sub-dataset 1 included cohorts that measured inflammation using CRP and fibrinogen (3 studies); sub-dataset 2 included cohorts that measured inflammation using CRP and IL-6 (4 studies); sub-dataset 3 included cohorts that measured inflammation using CRP and TNF- α (3 studies); and sub-dataset 4 included cohorts that measured inflammation by CRP and IL-1 β (2

studies).

For the bivariate analysis in each sub-dataset and cohort, means and frequencies were reported for all continuous and categorical characteristics by educational attainment. We used Chi-squared test or Fisher exact test for the categorical variables and T-test or Wilcoxon rank test for continuous variables to estimate bivariate associations.

For the multivariate analysis in each cohort, linear regression models were used to assess the relationship between educational attainment and each inflammatory marker. In a basic statistical model, effect estimates were adjusted for age (continuous) and sex (for NCDS, sex only as participants are the same age) (Model 1). To examine whether the association between educational attainment and inflammation was mediated by alcohol consumption, smoking status, sedentary behaviour and BMI, we defined a second model controlling for all potential intermediate factors (Model 2). Sedentary behaviour was not available in the NCDS, therefore Model 2 for this cohort adjusted only for alcohol consumption, smoking status and BMI. The high educational level group was used as the reference category; therefore a positive regression coefficient indicates an increased level of inflammatory marker in participants with lower educational attainment.

As proposed elsewhere (Berger et al., 2019), we used a meta-analysis in each sub-dataset, based on random effects modelling with a restricted maximum likelihood estimator using the *metafor* R package to combine the results from each cohort within each sub-dataset. The combined effect represents the mean of the population of ‘true effects’. Test of heterogeneity among studies was provided by the Q-statistic and I^2 measurement was calculated representing the proportion of total variability that can be attributed to heterogeneity (a higher value indicates a greater degree of heterogeneity).

In sensitivity analyses, we examined the SEP- inflammatory marker association separately in men and women. We also used an alternative statistical transformation for the values of each inflammatory marker. First, each inflammatory marker was categorised into quartiles. Then

we defined a binary indicator ‘high’ vs ‘low’ concentration based on the highest quartiles (4th quartile) vs. the three lower quartiles (1st – 3rd quartiles). To evaluate the association between SEP and the inflammatory markers (high concentration = 1 and low concentration = 0) we used a logistic regression and the results were expressed as odds ratios (OR) and 95% confidence intervals (95% CI). We performed additional sensitivity analyses restricting our study population to never smoked participants or participants with a normal BMI [18.5–24.9 kg/m²].

All statistical analyses were performed using R (R Core Team, 2014) version 3.1.2 within the Rstudio environment (RStudio Team, 2016).

3. Results

3.1. Study population

Table 1 presents a description of the variables for each sub-dataset. Sub-dataset 1 only included participants from the three British cohort studies (37.4%, 21.7%, 40.9% for Whitehall, ELSA and NCDS respectively) while sub-dataset 2 included participants from 4 out of the 6 cohorts (43.5%,38.7%;11.5%,6.2% for Whitehall, Colaus, YFS and Skipogh respectively). Sub-dataset 3 and 4 were mostly constituted of Swiss participants (79.6% and 100% respectively) with some participants from YFS in dataset 3. Comparisons between each sub-dataset are given in Supplementary Table 2 and show small to moderate differences of the participants’ key characteristics: the proportion of participants with a lower educational attainment ranged from 59.1% in sub-dataset 1 to 45.8% in sub-dataset 2. Irrespective of the sub-dataset considered, participants with a low educational attainment were more likely to smoke and to be overweight/obese and sedentary compared to the participants with a higher educational attainment.

Description of participants’ characteristics by cohort is provided in Supplementary Table 3 and 4. A total of 25 955 participants were

Table 1
Summary characteristics of the participants for each sub-dataset μ Smoking status in YFS was categorised as ever vs never, *sedentary not available in NCDS.

		Sub-dataset 1 (N = 17 470)	Sub-dataset 2 (N = 15 019)	Sub-dataset 3 (N = 8 485)	Sub-dataset 4 (N = 6751)
		CRP-Fibrinogen	CRP-IL6	CRP-TNFA	CRP-IL1B
		Mean (SD) or N(%)	Mean (SD) or N(%)	Mean (SD) or N(%)	Mean (SD) or N(%)
Cohorts		Whitehall, NCDS, ELSA	Whitehall, Colaus, YFS, Skipogh	Colaus, YFS, Skipogh	Colaus, Skipogh
Demographics					
Age		51.3 (8.8)	49.6 (10.2)	49.3 (12.5)	52.3 (12.1)
Sex	Men	9875 (56.53)	8549 (56.92)	4012 (47.28)	3244 (48.05)
	Women	7595 (42.47)	6470 (43.08)	4473 (52.72)	3507 (51.95)
Inflammatory level					
CRP (mg/L)		2.4 (5.0)	2.2 (3.9)	2.4 (3.7)	2.5 (3.8)
Fibrinogen (g/L)		2.8 (0.7)			
IL6 (pg/mL)			5.6 (35)		
TNFA (pg/mL)				16.4 (45.9)	
IL1B (pg/mL)					4.3 (16.3)
Socioeconomic position					
Educational attainment	High	4174 (23.89)	5260 (35.02)	2980 (35.12)	1900 (28.14)
	Middle	2979 (17.05)	2885 (19.21)	1132 (13.34)	1077 (15.95)
	Low	10,317 (59.06)	6874 (45.77)	4373 (51.54)	3774 (55.9)
Intermediate factors					
Alcohol consumption	Abstainer	4316 (24.71)	3748 (24.96)	2541 (29.95)	1862 (27.58)
	Low	9214 (52.74)	8914 (59.35)	4624 (54.5)	4227 (62.61)
	High	3940 (22.55)	2357 (15.69)	1320 (15.56)	662 (9.81)
Smoking status [‡]	Never	8076 (46.23)	6848 (45.60)	3616 (42.62)	2741 (40.6)
	Former	6171 (35.32)	4592 (30.57)	2185 (25.75)	2185 (32.37)
	Current	3223 (18.45)	2720 (18.11)	1825 (21.51)	1825 (27.03)
	Ever		859 (5.72)	859 (10.12)	
BMI	Underweight	75 (0.43)	129 (0.86)	94 (1.11)	80 (1.19)
	Normal weight	7888 (45.15)	7416 (49.38)	4007 (47.22)	3175 (47.03)
	Overweight	6825 (39.07)	5527 (36.8)	3060 (36.06)	2446 (36.23)
	Obese	2682 (15.35)	1947 (12.96)	1324 (15.6)	1050 (15.55)
Sedentary	No	8106 (46.40)	11,089 (73.83)	5862 (69.09)	4339 (64.27)
	Yes	2215 (12.68)	3930 (26.17)	2623 (30.91)	2412 (35.73)
	Missing	7149 (40.92)			

included (53.5% of men). YFS was the youngest cohort with a mean age of 37.7 (SD 5.0), while ELSA was the oldest with a mean age of 64.0 (SD 8.5). Whitehall had the smallest proportion of women (30.6%) compared to other cohort studies. The percentage of participants with a low educational attainment ranged from 34.5% (YFS) to 76.9% (NCDS). YFS had the largest proportion of participants (37.9%) that reported drinking more than 21 or 14 alcoholic units per week respectively for men and women. The proportion of participants who had never smoked varied from 39.9% (CoLaus and ELSA) to 50.5% (YFS). There was 28.2% of obese in ELSA and 39.4% of Skipogh participants were sedentary.

3.2. Association between educational attainment and each inflammatory marker in each sub-dataset

Table 2 and Fig. 1 present the linear regression results from the random effects meta-analyses of educational attainment and each inflammatory marker in each sub-dataset. Despite the moderate to high degree of heterogeneity in study-specific estimates, we observed an association between educational attainment and CRP in Model 1 in all sub-datasets: the difference in the log of CRP concentration between the lowest and the highest educational group was 40% [24%; 60%] in sub-dataset 1; 27% [18%; 36%] in sub-dataset 2; 27% [14%; 42%] in sub-dataset 3 and 35% [27%; 43%] in sub-dataset 4. After additional adjustment for alcohol consumption, smoking status, sedentary behaviour and BMI (Model 2), these observed associations were weakened: 19% [9%; 30%] in sub-dataset 1; 11% [7%; 15%] in sub-dataset 2; 11% [6%; 17%] in sub-dataset 3 and 12% [6%; 19%] in sub-dataset 4. Except in sub-dataset 3, we observed a weaker association between a medium level of education and CRP compared to those with a low level of education and CRP (Model 1) and, except in the sub-dataset 1, the association weakened further after controlling for the four intermediate factors (Model 2).

In the sub-dataset 1 including CRP and fibrinogen, we also observed an association between educational attainment and fibrinogen: participants with a low level of education had higher levels of fibrinogen compared to those with a high level of education, the difference was 5% [3%; 6%]. This association was attenuated after adjustment for the four intermediate factors in Model 2 (2% [1%; 3%]). We observed a weaker association between a medium level of education and fibrinogen (Model 1, 2% [0%; 4%]) which weakened further in Model 2 (1% [-1%; 2%]).

In sub-dataset 2 including CRP and IL6, results for IL-6 were weaker but directionally consistent: a higher level of IL-6 was found in participants with a lower educational attainment (Model 1, 7% [-1%; 16%]) mostly explained by the intermediate factors (Model 2, 2% [-1%; 6%]).

There was weak evidence of an association between education and TNF-α and IL-1β in sub-datasets 3 and 4 respectively.

3.3. Pattern of inflammatory markers by educational attainment in each cohort study

Association results by cohort study are detailed in Supplementary Table 5 and presented in Fig. 1. We observed an inverse association of educational attainment with CRP, fibrinogen and IL-6 in Whitehall, where participants with a low level of education were more likely to have higher levels of all three inflammatory markers compared to those with a high level of education (Model 1: β [95%CI] = 0.22 [0.15, 0.29] for CRP, β [95%CI] = 0.03 [0.02, 0.05] for fibrinogen, β [95%CI] = 0.08 [0.05, 0.12] for IL-6). These associations were attenuated after adjustment for intermediate factors (β [95%CI] = 0.10 [0.03, 0.17] for CRP, β [95%CI] = 0.02 [0.002, 0.03] for fibrinogen and β [95%CI] = 0.04 [0.01; 0.08] for IL-6).

Participants with a low level of education had a higher level of both CRP and fibrinogen compared to those with a high level of education in

NCDS (β [95%CI] = 0.056 [0.044, 0.068]) and in ELSA (β [95%CI] = 0.049 [0.03, 0.069]). These associations were attenuated after additional adjustment for intermediate factors (β [95%CI] = 0.025 [0.013, 0.037]) and in ELSA (β [95%CI] = 0.019 [-0.001, 0.038]).

In Skipogh in which CRP, IL-6, TNF-α and IL-1β were measured, only CRP and IL-6 were associated with educational attainment (Model 1: β [95%CI] = 0.259 [0.073, 0.444] for CRP and β [95%CI] = 0.213 [0.099; 0.327] for IL-6). These associations were attenuated upon adjustment for intermediate factors for IL-6 (Model 2: β [95%CI] = 0.108 [-0.076, 0.293] for CRP and β [95%CI] = 0.105 [-0.007, 0.217] for

Table 2 Results from the meta-analyses of the association between educational attainment and each inflammatory markers in each sub-dataset.

Sub-dataset 1 (N = 17 470)	Ref.: High	CRP		Fibrinogen	
		Medium	Low	Medium	Low
Model 1 [□]	% Change in	21% (9%; 35%)	40% (24%; 60%)	2% (0%; 4%)	5% (3%; 6%)
	β (95% CI)				
	P-value	< 0.001	< 0.001	0.063	< 0.001
	I ²	65.738	87.492	70.244	65.555
	P _H	0.043	< 0.001	0.034	0.043
Model 2 ^γ	% Change in	10% (3%; 18%)	19% (9%; 30%)	1% (-1%; 2%)	2% (1%; 3%)
	β (95% CI)				
	P-value	0.005	< 0.001	0.466	< 0.001
	I ²	30.685	75.076	43.945	0.000
	P _H	0.275	0.011	0.164	0.534
Sub-dataset 2 (N = 15 019)		CRP		IL-6	
		Medium	Low	Medium	Low
Model 1 [□]	% Change in	9% (3%; 15%)	27% (18%; 36%)	4% (-2%; 11%)	7% (-1%; 16%)
	β (95% CI)				
	P-value	0.002	< 0.001	0.207	0.091
	I ²	0.080	58.901	52.308	88.360
	P _H	0.214	0.067	0.109	< 0.001
Model 2 ^γ	% Change in	4% (-2%; 9%)	11% (7%; 15%)	3% (-3%; 9%)	2% (-1%; 6%)
	β (95% CI)				
	P-value	0.177	< 0.001	0.414	0.241
	I ²	0.013	0.000	47.679	39.609
	P _H	0.353	0.945	0.126	0.137
Sub-dataset 3 (N = 8 485)		CRP		TNF-α	
		Medium	Low	Medium	Low
Model 1 [□]	% Change in	6% (-5%; 18%)	27% (14%; 42%)	2% (-7%; 11%)	0% (-5%; 6%)
	β (95% CI)				
	P-value	0.284	< 0.001	0.740	0.981
	I ²	23.749	63.734	51.580	58.491
	P _H	0.151	0.052	0.126	0.088
Model 2 ^γ	% Change in	1% (-6%; 9%)	11% (6%; 17%)	1% (-7%; 10%)	-1% (-5%; 4%)
	β (95% CI)				
	P-value	0.743	< 0.001	0.741	0.730
	I ²	0.023	0.000	47.237	33.705
	P _H	0.267	0.834	0.156	0.279
Sub-dataset 4 (N = 6 751)		CRP		IL-1β	
		Medium	Low	Medium	Low
Model 1 [□]	% Change in	9% (-1%; 20%)	35% (27%; 43%)	6% (-1%; 15%)	2% (-3%; 8%)
	β (95% CI)				
	P-value	0.086	< 0.001	0.105	0.391
	I ²	13.953	0.000	0.000	0.000
	P _H	0.281	0.637	0.746	0.968
Model 2 ^γ	% Change in	3% (-6%; 14%)	12% (6%; 19%)	7% (-1%; 15%)	4% (-2%; 10%)
	β (95% CI)				
	P-value	0.511	< 0.001	0.081	0.204
	I ²	20.499	0.000	0.000	0.000
	P _H	0.262	0.959	0.697	0.819

Abb.: CI confidence interval, I² percentage of between study heterogeneity, P_H P-value of heterogeneity test.

[□] Model 1 adjusted for age and sex.

^γ Model 2 controlled for age, sex and additionally alcohol, smoking, body mass index and sedentary.

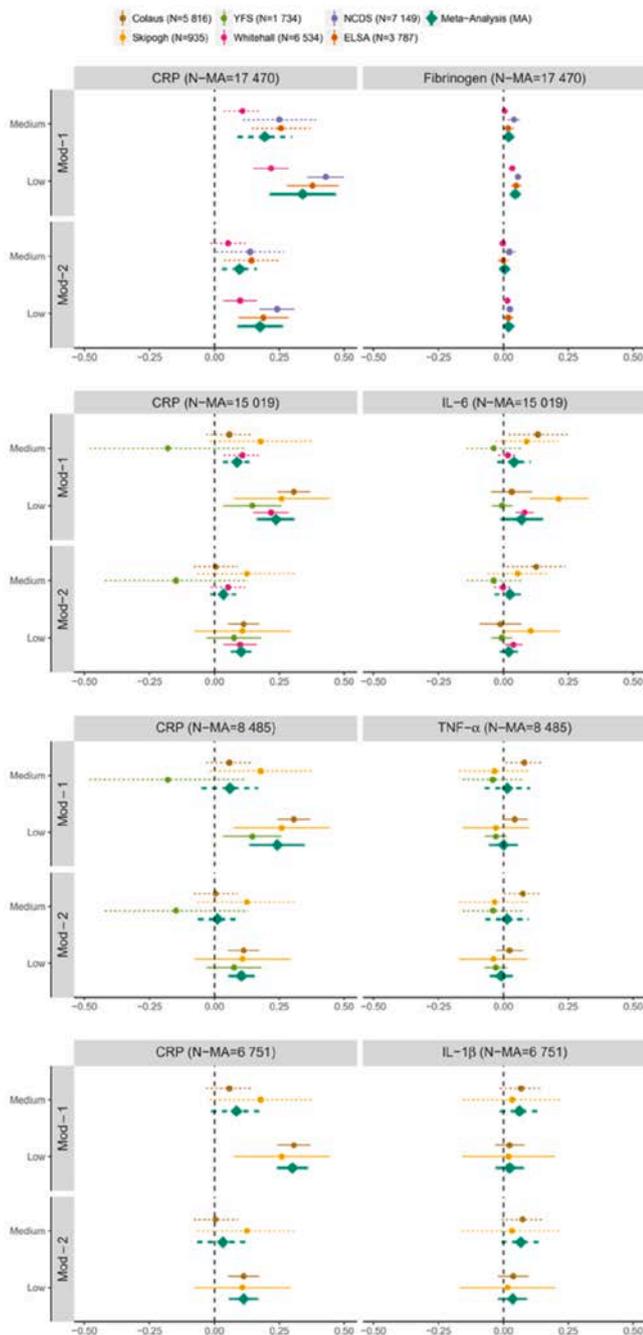


Fig. 1. Forest plot of regression coefficients [95% confidence interval] for the association between participant’s educational attainment and inflammatory markers concentration by cohort and in random effect meta-analysis framework for each sub-dataset for Model 1 (Mod-1), after adjustment for each intermediate factor (Mod-2). The high SEP group was used as reference, dotted lines represent the medium SEP group and solid lines the low SEP group.

IL-6).

In both Colaus and YFS, only the level of CRP was patterned by educational attainment whereas no associations between educational attainment and IL-6, TNF- α and IL1- β were observed in Colaus. Similarly in YFS, no associations were observed for the other inflammatory markers (IL-6 and TNF- α).

3.4. Sensitivity analyses

Sex-stratified results from random-effects meta-analysis are provided in Supplementary Table 6. In sub-dataset 1, the association

between educational attainment and CRP was attenuated upon adjustment for the four intermediate factors in men (β [95%CI] = 0.177 [0.102; 0.252]) but not in women (Model 2, β [95%CI] = 0.14 [-0.026; 0.306]). An association between educational attainment and CRP after adjustment on the four intermediate factors was observed in men and women in the other three sub-datasets.

In sub-dataset 1, we also observed an association between educational attainment and fibrinogen in both men (β [95%CI] = 0.044 [0.026, 0.062]) and women (β [95%CI] = 0.049 [0.036, 0.062]). Evidence of these associations, though attenuated, remained after adjustment for the four intermediate factors in Model 2 (β [95%CI] = 0.022 [0.011, 0.033] in men and β [95%CI] = 0.014 [0.001, 0.026] in women).

In both men and women, there was no association between educational attainment and IL-6, TNF- α and IL-1 β ($P > 0.05$ in all settings) in sub-datasets 2, 3 and 4 respectively.

Our main results were robust when using an alternative transformation of each inflammatory marker (see Methods, Supplementary Fig. 1 and Supplementary Table 7). In additional sensitivity analyses, restricting our study population to ‘never smoked’ participants or participants with a normal BMI size reduced statistical power and measures of association were subsequently weakened but did not materially affect our main results and conclusion (Supplementary Table 8 and 9).

4. Discussion

Our study aimed to capture the association between education, and the complex regulatory cascades involved in inflammation. We found that the magnitude of the relationship between educational attainment and five inflammatory biomarkers (CRP, fibrinogen, IL-6, IL-1 β and TNF- α) was variable, and by far the most socially patterned biomarker was CRP, followed by fibrinogen and to lesser extent IL-6, where low educational attainment was associated with higher inflammation. Results from the meta-analysis for TNF- α and IL-1 β were null and thus not consistent with the other inflammatory biomarkers. Country specific results suggest that the overall direction of the association was broadly similar across available biomarkers (Whitehall, ELSA and NCDs), although some heterogeneity in results was found between cohorts (Colaus, Skipogh and YFS).

The association between educational attainment and CRP or fibrinogen in adulthood was only partly attenuated after accounting for behavioural factors (smoking status, alcohol consumption and sedentary) and BMI suggesting that (i) educational attainment may be a consistent and important upstream independent variable for elevated inflammation and that (ii) educational attainment may also operate through intermediate pathways other than health behaviours which deserve attention in future analyses. These could for include, for example, psychosocial pathways such as adverse childhood experiences (Lacey et al., 2020), loneliness (Vingeliene et al., 2019; Smith et al., 2020), and social relationships (Yang et al., 2014; Smith et al., 2020) might play a role in the relationship between education and inflammation. Our aim in this paper was to describe the association between education and inflammation using different inflammatory biomarkers. Our rationale for adjusting the models was to describe and compare the patterns of adjustment, providing additional descriptive information about associations between education and inflammation. The covariates may be on the causal pathway between education and inflammation, however, to establish mechanisms and pathways further analyses using causal inference approaches is required.

4.1. Prior literature

Our findings are in agreement with data-driven analyses on diseases across all organ systems that have linked low SEP to increased risk of health conditions characterized by elevated systemic inflammation, such as diseases of the pancreas, liver, kidney, vascular and respiratory

system, lung cancer, and dementia (Kivimäki et al., 2020). In line with a number of studies on inflammatory markers, our results support studies reporting an association between disadvantaged SEP and high CRP levels (Nazmi and Victora, 2007; Muscatell et al., 2018) and further add to the empirical evidence supportive of the social patterning of fibrinogen (Davillas et al., 2017; Panagiotakos et al., 2004; Pollitt et al., 2008; Steinvil et al., 2008). The weaker association observed between educational attainment and IL-6 is in contrast with another recently published meta-analysis that reported a social gradient for IL-6 (Muscatell et al., 2018). It is noteworthy that their work relied mainly on North American studies, whereas only European studies were included in our study. Little is known about the association between SEP and the other inflammatory markers addressed in our work. The present results indicated no association between low educational attainment and higher level of IL-1 β or TNF- α in line with the somewhat conflicting results from previous research work (Fraga et al., 2015; Koster et al., 2006; Steptoe et al., 2002). Future research including other novel biomarkers of inflammation such as Glyc-A would also make important contributions to understanding the relationship between social factors and the inflammatory system.

4.2. Cohort specificities

Because educational attainment reflects the broad socio-economic, cultural and historical context in which individuals live and grow, we also described the educational patterning of inflammation in each cohort study from the UK, Switzerland and Finland covering different contextual times and place. Our findings were directionally consistent across country for both CRP and fibrinogen and to lesser extent for IL-6 suggesting some shared social to biological processes across these European populations. The discrepancies between cohorts observed for IL-1 β and TNF- α suggests that some more context specific mechanisms may be at play in the biological embodiment of the social environment. Alternatively, these discrepancies may reflect differences in study protocols and methodology between the cohort studies.

There is no consensus in the literature about the most relevant inflammatory markers to use for analysing the social patterning of inflammation. Our study shows that CRP and fibrinogen are consistently socially patterned suggesting that these two biomarkers have special status in the physiological cascade linking SEP with health. However, the fact that the relationship between SEP and inflammation varies according to biomarkers used deserves to be further studied.

4.3. Mechanisms

CRP and fibrinogen, both synthesised by the liver and non-specific, were found to be socially patterned in our study. These inflammatory markers do not provide information about the type of the inflammatory response, nor the initial trigger of the inflammatory cascade. However, in this framework it may be justified to hypothesise that CRP and fibrinogen are maintenance proteins with both immune and non-immune functions suggesting that elevated CRP may favour the organism's homeostasis and its somatic maintenance (Del Giudice and Gangestad, 2018). When inflammation is triggered, IL-6, TNF- α and IL-1 β are released into circulation (by macrophages and other cells) to exert their specific and various roles in inflammation. The different social patterning observed for these three inflammatory markers may reflect not only different functions but also their different sources of production. Our work suggests that the inflammatory response of the social environment may be a demonstration of one out of a variety of embodiment processes, in line with the general notion that humans respond also at a biological level to their social environment. However, this may explain only a proportion of the educational patterning in poor health outcomes, an issue that warrants further research.

4.4. Weaknesses and strengths

The main limitation of this study was that we did not have exactly the same set of inflammatory markers across the 6 cohort studies which constrained our methodological choices: the use of 4 sub-datasets made the comparisons of findings across the sub-dataset non-optimal. The six cohorts we used may not be representative of the general population and are susceptible to attrition leading to a potential under-representation of the most extreme socio-economic groups. Additionally, in each cohort study, health behaviours and educational attainment were self-reported which may result in misclassification. The variables were then harmonised which may indicate that some cohort specificities are attenuated or lost, leading to a certain degree of residual confounding. We cannot rule out the possibility that other factors may contribute to the mechanisms linking SEP and inflammation. These factors may include psychological factors, chemical exposures or infectious diseases that are related to both SEP and inflammation. Indeed, environmental pollutants as well as infectious and chronic disease prevalence across the lifecourse and other exposures may also influence the association between education and inflammation with country/cohort variability. From a social perspective information on race/ethnicity or migration status which may be confounding factors in an association between education and inflammation were not available and may contribute to unmeasured confounding.

Regarding the imputation of missing data, the last observation carried forward (LOCF) approach reduces within-individual variability which leads to underestimating error. This can result in falsely finding 'statistically significant' associations. Given the relatively low proportion of missing values (around or below 5% in all studies except SKIPOGH), the influence of missing data or the imputation method we used on the associations we found in our work is probably weak.

Important strengths include the opportunity of using harmonised data from 6 well-characterised cohort studies on participants from three different European countries. Four sub-datasets were derived from these 6 cohort studies allowing the pairwise comparison of the social patterning of CRP and any other inflammatory markers; fibrinogen in sub-dataset 1, IL-6 in sub-dataset 2, TNF- α in sub-dataset 3 and IL-1 β in sub-dataset 4. The availability of these data allows us to begin elucidating the complex association between the social environment and other inflammatory-related biomarkers beyond CRP from the complex regulatory cascades involved in inflammation. These relationships were modelled separately for each cohort study and examined jointly using a random effect meta-analysis. The stability of our results was further confirmed by the multiple sensitivity analyses carried out.

5. Conclusion

In this multi-cohort study, we provide a thorough examination of SEP inequalities in five inflammatory markers (CRP, fibrinogen, IL-6, IL-1 β and TNF- α) using data from several European countries. This work contributes to a better understanding of the socioeconomic patterning of the inflammatory response and suggests that CRP and fibrinogen may be particularly influenced by SEP. The finding that the relationship between educational attainment and each of five inflammatory biomarkers investigated was variable suggests that the way SEP is embodied through inflammation may involve complex processes. These results point to several promising directions to explore chronic inflammation, including the investigation among the full human set of inflammatory-related molecules and events to identify which ones are activated in response to the social environment, at which life stages and to which extent they contribute to explain social inequalities in inflammatory related-diseases. This study calls for further research to better understand how the inflammatory cascade, in response to the social environment, might alter health outcomes.

Funding sources

This study was supported by the European Commission Horizon

2020 grant number [633666] to Paolo Vineis. Mika Kivimäki was supported by NordForsk, the UK Medical Research Council [MRC S011676], the US National Institute on Aging [R01AG056477], the Academy of Finland [311492], and Helsinki Institute of Life Science during the conduct of the study. Cathal McCrory is supported by the Health Research Board (HRB) of Ireland under an Emerging Investigator Award (EIA-2017-012).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to The Centre for Longitudinal Studies, UCL Institute of Education for the use of these data and to the UK Data Archive and UK Data Service for making them available.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbi.2020.09.002>.

References

- Alwan, H., Pruijm, M., Ponte, B., Ackermann, D., Guessous, I., Ehret, G., Staessen, J.A., Asayama, K., Vuistiner, P., Younes, S.E., Paccaud, F., Wuerzner, G., Pechere-Bertschi, A., Mohaupt, M., Vogt, B., Martin, P.-Y., Burnier, M., Bochud, M., 2014. Epidemiology of masked and white-coat hypertension: the family-based SKIPOGH study. *PLoS One* 9, e92522. <https://doi.org/10.1371/journal.pone.0092522>.
- Bennett, J.M., Reeves, G., Billman, G.E., Sturmberg, J.P., 2018. Inflammation-nature's way to efficiently respond to all types of challenges: implications for understanding and managing "the epidemic" of chronic diseases. *Front. Med.* 5, 316. <https://doi.org/10.3389/fmed.2018.00316>.
- Berger, E., Castagné, R., Chadeau-Hyam, M., Bochud, M., d'Errico, A., Gandini, M., Karimi, M., Kivimäki, M., Krogh, V., Marmot, M., Panico, S., Preisig, M., Ricceri, F., Sacerdote, C., Steptoe, A., Stringhini, S., Tumino, R., Vineis, P., Delpierre, C., Kelly-Irving, M., 2019. Multi-cohort study identifies social determinants of systemic inflammation over the life course. *Nat. Commun.* 10, 773. <https://doi.org/10.1038/s41467-019-08732-x>.
- Blane, D., Kelly-Irving, M., d'Errico, A., Bartley, M., Montgomery, S., 2013. Social-biological transitions: how does the social become biological? *Longitud. Life Course Stud.* 4, 136–146. doi: 10.14301/lcs.v4i2.236.
- d'Errico, A., Ricceri, F., Stringhini, S., Carmeli, C., Kivimäki, M., Bartley, M., McCrory, C., Bochud, M., Vollenweider, P., Tumino, R., Goldberg, M., Zins, M., Barros, H., Giles, G., Severi, G., Costa, G., Vineis, P., Lifepath Consortium, 2017. Socioeconomic indicators in epidemiologic research: a practical example from the LIFEPATH study. *PLoS ONE* 12, e0178071. doi: 10.1371/journal.pone.0178071.
- Davillas, A., Benzeval, M., Kumari, M., 2017. Socio-economic inequalities in C-reactive protein and fibrinogen across the adult age span: findings from Understanding Society. *Sci. Rep.* 7, 2641. <https://doi.org/10.1038/s41598-017-02888-6>.
- Del Giudice, M., Gangestad, S.W., 2018. Rethinking IL-6 and CRP: why they are more than inflammatory biomarkers, and why it matters. *Brain. Behav. Immun.* 70, 61–75. <https://doi.org/10.1016/j.bbi.2018.02.013>.
- Eklund, C.M., 2009. Proinflammatory cytokines in CRP baseline regulation. *Adv. Clin. Chem.* 48, 111–136. [https://doi.org/10.1016/s0065-2423\(09\)48005-3](https://doi.org/10.1016/s0065-2423(09)48005-3).
- Fagundes, C.P., Glaser, R., Kiecolt-Glaser, J.K., 2013. Stressful early life experiences and immune dysregulation across the lifespan. *Brain. Behav. Immun.* 27, 8–12. <https://doi.org/10.1016/j.bbi.2012.06.014>.
- Firmann, M., Mayor, V., Vidal, P.M., Bochud, M., Pécoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K.S., Yuan, X., Danoff, T.M., Stirnadel, H.A., Waterworth, D., Mooser, V., Waeber, G., Vollenweider, P., 2008. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* 8, 6. <https://doi.org/10.1186/1471-2261-8-6>.
- Fraga, S., Marques-Vidal, P., Vollenweider, P., Waeber, G., Guessous, I., Paccaud, F., Barros, H., Stringhini, S., 2015. Association of socioeconomic status with inflammatory markers: a two cohort comparison. *Prev. Med.* 71, 12–19. <https://doi.org/10.1016/j.ypmed.2014.11.031>.
- Franceschi, C., Garagnani, P., Parini, P., Giuliani, C., Santoro, A., 2018. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* 14, 576–590. <https://doi.org/10.1038/s41574-018-0059-4>.
- Furman, D., Campisi, J., Verdini, E., Carrera-Bastos, P., Targ, S., Franceschi, C., Ferrucci, L., Gilroy, D.W., Fasano, A., Miller, G.W., Miller, A.H., Mantovani, A., Weyand, C.M., Barzilai, N., Goronzy, J.J., Rando, T.A., Effros, R.B., Lucia, A., Kleinstreuer, N., Slavich, G.M., 2019. Chronic inflammation in the etiology of disease across the life span. *Nat. Med.* 25, 1822–1832. <https://doi.org/10.1038/s41591-019-0675-0>.
- Gares, V., Panico, L., Castagne, R., Delpierre, C., Kelly-Irving, M., Lifepath Consortium, 2017. The role of the early social environment on Epstein Barr virus infection: a prospective observational design using the Millennium Cohort Study. *Epidemiol. Infect.* 145, 3405–3412. doi: 10.1017/S0950268817002515.
- Kivimäki, M., Batty, G.D., Pentti, J., Shipley, M.J., Sipilä, P.N., Nyberg, S.T., Suominen, S.B., Oksanen, T., Stenholm, S., Virtanen, M., Marmot, M.G., Singh-Manoux, A., Brunner, E.J., Lindbohm, J.V., Ferrie, J.E., Vahtera, J., 2020. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *Lancet Public Health* 5, e140–e149. [https://doi.org/10.1016/S2468-2667\(19\)30248-8](https://doi.org/10.1016/S2468-2667(19)30248-8).
- Kivimäki, M., Steptoe, A., 2018. Effects of stress on the development and progression of cardiovascular disease. *Nat. Rev. Cardiol.* 15, 215–229. <https://doi.org/10.1038/nrcardio.2017.189>.
- Koster, A., Bosma, H., Penninx, B.W.J.H., Newman, A.B., Harris, T.B., van Eijk, J.T.M., Kempen, G.I.J.M., Simonsick, E.M., Johnson, K.C., Rooks, R.N., Ayonayon, H.N., Rubin, S.M., Kritchevsky, S.B., Health ABC Study, 2006. Association of inflammatory markers with socioeconomic status. *J. Gerontol. A Biol. Sci. Med. Sci.* 61, 284–290.
- Lacey, R.E., Pinto Pereira, S.M., Li, L., Danese, A., 2020. Adverse childhood experiences and adult inflammation: single adversity, cumulative risk and latent class approaches. *Brain. Behav. Immun.* <https://doi.org/10.1016/j.bbi.2020.03.017>.
- Marmot, M.G., Smith, G.D., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E., Feeney, A., 1991. Health inequalities among British civil servants: the Whitehall II study. *Lancet Lond. Engl.* 337, 1387–1393. [https://doi.org/10.1016/0140-6736\(91\)93068-k](https://doi.org/10.1016/0140-6736(91)93068-k).
- Medzhitov, R., 2008. Origin and physiological roles of inflammation. *Nature* 454, 428–435. <https://doi.org/10.1038/nature07201>.
- Miller, G., Chen, E., Cole, S.W., 2009. Health psychology: developing biologically plausible models linking the social world and physical health. *Annu. Rev. Psychol.* 60, 501–524. <https://doi.org/10.1146/annurev.psych.60.110707.163551>.
- Muscattell, K.A., Brosso, S.N., Humphreys, K.L., 2018. Socioeconomic status and inflammation: a meta-analysis. *Mol. Psychiatry* 1. <https://doi.org/10.1038/s41380-018-0259-2>.
- Nazmi, A., Victora, C.G., 2007. Socioeconomic and racial/ethnic differentials of C-reactive protein levels: a systematic review of population-based studies. *BMC Public Health* 7, 212. <https://doi.org/10.1186/1471-2458-7-212>.
- Panagiotakos, D.B., Pitsavos, C.E., Chrysohoou, C.A., Skoumas, J., Toutouza, M., Belegrios, D., Toutouzas, P.K., Stefanadis, C., 2004. The association between educational status and risk factors related to cardiovascular disease in healthy individuals: the ATTICA study. *Ann. Epidemiol.* 14, 188–194. [https://doi.org/10.1016/S1047-2797\(03\)00117-0](https://doi.org/10.1016/S1047-2797(03)00117-0).
- Petrovic, D., de Mestral, C., Bochud, M., Bartley, M., Kivimäki, M., Vineis, P., Mackenbach, J., Stringhini, S., 2018. The contribution of health behaviors to socioeconomic inequalities in health: a systematic review. *Prev. Med.* 113, 15–31. <https://doi.org/10.1016/j.ypmed.2018.05.003>.
- Pollitt, R.A., Kaufman, J.S., Rose, K.M., Diez-Roux, A.V., Zeng, D., Heiss, G., 2008. Cumulative life course and adult socioeconomic status and markers of inflammation in adulthood. *J. Epidemiol. Community Health* 62, 484–491. <https://doi.org/10.1136/jech.2006.054106>.
- Power, C., Elliott, J., 2006. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int. J. Epidemiol.* 35, 34–41. <https://doi.org/10.1093/ije/dy183>.
- R Core Team, 2014. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing 73, 3–36.
- Raitakari, O.T., Juonala, M., Rönnemaa, T., Keltikangas-Järvinen, L., Räsänen, L., Pietikäinen, M., Hutri-Kähönen, N., Taittonen, L., Jokinen, E., Marniemi, J., Jula, A., Telama, R., Kähönen, M., Lehtimäki, T., Akerblom, H.K., Viikari, J.S.A., 2008. Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* 37, 1220–1226. <https://doi.org/10.1093/ije/dym225>.
- RStudio Team, 2016. RStudio: Integrated Development Environment for R.
- Smith, K.J., Gavey, S., Riddell, N.E., Kontari, P., Victor, C., 2020. The association between loneliness, social isolation and inflammation: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 112, 519–541. doi: 10.1016/j.neubiorev.2020.02.002.
- Steiniv, A., Shirom, A., Melamed, S., Tokes, S., Justo, D., Saar, N., Shapira, I., Berliner, S., Rogowski, O., 2008. Relation of educational level to inflammation-sensitive biomarker level. *Am. J. Cardiol.* 102, 1034–1039. <https://doi.org/10.1016/j.amjcard.2008.05.055>.
- Steptoe, A., Breeze, E., Banks, J., Nazroo, J., 2013. Cohort profile: the English longitudinal study of ageing. *Int. J. Epidemiol.* 42, 1640–1648. <https://doi.org/10.1093/ije/dys168>.
- Steptoe, A., Owen, N., Kunz-Ebrecht, S., Mohamed-Ali, V., 2002. Inflammatory cytokines, socioeconomic status, and acute stress reactivity. *Brain. Behav. Immun.* 16, 774–784. [https://doi.org/10.1016/s0889-1591\(02\)00030-2](https://doi.org/10.1016/s0889-1591(02)00030-2).
- Straub, R.H., 2017. The brain and immune system prompt energy shortage in chronic inflammation and ageing. *Nat. Rev. Rheumatol.* 13, 743–751. <https://doi.org/10.1038/nrrheum.2017.172>.
- Vineis, P., Avendano-Pabon, M., Barros, H., Chadeau-Hyam, M., Costa, G., Dijaurescu, M., Delpierre, C., D'Errico, A., Fraga, S., Giles, G., Goldberg, M., Zins, M., Kelly-Irving, M., Kivimäki, M., Lang, T., Layte, R., Mackenbach, J.P., Marmot, M., McCrory, C., Carmeli, C., Milne, R.L., Muennig, P., Nusselder, W., Polidoro, S., Ricceri, F., Robinson, O., Stringhini, S., 2017. The biology of inequalities in health: the LIFEPATH project. *Longitud. Life Course Stud.* 8, 33. doi: 10.14301/lcs.v8i4.448.
- Vingeliene, S., Hiyoshi, A., Lentjes, M., Fall, K., Montgomery, S., 2019. Longitudinal analysis of loneliness and inflammation at older ages: english longitudinal study of ageing. *Psychoneuroendocrinology* 110, 104421. <https://doi.org/10.1016/j.psyneuen.2019.104421>.
- Yang, Y.C., Schorpp, K., Harris, K.M., 2014. Social support, social strain and inflammation: evidence from a national longitudinal study of U.S. adults. 1982. *Soc. Sci. Med.* 107, 124–135. <https://doi.org/10.1016/j.socscimed.2014.02.013>.

ANNEXE 3 : INFLAMMATION CHRONIQUE ET DEVELOPPEMENT DU
CANCER DU SEIN ET DES LYMPHOMES NON-HODGKINIEN : ETUDE CAS-
CONTROLES DANS DEUX COHORTES PROSPECTIVES

SCIENTIFIC REPORTS



OPEN

Association between low-grade inflammation and Breast cancer and B-cell Myeloma and Non-Hodgkin Lymphoma: findings from two prospective cohorts

Eloise Berger¹, Cyrille Delpierre¹, Fatemeh Saberi Hosnijeh^{2,3}, Michelle Kelly-Irving¹, Lutzen Portengen², Ingvar A. Bergdahl⁴, Ann-Sofie Johansson⁵, Vittorio Krogh⁶, Domenico Palli⁷, Salvatore Panico⁸, Carlotta Sacerdote⁹, Rosario Tumino¹⁰, Soterios A. Kyrtopoulos¹¹, Paolo Vineis^{12,13,14}, Marc Chadeau-Hyam^{2,13,14}, Roel Vermeulen^{2,13}, Raphaële Castagné^{1,13,14} & EnviroGenoMarkers*

Chronic inflammation may be involved in cancer development and progression. Using 28 inflammatory-related proteins collected from prospective blood samples from two case-control studies nested in the Italian component of the European Prospective Investigation into Cancer and nutrition (n = 261) and in the Northern Sweden Health and Disease Study (n = 402), we tested the hypothesis that an inflammatory score is associated with breast cancer (BC) and B-cell Non-Hodgkin Lymphoma (B-cell NHL, including 68 multiple myeloma cases) onset. We modelled the relationship between this inflammatory score and the two cancers studied: (BC and B-cell NHL) using generalised linear models, and assessed, through adjustments the role of behaviours and lifestyle factors. Analyses were performed by cancer types pooling both populations, and stratified by cohorts, and time to diagnosis. Our results suggested a lower inflammatory score in B-cell NHL cases ($\beta = -1.28$, $p = 0.012$), and, to lesser, extent with BC ($\beta = -0.96$, $p = 0.33$) compared to controls, mainly driven by cancer cases diagnosed less than 6 years after enrolment. These associations were not affected by subsequent adjustments for potential intermediate confounders, notably behaviours. Sensitivity analyses indicated that our findings were not affected by the way the inflammatory score was calculated. These observations call for further studies involving larger populations, larger variety of cancer types and repeated measures of larger panel of inflammatory markers.

¹LEASP, UMR 1027, Inserm-Université Toulouse III Paul Sabatier, Toulouse, France. ²Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands. ³Immunology Department, Erasmus University Medical Center, Rotterdam, The Netherlands. ⁴Department of Biobank Research, Umeå University, Umeå, Sweden. ⁵Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden. ⁶Fondazione IRCCS- Istituto Nazionale dei Tumori, Milan, Italy. ⁷Istituto per lo Studio e la Prevenzione Oncologica (ISPO Toscana), Florence, Italy. ⁸Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy. ⁹Piedmont Reference Centre for Epidemiology and Cancer Prevention (CPO Piemonte), Turin, Italy. ¹⁰Cancer registry and Histopathology Unit, Azienda Ospedaliera 'Civile -M.P.Arezzo', Ragusa, Italy. ¹¹National Hellenic Research Foundation, Institute of Biology, Medicinal Chemistry and Biotechnology, Athens, Greece. ¹²HuGeF, Human Genetics Foundation, Torino, Italy. ¹³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ¹⁴MRC-PHE Centre for Environment and Health, Imperial College, London, London, UK. Roel Vermeulen and Raphaële Castagné jointly supervised this work. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to R.C. (email: raphaele.castagne@inserm.fr)

Inflammation is a key pathway in many diseases including cancer¹, a leading cause of death worldwide². More specifically inflammation is one of the enabling hallmarks of cancer³ and is thought to influence several phases of cancer development including initiation and progression⁴ through both intrinsic (genetic events) and extrinsic (soluble mediators) pathways⁵. Epidemiological, pharmacological and genetic research provide solid evidence that inflammation can increase cancer risk and promote tumour progression^{6,7}. Epidemiological studies have already examined the link between inflammation and risk of some cancers including bladder, oesophageal, ovarian, prostate or thyroid cancer⁶ but also colorectal cancer⁸ and others⁷.

While breast cancer (BC) is the most commonly diagnosed cancer among women worldwide, B-cell lymphomas (BCLs) are the most common hematopoietic cancers in both men and women in the developed world. Few epidemiological studies have investigated the association between blood levels of cytokines^{9,10}, CRP^{11–13} and/or immune related factor⁹ and risk of BC in prospective settings and results were inconsistent^{9–15}. Several prospective studies have also examined the association between B-cell Non-Hodgkin Lymphoma (B-cell NHL) and circulating levels of inflammatory markers providing strong evidence of a subtle role of inflammation in B cell NHL development and progression^{16–26}. These studies have suggested an increased risk of B cell NHL and/or its subtypes with increased blood levels of interleukin^{17–20,23–25}, chemokines^{16,25}, immune activation markers^{16,17,22} and growth factor¹⁹. While other studies reported negative association of interleukin with risk of B cell NHL^{18,23}.

The prevalence of these two cancers and the conflicting results regarding the role of inflammation underline that BC and B-cell NHL are particularly relevant to explore the role of inflammation in cancer initiation.

Previous studies generally focus on the concentration levels of individual biomarkers, and since these biomarkers are systemic, associations between individual biomarkers and cancer may not be always apparent. The use of lower resolution summary scores may improve statistical power and reveal associations with cancer risk^{27,28}. In the present work, we investigate inflammatory scores to ascertain their link with cancers onset. We used data from the EnviroGenoMarkers (EGM) project designed as two nested case-control studies including breast cancer (BC) and B-cell Non-Hodgkin Lymphoma (B-cell NHL, including multiple myeloma) cases within two cohorts. Specifically, we constructed our inflammatory score using a large panel of cytokines, chemokines and growth factors measured in prospectively collected peripheral blood samples in 268 participants from the Italian component of the European Prospective Investigation into Cancer and Nutrition (EPIC-Italy) and 492 participants from the Northern Sweden Health and Disease Study (NSHDS). Making the most of the EGM dataset, we examined participant's inflammatory status through the inflammatory score (and a corresponding unsupervised alternative) in relation to cancer onset as a common mechanism in disease initiation and progression. These associations were investigated for each type of cancer (breast cancer and B-cell NHL separately), and after stratification by time-to-diagnosis (TtD).

Results

Study population. Key characteristics of the study population are reported in Table 1, participants were predominantly from Sweden (60.6%) with a mean age of 52.8 years (SD = 7.7). Participants from EPIC-Italy mostly women (74.3%) tended to report a lower education level, more physical activity, and have a higher alcohol consumption compared to participants from NSHDS. Our study population includes 90 breast cancer (54.4% from EPIC-Italy) and 248 B-cell NHL (33.8% from EPIC-Italy) prospective cases which were diagnosed from 2 to 15.5 years after inclusion.

Pre-diagnostic inflammatory score and breast cancer. The breast cancer population includes a total of 167 women (56.3% from EPIC-Italy) with a slightly higher proportion of cases in both cohorts after excluding individuals with missing data (Supplementary Table S1). The median time to diagnosis was 5.84 years and showed limited heterogeneity in both cohorts (5.35 years in EPIC-Italy and 6.43 years in NSHDS). BC cases from both cohorts were more inactive compared to controls ($p = 0.05$ in EPIC-Italy and NSHDS). In NSHDS, BC cases had lower BMI, experienced fewer pregnancies whereas no significant differences between cases and controls for any of the other baseline variables investigated were observed in EPIC-Italy (Supplementary Table S1). Women with BC from EPIC-Italy differed from those from NSHDS by age at menarche, parity, the use of hormone replacement therapy and their menopausal status (Supplementary Table S2).

Our analyses based on the full BC population suggested a lower inflammatory score in BC cases compared to controls (model 1: $\beta = -0.96$, $p = 0.333$, Table 2A), but that difference did not reach statistical significance. Consistent estimates and similar conclusions can be drawn for the analyses stratified by cohort (model 1: $\beta = -1.12$, $p = 0.436$ and $\beta = -0.64$, $p = 0.640$ respectively for EPIC-Italy and NSHDS, Table 2A). Adjusting for either behavioural, socioeconomic, or hormonal factors, or in the fully adjusted models, effect size estimates showed consistent signs, and associations did not reach statistical significance ($p > 0.09$ across all models investigated, Table 2A, Supplementary Fig. S1). Stratification by time to diagnosis revealed statistically significant lower inflammatory score in BC cases diagnosed less than 6 years after enrolment ($\beta = -2.88$, $p = 0.032$, Table 2A). Results for each cohort separately showed consistent direction of association, but the lower inflammatory score in BC cases diagnosed less than 6 years after enrolment was borderline statistically significant in EPIC-Italy ($\beta = -3.52$, $p = 0.057$) and non-significant in NSHDS ($\beta = -3.58$, $p = 0.196$, Table 2A, Supplementary Fig. S1). No associations were observed in the pooled population or by cohort for BC cases diagnosed 6 years after enrolment ($\beta = -0.06$, $p = 0.970$, Table 2A).

Analyses using the first principal component (PC1, explaining 32.6% of the variance) as an alternative inflammatory score showed that our conclusions were robust to our definition of the inflammatory score and we identified the same significant association indicating a lower inflammatory score in the BC cases diagnosed less than 6 years after enrolment ($\beta = -1.55$, $p = 0.029$, Table 2B, Supplementary Fig. S2).

		All individuals (n = 663)		EPIC-Italy (n = 261)		NSHDS (n = 402)		P-values
		N	% or mean(sd)	N	% or mean(sd)	N	% or mean(sd)	
Gender	Male	244	36,8%	67	25,7%	177	44,0%	<0.001
	Female	419	63,2%	194	74,3%	225	56,0%	
Type	Case	338	51,0%	133	51,0%	205	51,0%	—
	Control	325	49,0%	128	49,0%	197	49,0%	
Disease	Breast cancer (cases)	167 (90)	25,2%	94 (49)	36,0%	73 (41)	18,2%	<0.001
	NHL (cases)	496 (248)	74,8%	167 (84)	64,0%	329 (164)	81,8%	
Age	Global	663	52.8(7.7)	261	53.5(8.1)	402	52.4(7.5)	0.430
BMI (kg/m ²)	Global	663	26.0(3.9)	261	25.8(3.6)	402	26.2(4.1)	0.329
Phase	1	355	53,5%	194	74,3%	161	40,1%	<0.001
	2	308	46,5%	67	25,7%	241	59,9%	
Center	Central	116	17,5%	116	44,4%	0	0,0%	—
	North	104	15,7%	104	39,9%	0	0,0%	
	South	41	6,2%	41	15,7%	0	0,0%	
	Umea	402	60,6%	0	0,0%	402	100,0%	
Smoking status	Never	326	49,2%	127	48,7%	199	49,5%	0.807
	Former	183	27,6%	70	26,8%	113	28,1%	
	Current	154	23,2%	64	24,5%	90	22,4%	
Alcohol (g/day)	Global	663	7.1(12.5)	261	12.5(17.6)	402	3.6(5.0)	<0.001
Physical activity	Inactive	179	27,1%	70	26,8%	109	27,1%	0.007
	Moderately inactive	266	40,1%	106	40,6%	160	39,8%	
	Moderately active	166	25,0%	54	20,7%	112	27,9%	
	Active	52	7,8%	31	11,9%	21	5,2%	
Education	None/primary	270	40,7%	146	55,9%	124	30,8%	<0.001
	Professional/technical	132	19,9%	30	11,5%	102	25,4%	
	Secondary	147	22,2%	55	21,1%	92	22,9%	
	University/college	114	17,2%	30	11,5%	84	20,9%	
Inflammatory score	Global	663	7.0(5.9)	261	6.8(6.1)	402	7.0(5.7)	0.244
PC1	Global	663	−0.01(3.0)	261	−0.003(3.0)	402	−0.02(3.1)	0.807

Table 1. Comparison of baseline characteristics of EPIC-Italy and NSHDS. P-values are estimated with χ^2 tests, Student T-tests or Wilcoxon rank sum tests.

Pre-diagnostic inflammatory score and B-cell NHL. Key characteristics of the B-cell NHL study population are given in Supplementary Table S3. Of the 248 β -cell malignancies cases diagnosed during follow-up, 84 arose from the Italian population and 164 from Sweden. The median time to diagnosis was 6.09 years (5.53 in EPIC-Italy and 6.38 in NSHDS). Cases and controls were on average 53.1 years old (Supplementary Table S3).

Our analyses of all B-cell NHL cases (i.e. pooling all histological subtypes) indicated a lower inflammatory score compared to controls (model 1: $\beta = -1.28$, $p = 0.012$ Table 3A). Consistent effect size estimates were observed in both cohorts separately, but the lower inflammatory score in B-cell NHL cases was only found statistically significant in the Swedish population (model 1: $\beta = -0.89$, $p = 0.310$ and $\beta = -1.48$, $p = 0.019$ for EPIC-Italy and NSHDS, respectively, Table 3A). The adjustment for socioeconomic variables, as for behaviours, lightly weakened these associations, but in the fully adjusted model, the lower inflammatory score in B-cell NHL cases was still significant in the pooled population (Table 3A – Supplementary Fig. S3).

After stratification on time to diagnosis, we observed a borderline significant lower inflammatory score in cases diagnosed less than 6 years after inclusion in the pooled B-cell NHL population ($\beta = -1.27$, $p = 0.051$, Table 3A). A consistent estimate of the effect size was observed in cases diagnosed more than six years after enrollment, without the difference reaching statistical significance ($\beta = -1.06$, $p = 0.108$). B-cell NHL cases diagnosed more than 6 years after enrolment showed significantly lower inflammatory score only in the NSHDS samples. As before, results using PC1 as an alternative inflammatory score suggested highly consistent results (Table 3B, Supplementary Fig. S4).

Discussion

Our analyses identified a general lower burden of inflammation in prospective BC and B-cell NHL cancer cases. Although non-significant, BC cases had a lower inflammatory score compared to controls in the pooled population and in both cohort. This difference was mainly driven by BC cases diagnosed less than 6 years after enrollment, for whom the inflammatory score was significantly lower than in controls. No significant differences were observed after stratification by BC histological subtypes in both oestrogen receptor positive and negative BC (Supplementary Table S4).

	Full Breast Cancer population										TtD ≤ 6 years*		TtD > 6 years*	
	Model 1		Model 1 + Behaviours		Model 1 + SEP		Model 1 + Hormonal		Fully adjusted model		Fully adjusted model		Fully adjusted model	
	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value
(A) Inflammatory score														
All	-0.96 (0.99)	0.333	-1.28 (1.01)	0.208	-1.27 (1.02)	0.215	-1.35(1.05)	0.201	-1.72(1.09)	0.118	-2.88(1.32)	0.032	-0.06(1.43)	0.970
EPIC-Italy	-1.12 (1.43)	0.436	-2.03 (1.47)	0.170	-1.45 (1.45)	0.318	-1.33(1.57)	0.400	-2.36(1.63)	0.151	-3.52(1.81)	0.057	1.03(2.67)	0.701
NSHDS	-0.64 (1.37)	0.640	-0.25 (1.54)	0.871	-0.86 (1.47)	0.558	-1.72(1.57)	0.278	-1.75 (1.88)	0.356	-3.58 (2.70)	0.196	-2.56 (1.88)	0.181
(B) Principal component 1														
All	-0.56 (0.53)	0.297	-0.74 (0.54)	0.175	-0.64 (0.55)	0.241	-0.87(0.55)	0.166	-1.00(0.57)	0.083	-1.55(0.70)	0.029	-0.09(0.75)	0.906
EPIC-Italy	-0.75 (0.77)	0.335	-1.25 (0.81)	0.123	-0.87 (0.79)	0.273	-0.83(0.84)	0.322	-1.30(0.88)	0.147	-1.90(1.04)	0.074	0.83(1.42)	0.564
NSHDS	-0.17 (0.71)	0.810	-0.004(0.77)	0.996	-0.23 (0.76)	0.760	-0.93 (0.81)	0.253	-0.85 (0.92)	0.361	-1.45 (1.22)	0.242	-1.29 (1.02)	0.214

Table 2. Association of pre-diagnostic inflammatory score (A) and the first PC (B) with breast cancer case/control. Pooled cohorts (n = 167), EPIC-Italy (n = 94) and NSHDS (n = 73). Model 1 is adjusted for age and cohort in the overall BC population; for age and center in EPIC-Italy and age in NSHDS. β coefficient regression estimates inflammatory score's (A) or PC1's (B) difference in cases compared to controls. *Additional analyses stratified by time from blood collection to cases' diagnosis included for 'less than 6 years' strata 49 BC cancer cases: 32 in EPIC-Italy and 17 in NSHDS; for 'higher than 6 years' strata 41 BC cancer cases: 17 in EPIC-Italy and 24 in NSHDS.

	Full Lymphoma population								TtD ≤ 6 years*		TtD > 6 years*	
	Model 1		Model 1 + Behaviours		Model 1 + SEP		Fully adjusted model		Fully adjusted model		Fully adjusted model	
	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value	β (SE)	P-value
(A) Inflammatory score												
All	-1.28 (0.51)	0.012	-1.19 (0.51)	0.021	-1.21 (0.51)	0.019	-1.11 (0.52)	0.032	-1.27 (0.65)	0.051	-1.06 (0.66)	0.108
EPIC-Italy	-0.89 (0.87)	0.310	-0.93 (0.89)	0.299	-0.96 (0.88)	0.277	-1.02 (0.90)	0.258	-1.87 (1.08)	0.087	0.05 (1.20)	0.965
NSHDS	-1.48 (0.63)	0.019	-1.31 (0.64)	0.040	-1.36 (0.63)	0.031	-1.19 (0.64)	0.062	-0.71 (0.82)	0.389	-1.61 (0.79)	0.043
(B) Principal component 1												
All	-0.77 (0.26)	0.004	-0.71 (0.26)	0.007	-0.71 (0.26)	0.007	-0.65 (0.26)	0.013	-0.76 (0.32)	0.018	-0.60 (0.33)	0.073
EPIC-Italy	-0.59 (0.40)	0.138	-0.66 (0.40)	0.101	-0.62 (0.40)	0.126	-0.70 (0.40)	0.083	-1.14 (0.47)	0.017	-0.13 (0.54)	0.816
NSHDS	-0.85 (0.34)	0.012	-0.73 (0.34)	0.032	-0.76 (0.34)	0.024	-0.64 (0.34)	0.059	-0.48 (0.43)	0.263	-0.82 (0.42)	0.053

Table 3. Association of pre-diagnostic inflammatory score (A) and the first PC (B) with B-cell non-Hodgkin lymphoma case/control. Pooled cohorts (n = 496), EPIC-Italy (n = 167) and NSHDS (n = 329) Model 1 is adjusted for age, gender, phase and cohort in the overall all B-cell NHL population; for age, gender, phase and center in EPIC-Italy and age, gender, phase in NSHDS. β coefficient regression estimates inflammatory score's (A) or PC1's (B) difference in cases compared to controls. *Additional analyses stratified by time from blood collection to cases' diagnosis included for 'less than 6 years' strata 129 B-cell NHL cancer cases: 49 in EPIC-Italy and 80 in NSHDS; for 'higher than 6 years' strata 119 B-cell NHL cancer cases: 35 in EPIC-Italy and 84 in NSHDS.

B cell NHL cases also had a significantly lower inflammatory score compared to controls in the pooled population and in NSHDS. Additional analyses showed a lower inflammatory score in all histological subtype cases, statistically significant in the largest subtype group of multiple myeloma (27.4% of B-cell NHL cases, Supplementary Table S5). Our results were robust to the definition of the inflammatory score and were not affected by adjusting for the main potential confounders.

Several limitations of this study should be considered. Our study population remains limited in size especially the stratified analyses by cohort, by histological subtypes and time to diagnoses. Our results showed overall similar trends but reduced strength of associations, which might indicate limited statistical power. The two cohorts we used may be heterogeneous in terms of population genetics, exposure profiles, lifestyle factors and study design-related source of variability, explaining some result differences between the two cohorts. Only BC and B-cell NHL cases were included in the EGM project, hence limiting our ability to investigate the role of inflammation in various cancer types. Additionally, a minimum of two years elapsed between enrolment and diagnosis. Since cancer development is likely to have a long latency period, over many years, we cannot rule out the presence of abnormal cells and cancer cases may already have been in a preclinical state at enrolment. Furthermore, we measured the inflammatory proteins at a single time point which may not reflect the long term inflammatory status of an individual and does not allow us to consider the speed of cancer progression in relation to longitudinal

inflammatory burden. Finally, we cannot rule out the possibility that other factors may contribute to the association between inflammation and cancer risk. Furthermore variables relating to the inflammatory system, such as infection status and/or medication could not be accounted for.

Our study has a number of strengths. First, the prospective design limits both recall and reverse causation bias which can be induced by the disease itself, treatments or lifestyle changes after diagnosis. We also had access to a wide range of information about lifestyle factors and hormonal status which allowed us to adjust for important potential confounders. Additionally, the availability of two cohorts allowed for independent confirmation of the observed signals and comparison of the inflammatory score and cancer association results between two countries. Moreover, compared to most previous prospective studies, we measured and combined into a score a large panel of inflammatory proteins to capture the inflammatory load. We know of two studies which have examined several markers of inflammation both individually and jointly as an inflammatory score and have reported that the use of a score is a robust method to explore the association between inflammatory load and future risk of cancer^{27,29}. Allin *et al.* reported a positive association between an inflammatory score derived from three inflammatory biomarkers (CRP, fibrinogen and whole blood leukocyte count) and future risk of colorectal, lung and BC during a median follow-up period of 4.8 years in a Danish general population²⁷. To test the robustness of the definition of the inflammatory score, we defined, as an unsupervised alternative, a score using the first principal component obtained from the 28 inflammatory related proteins. Our results and conclusions remained stable with either score.

As a general pattern, behavioural factors and education did not seem to explain a substantial part of case-control differences in the inflammatory score. An unhealthy lifestyle, including smoking³⁰, heavy drinking³¹, physical inactivity³² and body mass index is related to higher chronic inflammation and adverse health outcomes³³. In our study, after controlling for behavioural and education, no significant changes were observed in the associations indicating that these variables do not act as confounder in the observed association between inflammation and cancer, suggesting that inflammation may be causally linked to cancer risk.

Our study remains however restricted to a limited set of potential confounders and mediators, hence hampering the generalisability of our conclusion.

A vast majority of the studies looking at the association between inflammation and BC risk used the CRP and a recent meta-analysis concludes a modest statistically significant positive association between CRP concentration and breast cancer risk^{11,34}. Our results may appear to be contradictory. This may be due to our approach which summarises the association of 28 inflammatory related proteins with disease risk using a score to better capture the inflammatory status and improve statistical power but also attributable to the timing in which blood samples were collected and cancer diagnosed, or to other factors or confounders. Several epidemiological studies have explored the link between circulating marker of adiposity and inflammation in relation to BC risk with the underlying idea that obesity associated inflammation may increase the risk of cancer^{9,34–36}. The release of inflammatory mediators may then take place in the adipocytes and might not be apparent at the same level in the circulatory system^{36,37}. The time from blood collection to cancer diagnosis may change according to studies and influence results. In our results a lower inflammatory score in BC cases compared to controls was only observed in those diagnosed less than 6 years after enrolment. Early BC cases may present inflammatory differences compared to those developing BC cancer later or may be already in a pre-disease stage characterized by a lower inflammatory score suggesting that inflammatory status differs according to BC cancer evolution time.

Numerous studies have investigated the association between pre-diagnostic inflammation and NHL using cytokines, chemokines and other immune markers independently. These studies showed a deregulation of these biomarkers with future risk of NHL suggesting subclinical dysfunction as a consistent risk factor^{16–21,24,38}. In a previous study in a subset of EPIC-Italy's population, the authors reported association between lower levels of IL-2, INF- γ and upper levels of ICAM and risk of NHL¹⁸. In another study, it has been shown that lower levels of leptin was associated with increased risk of NHL and upper levels of IL-10 increase risk of B-NHL; they also found no association between a composite score from a principal component and future risk of NHL²⁰. A case control study in the PLCO trial reported an increased risk of NHL with elevated serum levels of TNF-R1 and sCD27¹⁷. Our results demonstrate that future cases of NHL had a lower inflammatory score compared to controls. This provides further support for the subtle status of deregulated immune response or failure to modulate the immune response appropriately in relation to NHL risk. It is also established that NHL is one of the most common malignancies occurring after immunosuppressive therapy use due to transplantation^{39,40}.

Inflammation has multifaceted role in cancer including initiation, promotion and progression. Deregulation of biological processes can lead towards chronic inflammation or play in immunosuppressive roles. Our results from a case-control study nested within 2 Europeans cohorts focusing on 2 specific cancers (BC and B-cell NHL) suggest an association between the presence of cancer and a lower inflammatory score and support the complex role of chronic inflammation in cancer development. The use of the PC1 aimed to measure the inflammation variability captured by the 28 proteins. These findings call for further studies to better understand how inflammatory changes evolve involving larger populations, larger variety of cancer types and repeated measures of larger panels of inflammatory markers.

Methods

Study population. The Envirogenomarkers study (EGM, www.envirogenomarkers.net) was designed as two nested case-control studies and includes participants from the Italian component of the European Prospective Investigation into Cancer and nutrition⁴¹ (EPIC-Italy) and the Northern Sweden Health and Disease Study⁴² (NSHDS), which have been described previously^{43,44}. The Envirogenomarkers project and its associated studies, experimental protocols and methods were approved by the Regional Ethical Review Board of the Umeå Division of Medical Research, for the Swedish cohort, and the Florence Health Unit Local Ethical Committee, for the Italian cohort. All methods were carried out in accordance with the approved guidelines.

The EGM research project was to identify novel biomarkers associated with chronic diseases in which the environment might play an important role. EGM project was therefore focused on breast cancer and B-cell NHL. Exposure of environmental pollutants including polychlorinated biphenyls, polycyclic aromatic hydrocarbons, cadmium or lead were also collected.

EPIC-Italy includes 47,749 healthy participants aged 35–70 years old from 5 different areas: Turin, Varese, Florence, Naples, and Ragusa. Anthropometric measurement, lifestyle factors and blood samples were collected at recruitment (1993–1998). Standardized procedures were used to identify newly diagnosed cases of cancer based on automated linkages to cancer and mortality registries, municipal population offices and hospital discharge systems. In Naples follow-up information was collected through periodic personal contact. All participants signed a written informed consent and the ethical review boards of the International Agency for Research on Cancer, and of the collaborating institutions responsible for subject recruitment in each of the EPIC recruitment centers approved the study.

NSHDS is a prospective cohort study that included 95,000 participants from the general population from 1990 up to January 2008. It includes three sub-cohorts, of which the largest is the Västerbotten Intervention Programme (VIP) that mainly recruits individuals aged 40, 50 or 60 years. At initial recruitment, subjects were asked to complete a self-administered questionnaire to collect demographic, medical and lifestyle information and a separate self-administered food frequency questionnaire. Informed consent was obtained from all participants and a medical examination was conducted during which a blood sample was taken. Incident cancers occurring during follow-up were identified by linkage with the Swedish Cancer Registry and the local Northern Sweden Cancer Registry. NSHDS study was approved by the Local Ethical Committee in Umeå, Sweden. The sample selection strategy is described in Supplementary Fig. S5.

Outcome. Our study population includes breast cancer (BC) and B-cell non-Hodgkin lymphoma (B-cell NHL) cases who were healthy at blood collection and were clinically diagnosed from 2 (minimum chosen duration to avoid the inclusion of cancer cases at enrolment) to 15.5 years after inclusion. Only invasive breast cancer cases were included in the study. All eligible B-cell NHL cases, including multiple myeloma were included. For each case, one suitable control was selected among participants in each cohort who were alive and free of cancer and were matched by center, gender, date of blood collection (+/–6 months), and age at recruitment (+/–2.5 years). We considered all B-cell NHL subtype together and in the text B-cell NHL includes multiple myeloma.

Laboratory analyses. For each participant, a blood sample was collected at enrolment and stored in citrate (Italy) or EDTA (Sweden). Within two hours, plasma was separated and placed in cold storage. Biosamples underwent inflammatory profiling in two distinct phases and 32 of inflammation-related proteins were measured using the milliplex HCYTOMAG-60K and HSCYTOMAG-60SK kits (Millipore, Billerica, MA), according to the manufacturer's protocol⁴⁵. Four analytes were excluded from further statistical analyses due to high rates of non-detection (>75%). Finally, 12 cytokines, 10 chemokines and 6 growth and angiogenic factors were measured for all participants (Supplementary Table S6). Protein levels below the level of detection were imputed based on a maximum likelihood estimation method exploiting the correlation structure across proteins to draw the missing values⁴⁶. Levels of proteins were log-transformed to normalise their distributions. As previously proposed⁴⁷, linear mixed models efficiently correct for technically-induced variation, which is potentially diluting the effects of interest. In practice, we adopted a two-step procedure first fitting a linear mixed model regressing protein levels (response) including a random intercept depending on technical covariates (microtiter plate). In a second step, we subtracted from the measured protein levels the random effect estimates measuring the variation linked to technical covariates to obtain denoised data (i.e. without the potential technical artefact)⁴⁸.

Inflammatory measure. As proposed previously⁴⁸, and under the working hypothesis of a consistent contribution of each inflammatory marker to the overall inflammatory burden, we summarized the individual inflammatory status as a score derived from the 28 proteins assayed: for each protein, we defined a dichotomized indicator: “high concentration” = 1, and “low concentration” = 0 based on the highest quartile of the denoised protein concentrations, and summed across the 28 proteins these binary indicators⁴⁸. Principal component analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set⁴⁹. PCA of these 28 inflammatory markers concentrations shows that 19 PCs explained more than 90% of the total variation seen in the dataset (Supplementary Fig. S6). As a continuous and hypothesis-free alternative score, we used the scores of the first principal component (PC1, explaining 32.6% of the total variance of the 28 proteins).

Statistical analysis. We compared baseline characteristics of the total population and by cancer sub-population using the Chi-squared test or Fisher exact test for the categorical variables and T-test or Wilcoxon rank test for continuous variables. Categorical variables are reported as percentages while continuous variables are reported as mean (and standard deviation).

Our inflammatory score and its continuous alternative (PC1) were regressed against the disease outcome (case-control status and in BC and B-cell NHL population, separately) on the pooled population. We used the same model adjusting for potential confounders collected at enrolment for each type of cancer: age, gender, phase, cohort and center (Model 1). Potential confounders included behavioural factors: body mass index (BMI, continuous variable, kg/m²); smoking status (categorical: current, former, never); alcohol consumption (continuous, g/day); physical activity (categorical: inactive, moderately inactive, moderately active, active); and education (categorical: none/primary, professional/technical, secondary, university/college) as a proxy for socioeconomic status. Based on model 1, we controlled separately for behavioural factors (hereafter referred to as model 1 + behaviours) and for participant's educational level (hereafter referred to as model 1 + socioeconomic position). The

fully adjusted model included behavioural and socioeconomic factors listed above and was additionally adjusted for the reproductive/hormonal variables for BC analyses only. The latter comprised menopausal status (categorical: post-menopausal, pre-menopausal, unknown); contraceptive usage (categorical: yes, no); age at menarche (binary: ≤ 12 years old or > 12 years old); menopausal hormone use (categorical: yes, no) and parity (quantitative discrete: 0, 1, 2, 3, > 4).

Sensitivity analyses. We fitted our models on data for each cohort separately (EPIC-Italy and NSHDS). To evaluate the impact of follow-up time and consider a potential effect of the preclinical phase of the disease, analyses were also stratified on the time to diagnosis as defined by the time elapsed between blood collection and clinical diagnosis: below or above the median time to diagnosis (6 years) across the entire population of cancer cases. We used the same time to diagnosis cut-off (below or above 6 years) for each type of cancer since the median time to diagnosis vary little across cancer type (5.84 years in BC and 6.09 years in B-cell NHL). In these analyses, cases were compared to all controls to preserve statistical power. Analyses were also performed by BC and B-cell NHL histological subtypes.

Statistical analyses were performed in R v3.2⁵⁰ using the RStudio environment v.0.99.484. All p-values with $p < 0.05$ were considered statistically significant.

Availability of data and materials. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Scrive, R., Vasile, M., Bartosiewicz, I. & Valesini, G. Inflammation as “common soil” of the multifactorial diseases. *Autoimmun. Rev.* **10**, 369–374 (2011).
- Global Burden of Disease Cancer Collaboration *et al.* The Global Burden of Cancer 2013. *JAMA Oncol.* **1**, 505–527 (2015).
- Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
- Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* **140**, 883–899 (2010).
- Allavena, P., Garlanda, C., Borrello, M. G., Sica, A. & Mantovani, A. Pathways connecting inflammation and cancer. *Curr. Opin. Genet. Dev.* **18**, 3–10 (2008).
- Mantovani, A., Allavena, P., Sica, A. & Balkwill, F. Cancer-related inflammation. *Nature* **454**, 436–444 (2008).
- Balkwill, F., Charles, K. A. & Mantovani, A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* **7**, 211–217 (2005).
- Lasry, A., Zinger, A. & Ben-Neriah, Y. Inflammatory networks underlying colorectal cancer. *Nat. Immunol.* **17**, 230–240 (2016).
- Dias, J. A. *et al.* Low-Grade Inflammation, Oxidative Stress and Risk of Invasive Post-Menopausal Breast Cancer - A Nested Case-Control Study from the Malmö Diet and Cancer Cohort. *PLoS One* **11**, e0158959 (2016).
- Tamimi, R. M. *et al.* Circulating colony stimulating factor-1 and breast cancer risk. *Cancer Res.* **68**, 18–21 (2008).
- Touvier, M. *et al.* Association between prediagnostic biomarkers of inflammation and endothelial function and cancer risk; a nested case-control study. *Am. J. Epidemiol.* **177**, 3–13 (2013).
- Wulaningsih, W. *et al.* Prediagnostic serum inflammatory markers in relation to breast cancer risk, severity at diagnosis and survival in breast cancer patients. *Carcinogenesis* **36**, 1121–1128 (2015).
- Guo, L. *et al.* C-reactive protein and risk of breast cancer: A systematic review and meta-analysis. *Sci. Rep.* **5**, 10508 (2015).
- Allen, M. D. & Jones, L. J. The role of inflammation in progression of breast cancer: Friend or foe? (Review). *Int. J. Oncol.* **47**, 797–805 (2015).
- Tobias, D. K. *et al.* Markers of Inflammation and Incident Breast Cancer Risk in the Women’s Health Study. *Am. J. Epidemiol.* **187**, 705–716 (2018).
- De Roos, A. J. *et al.* Markers of B-cell activation in relation to risk of non-Hodgkin lymphoma. *Cancer Res.* **72**, 4733–4743 (2012).
- Purdue, M. P. *et al.* Prediagnostic Serum Levels of Cytokines and Other Immune Markers and Risk of Non-Hodgkin Lymphoma. *Cancer Res.* **71**, 4898–4907 (2011).
- Hosnijeh, F. S. *et al.* Plasma Cytokines and Future Risk of Non-Hodgkin Lymphoma (NHL): A Case-Control Study Nested in the Italian European Prospective Investigation into Cancer and Nutrition. *Cancer Epidemiol. Prev. Biomark.* **19**, 1577–1584 (2010).
- Birmann, B. M. *et al.* Prediagnosis biomarkers of insulin-like growth factor-1, insulin, and interleukin-6 dysregulation and multiple myeloma risk in the Multiple Myeloma Cohort Consortium. *Blood* **120**, 4929–4937 (2012).
- Conroy, S. M. *et al.* Non-Hodgkin Lymphoma and Circulating Markers of Inflammation and Adiposity in a Nested Case-Control Study: The Multiethnic Cohort. *Cancer Epidemiol. Prev. Biomark.* **22**, 337–347 (2013).
- Vermeulen, R. *et al.* Circulating soluble CD30 and future risk of lymphoma; evidence from two prospective studies in the general population. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 1925–1927 (2011).
- Purdue, M. P. *et al.* A prospective study of serum soluble CD30 concentration and risk of non-Hodgkin lymphoma. *Blood* **114**, 2730–2732 (2009).
- Gu, Y. *et al.* Circulating cytokines and risk of B-cell non-Hodgkin lymphoma: a prospective study. *Cancer Causes Control CCC* **21**, 1323–1333 (2010).
- Edlefsen, K. L. *et al.* Cytokines in serum in relation to future non-Hodgkin lymphoma risk: Evidence for associations by histologic subtype. *Int. J. Cancer* **135**, 913–922 (2014).
- Purdue, M. P. *et al.* A prospective study of 67 serum immune and inflammation markers and risk of non-Hodgkin lymphoma. *Blood* **122**, 951–957 (2013).
- Kristinsson, S. Y. *et al.* Hodgkin lymphoma risk following infectious and chronic inflammatory diseases: a large population-based case-control study from Sweden. *Int. J. Hematol.* **101**, 563–568 (2015).
- Allin, K. H., Bojesen, S. E. & Nordestgaard, B. G. Inflammatory biomarkers and risk of cancer in 84,000 individuals from the general population. *Int. J. Cancer* **139**, 1493–1500 (2016).
- Morrison, L. *et al.* Inflammatory biomarker score and cancer: A population-based prospective cohort study. *BMC Cancer* **16**, (2016).
- Van Hemelrijck, M. *et al.* Association between levels of C-reactive protein and leukocytes and cancer: three repeated measurements in the Swedish AMORIS study. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **20**, 428–437 (2011).
- Rom, O., Avezov, K., Aizenbud, D. & Reznick, A. Z. Cigarette smoking and inflammation revisited. *Respir. Physiol. Neurobiol.* **187**, 5–10 (2013).
- González-Reimers, E., Santolaria-Fernández, F., Martín-González, M. C., Fernández-Rodríguez, C. M. & Quintero-Platt, G. Alcoholism: a systemic proinflammatory condition. *World J. Gastroenterol.* **20**, 14660–14671 (2014).
- Simpson, R. J. *et al.* Exercise and the aging immune system. *Ageing Res. Rev.* **11**, 404–420 (2012).

33. Ding, D., Rogers, K., van der Ploeg, H., Stamatakis, E. & Bauman, A. E. Traditional and Emerging Lifestyle Risk Behaviors and All-Cause Mortality in Middle-Aged and Older Adults: Evidence from a Large Population-Based Australian Cohort. *PLoS Med.* **12**, e1001917 (2015).
34. Ford, N. A., Devlin, K. L., Lashinger, L. M. & Hursting, S. D. Deconvoluting the Obesity and Breast Cancer Link: Secretome, Soil and Seed Interactions. *J. Mammary Gland Biol. Neoplasia* **18**, (2013).
35. Crespi, E., Bottai, G. & Santarpia, L. Role of inflammation in obesity-related breast cancer. *Curr. Opin. Pharmacol.* **31**, 114–122 (2016).
36. Iyengar, N. M., Hudis, C. A. & Dannenberg, A. J. Obesity and inflammation: new insights into breast cancer development and progression. *Am. Soc. Clin. Oncol. Educ. Book Am. Soc. Clin. Oncol. Meet.* 46–51, https://doi.org/10.1200/EdBook_AM.2013.33.46 (2013).
37. Dumeaux, V. *et al.* Peripheral blood cells inform on the presence of breast cancer: a population-based case-control study. *Int. J. Cancer* **136**, 656–667 (2015).
38. Gu, Y. *et al.* Circulating cytokines and risk of B-cell non-Hodgkin lymphoma: a prospective study. *Cancer Causes Control CCC* **21**, 1323–1333 (2010).
39. Lateef, N. *et al.* Malignancies After Heart Transplant. *Research Gate* **14** (2015).
40. Wang, Y.-J. *et al.* Malignancy After Heart Transplantation Under Everolimus Versus Mycophenolate Mofetil Immunosuppression. *Transplant. Proc.* **48**, 969–973 (2016).
41. Palli, D. *et al.* A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori* **89**, 586–593 (2003).
42. Hallmans, G. *et al.* Cardiovascular disease and diabetes in the Northern Sweden Health and Disease Study Cohort - evaluation of risk factors and their interactions. *Scand. J. Public Health Suppl.* **61**, 18–24 (2003).
43. Hebels, D. G. A. J. *et al.* Performance in omics analyses of blood samples in long-term storage: opportunities for the exploitation of existing biobanks in environmental health research. *Environ. Health Perspect.* **121**, 480–487 (2013).
44. Georgiadis, P. *et al.* Omics for prediction of environmental health effects: Blood leukocyte-based cross-omic profiling reliably predicts diseases associated with tobacco smoking. *Sci. Rep.* **6**, 20544 (2016).
45. de Jager, W., te Velthuis, H., Prakken, B. J., Kuis, W. & Rijkers, G. T. Simultaneous Detection of 15 Human Cytokines in a Single Sample of Stimulated Peripheral Blood Mononuclear Cells. *Clin. Diagn. Lab. Immunol.* **10**, 133–139 (2003).
46. Lubin, J. H. *et al.* Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ. Health Perspect.* **112**, 1691–1696 (2004).
47. Chadeau-Hyam, M. *et al.* Prediagnostic transcriptomic markers of Chronic lymphocytic leukemia reveal perturbations 10 years before diagnosis. *Ann. Oncol.* **25**, 1065–1072 (2014).
48. Castagné, R. *et al.* A life course approach to explore the biological embedding of socioeconomic position and social mobility through circulating inflammatory markers. *Sci. Rep.* **6**, 25170 (2016).
49. Ringné, M. What is principal component analysis? *Nat. Biotechnol.* **26**, 303–304 (2008).
50. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 73, 3–36 (2014).

Acknowledgements

The authors wish to thank all the centers that took part in the study and the additional members of the EnviroGenoMarkers Consortium. For further information on the EnviroGenoMarkers Consortium please visit <http://www.envirogenomarkers.net/>. This work has been carried out as part of the 7th European Framework Programme (FP7) Envirogenomarkers [226756 to S.K.] project. This work was supported by the Institut National du Cancer [SHSESP14-082 to R.C.]. This work was funded by the European Commission as part of the Lifepath project [H2020 grant number 633666]. This work was supported by La Ligue nationale contre le cancer [Equipe Labellisée LIGUE 2017/CD]).

Author Contributions

S.A.K. and P.V. conceived of the EnviroGenoMarkers project. E.B. performed the statistical analysis and wrote the first draft of manuscript. R.C., C.D., M.K.-I., M.C.-H. and R.V. helped to assist in editing of and revising the manuscript. F.S.-H, L.P., I.A.B, A.J., V.K., D.P. S.P., C.S., R.T. contributed to data acquisition, analysis, and interpretation. All authors critically revised the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-29041-1>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Consortia

EnviroGenoMarkers

Beatrice Melin⁵, Per Lenner⁵, Benedetta Bendinelli⁷, Maria Botsivali¹¹, Aristotelis Chatziioannou¹¹, Ioannis Valavanis¹¹, Barbara Bodinier¹⁴, Javiera Garrido-Manriquez¹⁴, Toby J. Athersuch^{14,15}, Benoît Liquet^{16,17}, Henk Lokhorst¹⁸, Panagiotis Georgiadis¹⁹, Jos C. S. Kleinjans²⁰, Theo M. C. M. de Kok²⁰, Hector C. Keun²¹, Rachel Kelly²², Goran Hallmans²³, Euripides G. Stephanou²⁴, Antonis Myridakis²⁴, Manolis Kogevinas²⁵, Lucia Fazzo²⁶, Marco De Santis²⁶, Pietro Comba²⁶, Hannu Kiviranta²⁷, Panu Rantakokko²⁷, Riikka Airaksinen²⁷, Paivi Ruokojarvi²⁷, Mark Gilthorpe²⁸, Sarah Fleming²⁸, Thomas Fleming²⁸, Yu-Kang Tu²⁸, Thomas Lundh²⁹, Kuo-Liong Chien³⁰, Wei J. Chen³⁰, Wen-Chung Lee³⁰, Chuhsing Kate Hsiao³⁰, Po-Hsiu Kuo³⁰, Hung Hung³⁰ & Shu-Fen Liao³⁰

¹⁵Division of Computational and Systems Medicine, Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK. ¹⁶Laboratoire de Mathématiques et de leurs Applications, Université de Pau et des Pays de l'Adour, UMR CNRS 5142, Pau, France. ¹⁷ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology (QUT), Brisbane, Australia. ¹⁸Department of Hematology, University Medical Center Utrecht, Utrecht, The Netherlands. ¹⁹National Hellenic Research Foundation, Institute of Biology, Pharmaceutical Chemistry and Biotechnology, Athens, Greece. ²⁰Department of Toxicogenomics, Maastricht University, Maastricht, The Netherlands. ²¹Division of Cancer, Department of Surgery and Cancer, Imperial College London, Institute of Reproductive and Developmental Biology (IRDB), Hammersmith Hospital, London, UK. ²²Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA. ²³Nutrition Research, Department of Public Health and Clinical Medicine, and Department of Biobank Research, Umeå University, Umeå, Sweden. ²⁴Environmental Chemical Processes Laboratory, University of Crete, Heraklion, Greece. ²⁵ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. ²⁶Istituto Superiore di Sanita, Rome, Italy. ²⁷National Institute for Health and Welfare, Kuopio, Finland. ²⁸University of Leeds, Leeds, United Kingdom. ²⁹Lund University, Lund, Sweden. ³⁰National Taiwan University, Taipei, Taiwan.

ANNEXE 4 - VARIABLES UTILISEES POUR L'IMPUTATION

Travail 3	Travail 4	Travail 5**	
		Imputation des covariables sur l'ensemble de la base (N=83 436)	Imputation des caractéristiques sur le sous ensemble de cancer du sein (N=6701)
Métier du père	Métier du père	Métier du père	Métier du père
Education	Education	Education	Education
Education du conjoint	Profession	Profession	Profession
Profession	Diagnostic de cancer du sein	Diagnostic de cancer du sein	Age
Diagnostic de cancer du sein	Age	Age	Alcool
Age	Tabac	Tabac	Tabac
Tabac	Activité physique	Activité physique	Régime occidental
Activité physique	Alcool	Alcool	Activité physique
Alcool	Régime occidental	Régime occidental	Taille
Régime occidental	Taille	Taille	Poids
Régime méditerranéen	Poids	Poids	IMC
Poids	IMC	IMC	Age aux premières règles
Taille	Antécédents familiaux cancers du sein-ovaires-utérus	Age aux premières règles	Antécédents familiaux cancers du sein-ovaires-utérus
IMC	Utilisation THM	Antécédents familiaux cancers du sein-ovaires-utérus	Contraceptif oral
Antécédents familiaux cancers du sein-ovaires-utérus	Contraceptif oral	Contraceptif oral	Utilisation THM
Contraceptif oral	Age à la première grossesse	Utilisation THM	Nombre de grossesses
Age au premier allaitement	Allaitement	Nombre de grossesses	Age à la première grossesse

Utilisation THM	Durée d'allaitement	Age à la première grossesse	Allaitement
Nombre de grossesses	Nombre de cycle avant la première grossesse	Allaitement	Durée d'allaitement
Age à la première grossesse	Nombre de grossesses	Durée d'allaitement	Durée de vie reproductive
Allaitement	Durée de vie reproductive	Durée de vie reproductive	Nombre de cycle avant la première grossesse
Durée d'allaitement	Dépistage de cancer du sein	Nombre de cycle avant la première grossesse	Dépistage de cancer du sein
Durée de vie reproductive	Statut ménopausique au moment du prélèvement	Dépistage de cancer du sein	Statut ménopausique à au diagnostic
Nombre de cycles avant la première grossesse	CRP*	Statut ménopausique à l'inclusion	Récepteurs aux oestrogènes
Dépistage de cancer du sein	HDL*		Récepteurs à la progestérone
Statut ménopausique à l'inclusion	TG*		Protéine HER2
	LDL*		Stade
	Cholestérol*		Grade
	Fréquence cardiaque*		
	PAS*		
	PAD*		
	Œstradiol*		
	Testostérone*		
	SHBG*		

*Variables en continu - Non imputées

**Les covariables utilisées pour le travail 5 ont été imputées à partir des informations de la population globale et les caractéristiques ont été imputées spécifiquement sur la sous population de cancer du sein

ANNEXE 5 - ARTICLE RELATIF AU TRAVAIL

RESEARCH ARTICLE

Open Access



The impact of lifecourse socio-economic position and individual social mobility on breast cancer risk

Eloïse Berger^{1*} , Noële Maitre^{2,3}, Francesca Romana Mancini^{2,3}, Laura Baglietto^{2,3,4}, Vittorio Perduca^{2,3,5}, Hélène Colineaux^{1,6}, Sabina Sieri⁷, Salvatore Panico⁸, Carlotta Sacerdote⁹, Rosario Tumino¹⁰, Paolo Vineis^{11,12}, Marie-Christine Boutron-Ruault^{2,3}, Gianluca Severi^{2,3,13}, Raphaële Castagné^{1†} and Cyrille Delpierre^{1†}

Abstract

Background: Women with an advantaged socioeconomic position (SEP) have a higher risk of developing breast cancer (BC). The reasons for this association do not seem to be limited to reproductive factors and remain to be understood. We aimed to investigate the impact of lifecourse SEP from childhood and social mobility on the risk of BC considering a broad set of potential mediators.

Methods: We used a discovery-replication strategy in two European prospective cohorts, E3N ($N = 83,436$) and EPIC-Italy ($N = 20,530$). In E3N, 7877 women were diagnosed with BC during a median 24.4 years of follow-up, while in EPIC-Italy, 893 BC cases were diagnosed within 15.1 years. Hazard ratios (HR) were estimated using Cox proportional hazard models on imputed data.

Results: In E3N, women with higher education had a higher risk of BC (HR [95%CI] = 1.21 [1.12, 1.30]). This association was attenuated by adjusting for reproductive factors, in particular age at first childbirth (HR[95%CI] = 1.13 [1.04, 1.22]). Health behaviours, anthropometric variables, and BC screening had a weaker effect on the association. Women who remained in a stable advantaged SEP had a higher risk of BC (HR [95%CI] = 1.24 [1.07; 1.43]) attenuated after adjustment for potential mediators (HR [95%CI] = 1.13 [0.98; 1.31]). These results were replicated in EPIC-Italy.

Conclusions: These results confirm the important role of reproductive factors in the social gradient in BC risk, which does not appear to be fully explained by the large set of potential mediators, including cancer screening, suggesting that further research is needed to identify additional mechanisms.

Keywords: Lifecourse socio-economic position, Social mobility, Breast cancer, Prospective cohorts

* Correspondence: eloise.berger@univ-tlse3.fr

†Raphaële Castagné and Cyrille Delpierre: these authors jointly supervised this work.

¹UMR LEASP, Université de Toulouse III, UPS, Inserm, Toulouse, France
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Women with an advantaged socioeconomic position (SEP) have a higher risk of developing breast cancer (BC) compared to their disadvantaged counterparts [1, 2].

The social inequalities in BC risk could be partly explained by socially stratified distribution of known BC risk factors [3–7]. In particular age at first childbirth and parity seem to explain a large part of the association between SEP and the risk of BC [8–15]. However, an independent association between SEP and risk of BC has also been observed after controlling for those reproductive factors [8, 9, 11, 14].

Altogether these studies highlight the importance of reproductive factors in the social gradient of BC but also suggest that other pathways and mechanisms are involved and remain to be characterized. The few studies that have additionally considered health behaviours or anthropometric factors have shown that women with higher early life SEP had a higher risk of BC, partly mediated by age at first childbirth and the number of children but not by health behaviours [9]. In addition, several studies suggest that screening could act as a mediator in the association between SEP and the risk of BC [8, 12, 16, 17].

Most available studies so far have mainly focused on either young adulthood or late adulthood SEP [8, 10–13, 15–18], and few have examined SEP at different time points within a lifecourse framework [9, 14, 17, 19]. Studying the respective and combined effect of childhood and adult SEP is needed because they may reveal different mechanisms involved in the social gradient of BC incidence.

One of the major limitations of previous studies is the lack of simultaneous consideration of all potential mediators identified over the last two decades. It remains unclear, whether all those mediators influence the association between SEP and the risk of BC, and which ones are the main drivers of the association. We therefore aimed to investigate the potential and mutual impact of a large set of reproductive factors, anthropometric characteristics, and health behaviours, within a lifecourse framework. The purpose was to simultaneously evaluate all potential mediators, and to establish to which extent they can explain the social inequalities associated with breast cancer risk.

More specifically, we first assessed the relationship between SEP, from childhood to adulthood, and the future risk of BC in the E3N cohort. We further examined the impact of health behaviours, anthropometric characteristics, reproductive factors, family history of hormone-related cancer, and BC screening on these relationships. Third, we investigated the lifecourse influence of each SEP by sequentially controlling for time-ordered SEP, and investigated the impact of social mobility on BC risk. Finally, to assess the robustness and test the generalisability of our results, we conducted an independent replication study in the EPIC-Italy cohort.

Methods

Study populations

E3N and EPIC-Italy are two cohorts included in the European Prospective Investigation into Cancer and Nutrition (EPIC) study and have been described in detail elsewhere [20, 21]. Additional information is available in Additional file 1.

Briefly, E3N includes 98,995 women aged 38–66 and insured by the Mutuelle Générale de l'Education Nationale (MGEN), a national health insurance plan that primarily covers teachers. At inclusion, participants provided a written informed consent for the study and filled in a questionnaire that collected information about anthropometric measures, lifestyle / behaviours, SEP, and health. Self-administered questionnaires have been subsequently sent every 2–3 years since 1990. The June 1993 questionnaire (Q3) included a detailed diet history questionnaire.

EPIC-Italy represents a total of 34,152 volunteers aged 30–75 years at inclusion and recruited from four centers in Italy in 1993–1998. At inclusion, participants have filled in a questionnaire that collected information about anthropometric measurements, lifestyle / behaviours, SEP, and health.

Lifecourse SEP

We selected lifecourse SEP among variables available in both cohorts. SEP was measured at three time points from childhood to adulthood based on self-reported information and classified as described elsewhere [22], with some adaptation as detailed in Additional file 2.

Self-reported father's occupation was used as a proxy for childhood SEP. E3N women were born 1925–1950 and grew up in the 1940–60s, when the father's occupation could be considered a good indicator of the household's socio-economic conditions. We applied 3 E-SEC categories: less advantaged occupations [lower clerical, services, and sales workers; skilled workers; semi and unskilled workers (Class 7–9 ESEC)]; medium occupations [small employers and self-employed; farmers; lower supervisors and technicians (Class 4, 5, and 6 ESEC)] and more advantaged occupations [higher professionals and managers, lower professionals and managers; higher clerical, services and sales workers (Class 1–3 ESEC)].

SEP in young adulthood was measured using the participant's education categorised in 3 groups: low level [primary or lower secondary school], middle level [higher secondary school], and high level attainment [tertiary education].

Adulthood SEP was measured by the women's own occupation, following the same categorisation as for the father's occupation.

In EPIC-Italy, there was a large proportion of housewives, thus we used the highest household occupation to define adulthood SEP.

Follow up and outcomes

In E3N, self-reported BC diagnosis was confirmed through pathological reports. Women were followed from inclusion (Q1, 1990) to the date of BC diagnosis last filled in questionnaire, or end of the study (2014), whichever occurred first.

In EPIC-Italy, cancer cases were identified through automated linkage to cancer and mortality registries, or through active follow-up, and confirmed with histological reports or hospital discharge systems. Follow up time started at inclusion (1993–1998) to the last follow up: December 2010 for women from Naples, Turin, and Ragusa, and to December 2009 for women from Varese.

Covariates

Description of all covariates is provided in Additional file 3. The following factors have been considered as intermediate variables that could mediate the relationship between SEP and the risk of BC, and categorised in two blocks: i) Health behaviours and anthropometric characteristics including *alcohol consumption; smoking status; physical activity; Western diet pattern; weight; and height*; ii) reproductive factors including *menopausal status; use of menopausal hormone therapy (MHT); age at first childbirth; breastfeeding; reproductive lifespan*. We also included *family history of a hormone-related cancer (ovarian and breast); and BC screening*.

Description of all variables used is provided in Fig. 1.

In EPIC-Italy, available data were not exactly the same as those available in E3N. We selected and considered i) health behaviours and anthropometric characteristics: *alcohol consumption, smoking status, physical activity* (adapted from [23]), *Mediterranean diet* [24], *height, and weight*; and ii) reproductive factors: *age at first childbirth, and menopausal status*.

Statistical analyses in E3N

Descriptive analyses

Baseline characteristics of the population were compared by BC status and for each SEP indicator. Chi-squared test or Fisher’s exact test for categorical variables, and T-test or Wilcoxon rank test for continuous variables were used to estimate bivariate associations with SEP and log rank for the association with the risk of BC. Only those associated with both SEP and BC (*p*-value < 0.2) were considered in the multivariate and lifecourse analyses, except for smoking which was systematically included since it has been found associated with both SEP and BC in the literature.

Multivariate analyses

Cox proportional hazard models were used to investigate the relationship between SEP and BC risk. We defined a first model only adjusted for age (Model1). From Model 1, we adjusted for each potential mediator from each block of variables and also by block of variables as defined above. Finally, we defined a fully adjusted model

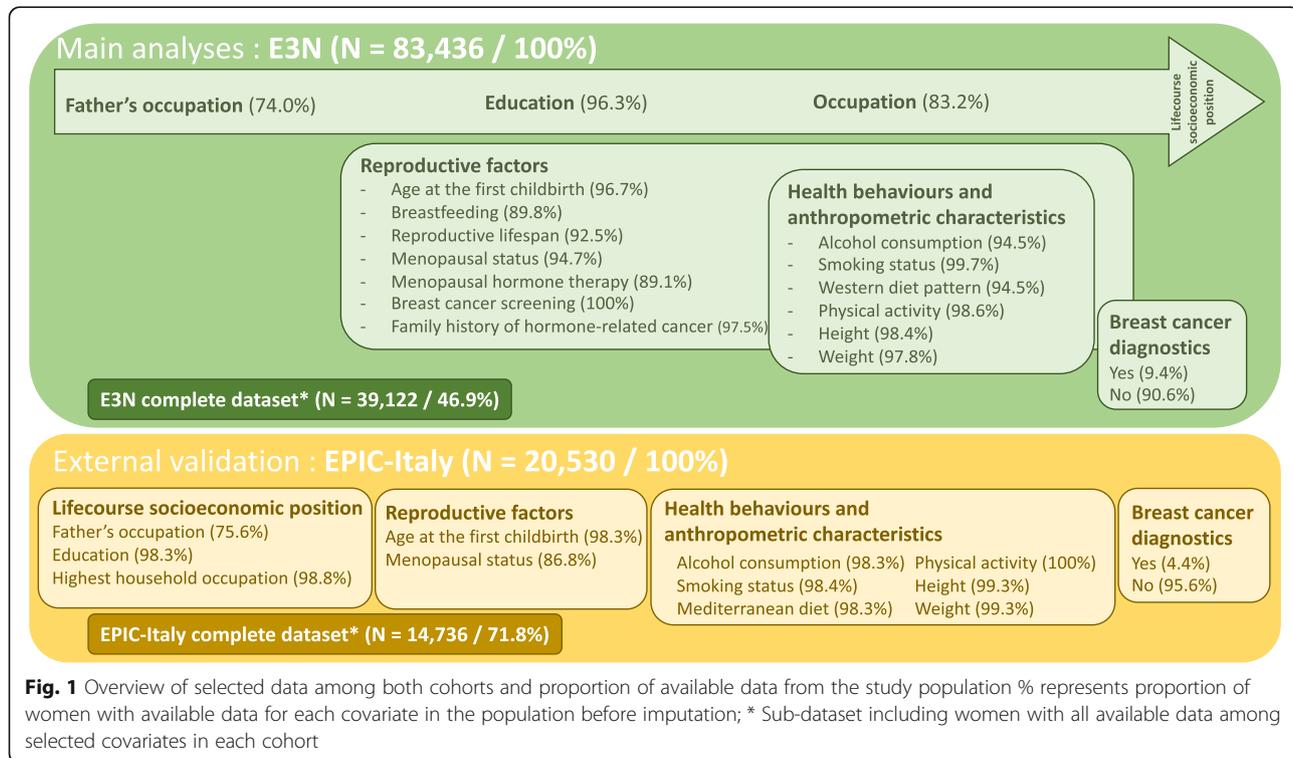


Fig. 1 Overview of selected data among both cohorts and proportion of available data from the study population % represents proportion of women with available data for each covariate in the population before imputation; * Sub-dataset including women with all available data among selected covariates in each cohort

including both blocks of variables. For each of the three lifecourse SEP, the disadvantaged category was used as the reference.

Given the large population of E3N, even the smallest deviation can lead to a statistical violation of the proportional risk assumptions as we observed for some variables; nevertheless, no trend seems to emerge looking at residuals graphically. So we used all variables as described previously. (Additional file 4).

Lifecourse analyses

To mimic lifecourse experiences, we sequentially adjusted for each SEP resulting in four time-sequenced models as:

- Model A: age + father's occupation
- Model B: Model A + education
- Model C: Model A + education + occupation
- Model D: Model A + education + occupation + covariates

Social mobility

A multiplicative interaction term was introduced for the father's occupation and the adult occupation, hence defining 5 classes: 'Stable disadvantaged SEP' (reference); 'Downward mobility'; 'Stable medium SEP'; 'Upward mobility'; 'Stable advantaged SEP'.

Missing data

To control for possible bias due to missing data, they were imputed using multiple imputation in the overall population with the MICE R package [25]. Fifteen imputations were conducted taking the missing-at-random assumption. All variables tested in the first study step were imputed, including lifecourse SEP. We used Rubin's combination rules to obtain Cox regression estimates from the multiple imputed data.

Sensitivity analyses

We restricted our analyses to women with complete information on SEP and covariates to test the robustness of our results ($N = 39,122$). To take into account potential specificities of invasive BC compared to in situ BC, we also ran Cox proportional hazards regression in women with invasive BC only (87.6% of all incident BC cases in E3N).

Replication analyses in EPIC-Italy

The same multivariate, lifecourse, and social mobility analyses were replicated in EPIC-Italy. Baseline model (Model1) in EPIC-Italy was further adjusted for center because of the study design. Multiple imputations were also performed in the overall population.

Statistical analyses were performed using R. version 1.2.1114 within R studio version 1.2.5001.

Results

E3N cohort

From the entire cohort, women with prevalent cancer, those with another cancer than BC as well as those with missing or inconsistent data on date or status of diagnosis were excluded from the analyses leaving a total of 83,436 women.

Selected characteristics for E3N by BC status are provided in Table 1. During a 24.4-year median follow-up time, 7877 women were diagnosed with BC. Compared to women who did not develop BC, cases were more likely to have an advantaged SEP along the lifecourse, to be older, to be heavy drinkers, to have a high adherence to a Western diet pattern, to be inactive or have low physical activity, and to be in the highest tercile of height and weight. Women diagnosed with BC were also more likely to have a family history of a hormone-related cancer, to be MHT ever user, to have fewer children, to have been older at first childbirth, not to have breastfed, to have a longer reproductive lifespan, to be in pre-menopause at baseline, and to have ever participated in mammographic screening before inclusion. Characteristics by SEP are available in Additional file 5).

Association between SEP and the risk of BC

Women whose fathers had an advantaged SEP had a higher BC risk compared with women whose fathers had a disadvantaged SEP (M1: HR [95%CI] = 1.10 [1.02, 1.18], Table 2a). Associations were slightly attenuated when adjusting for health behaviours and anthropometric characteristics (HR [95%CI] = 1.08 [1.01, 1.16]), reproductive factors (HR [95%CI] = 1.05 [0.98, 1.13]), or for all covariates (M2, HR [95%CI] = 1.04 [0.97, 1.12]). Women whose fathers had a medium SEP were not at higher risk for BC.

We found a positive association between women's education and BC risk (M1: HR [95%CI] = 1.21 [1.12, 1.30], Table 2b). Controlling for health behaviours and anthropometric characteristics only slightly attenuated the associations (HR [95%CI] = 1.18 [1.09, 1.27]) while controlling for reproductive factors, in particular age at first childbirth, affected it more strongly (HR[95%CI] = 1.06 [0.99, 1.15]). HRs were further reduced after controlling for all covariates (M2: HR [95%CI] = 1.06 [0.98, 1.14]).

Compared with their disadvantaged counterparts, women with an advantaged occupation had a higher risk of BC (HR[95%CI] = 1.08 [1.00, 1.17], Table 2c). HRs were slightly attenuated after adjustment for either each block of covariates or both (M2, HR[95%CI] = 1.05 [0.97, 1.14], Table 2c). No evidence of an increased risk was observed for women with medium SEP.

Table 1 Characteristics of women with available data from E3N according to BC status

Variables	BC		p
	No	Yes	
Father's occupation (3 cl)	n = 55,766	n = 5957	0.025
Advantaged, n(%)	9294 (16.67)	1075 (18.05)	
Medium, n(%)	23,619 (42.35)	2492 (41.83)	
Disadvantaged, n(%)	22,853 (40.98)	2390 (40.12)	
Education (3 cl)	n = 72,773	n = 7577	< 0.001
High, n(%)	26,096 (35.86)	2913 (38.45)	
Middle, n(%)	36,422 (50.05)	3745 (49.43)	
Low, n(%)	10,255 (14.09)	919 (12.13)	
Occupation (3 cl)	n = 62,726	n = 6700	0.017
Advantaged, n(%)	10,962 (17.48)	1250 (18.66)	
Medium, n(%)	40,679 (64.85)	4332 (64.66)	
Disadvantaged, n(%)	11,085 (17.67)	1118 (16.69)	
Age	n = 75,559	n = 7877	< 0.001
< 47.9y, n(%)	39,491 (52.27)	3837 (48.71)	
≥ 47.9y, n(%)	36,068 (47.73)	4040 (51.29)	
Alcohol consumption	n = 71,357	n = 7455	< 0.001
Abstainer, n(%)	7040 (9.87)	678 (9.09)	
Moderate, n(%)	38,720 (54.26)	4167 (55.9)	
High, n(%)	10,975 (15.38)	1356 (18.19)	
Not responding to Q3, n(%)	14,622 (20.49)	1254 (16.82)	
Smoking status	n = 75,333	n = 7850	0.693
Never, n(%)	50,531 (67.08)	5269 (67.12)	
Former, n(%)	15,204 (20.18)	1560 (19.87)	
Ever, n(%)	9598 (12.74)	1021 (13.01)	
Western diet pattern	n = 71,358	n = 7455	< 0.001
[-2.93,-0.511], n(%)	18,876 (26.45)	1921 (25.77)	
(-0.511,0.307], n(%)	18,868 (26.44)	2089 (28.02)	
(0.307,7.78], n(%)	18,991 (26.61)	2191 (29.39)	
Not responding to Q3, n(%)	14,623 (20.49)	1254 (16.82)	
Physical activity (MET)	n = 74,540	n = 7768	< 0.001
(7.57,35.9], n(%)	25,040 (33.59)	2395 (30.83)	
(4.66,7.57], n(%)	24,635 (33.05)	2630 (33.86)	
[0,4.66], n(%)	24,865 (33.36)	2743 (35.31)	
Height	n = 74,316	n = 7759	< 0.001
[135,160], n(%)	33,983 (45.73)	3355 (43.24)	
(160,164], n(%)	16,551 (22.27)	1732 (22.32)	
(164,190], n(%)	23,782 (32)	2672 (34.44)	
Weight	n = 73,847	n = 7715	0.076
[29,55], n(%)	25,135 (34.04)	2530 (32.79)	
(55,62], n(%)	24,737 (33.5)	2611 (33.84)	
(62,163], n(%)	23,975 (32.47)	2574 (33.36)	
Previous ovary cancer (1grade)	n = 75,559	n = 7877	0.002
No, n(%)	74,709 (98.88)	7757 (98.48)	

Table 1 Characteristics of women with available data from E3N according to BC status (*Continued*)

Variables	BC		<i>p</i>
Yes, <i>n</i> (%)	850 (1.12)	120 (1.52)	
Previous breast cancer (1grade)	<i>n</i> = 73,693	<i>n</i> = 7673	< 0.001
No, <i>n</i> (%)	31,622 (42.91)	3039 (39.61)	
Yes, <i>n</i> (%)	5240 (7.11)	886 (11.55)	
Not available, <i>n</i> (%)	36,831 (49.98)	3748 (48.85)	
MHT use	<i>n</i> = 67,245	<i>n</i> = 7111	< 0.001
No, <i>n</i> (%)	58,868 (87.54)	6021 (84.67)	
Yes, <i>n</i> (%)	6029 (8.97)	806 (11.33)	
Undefined, <i>n</i> (%)	2348 (3.49)	284 (3.99)	
Number of full term pregnancy	<i>n</i> = 75,551	<i>n</i> = 7877	< 0.001
3+, <i>n</i> (%)	22,212 (29.4)	2000 (25.39)	
1–2, <i>n</i> (%)	44,491 (58.89)	4827 (61.28)	
0, <i>n</i> (%)	8848 (11.71)	1050 (13.33)	
Age at the first childbirth	<i>n</i> = 73,018	<i>n</i> = 7660	< 0.001
[14, 23], <i>n</i> (%)	26,390 (36.14)	2445 (31.92)	
(23,26], <i>n</i> (%)	19,487 (26.69)	1959 (25.57)	
(26,59], <i>n</i> (%)	18,293 (25.05)	2206 (28.8)	
No preg, <i>n</i> (%)	8848 (12.12)	1050 (13.71)	
Combined age and number of pregnancy	<i>n</i> = 73,016	<i>n</i> = 7660	< 0.001
Early first birth and high number of children, <i>n</i> (%)	11,119 (15.23)	938 (12.25)	
High number of children, <i>n</i> (%)	10,150 (13.9)	997 (13.02)	
Late first birth and few number of children, <i>n</i> (%)	14,575 (19.96)	1794 (23.42)	
Low number of children, <i>n</i> (%)	28,324 (38.79)	2881 (37.61)	
No pregnancy, <i>n</i> (%)	8848 (12.12)	1050 (13.71)	
Breastfeeding	<i>n</i> = 67,708	<i>n</i> = 7189	0.019
Yes, <i>n</i> (%)	40,113 (59.24)	4156 (57.81)	
No, <i>n</i> (%)	27,595 (40.76)	3033 (42.19)	
Reproductive lifespan	<i>n</i> = 69,697	<i>n</i> = 7508	< 0.001
[27; 41[, <i>n</i> (%)	56,083 (80.47)	5964 (79.44)	
< 27, <i>n</i> (%)	1161 (1.67)	80 (1.07)	
≥ 41, <i>n</i> (%)	12,453 (17.87)	1464 (19.5)	
Breast cancer screening	<i>n</i> = 75,559	<i>n</i> = 7877	< 0.001
Yes, <i>n</i> (%)	52,720 (69.77)	6012 (76.32)	
No, <i>n</i> (%)	22,839 (30.23)	1865 (23.68)	
Postmenopausal status	<i>n</i> = 71,368	<i>n</i> = 7675	0.016
Pre-menopausal, <i>n</i> (%)	41,683 (58.41)	4593 (59.84)	
Post-menopausal, <i>n</i> (%)	29,685 (41.59)	3082 (40.16)	

P-values are estimated with log rank tests

Regarding the impact of other covariates on the risk of BC, higher age, high alcohol consumption, higher adherence to a Western diet pattern, lower physical activity, higher height, a family history of hormone-related cancer, use of MHT, and higher age at first childbirth were associated with a higher risk of BC in the fully adjusted model. Inversely, women with a shorter reproductive lifespan, no

adherence to BC screening, and who were in post-menopause had a lower risk of BC (Additional file 6).

Lifecourse SEP and the risk of BC

When we sequentially adjusted for each lifecourse SEP, only education was associated with BC risk. HRs for education were not affected when adjusting for childhood SEP

Table 2 Cox proportional hazard regression of BC risk using data from multiple imputation in E3N (N = 83,436)

		A. Father's occupation ^a		B. Education ^b		C. Occupation ^a	
		Medium	Advantaged	Middle	High	Medium	Advantaged
		HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
M1		1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.01, 1.17]	1.21 [1.12, 1.30]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
Health behaviours And Anthropometric factors	M1 + Alcohol consumption	1.00 [0.95, 1.06]	1.09 [1.02, 1.17]	1.09 [1.02, 1.17]	1.20 [1.11, 1.29]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Smoking status	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.02, 1.17]	1.21 [1.12, 1.30]	1.03 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Western diet pattern	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.10 [1.02, 1.18]	1.21 [1.13, 1.31]	1.03 [0.97, 1.09]	1.09 [1.00, 1.18]
	M1 + Physical activity	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.00, 1.16]	1.18 [1.10, 1.28]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + Height	1.00 [0.94, 1.05]	1.09 [1.01, 1.17]	1.09 [1.01, 1.17]	1.19 [1.11, 1.28]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + Weight	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.10 [1.02, 1.18]	1.21 [1.13, 1.31]	1.03 [0.97, 1.09]	1.08 [1.00, 1.17]
	M1 + all health behaviours and anthropometrics factors	1.00 [0.94, 1.05]	1.08 [1.01, 1.16]	1.08 [1.00, 1.16]	1.18 [1.09, 1.27]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
Reproductive factors	M1 + Family history of ovarian cancer	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.01, 1.17]	1.20 [1.12, 1.30]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + Family history of breast cancer	1.00 [0.94, 1.05]	1.09 [1.01, 1.17]	1.09 [1.01, 1.17]	1.19 [1.11, 1.29]	1.02 [0.96, 1.09]	1.08 [1.00, 1.17]
	M1 + MHT use	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.08 [1.01, 1.16]	1.20 [1.11, 1.29]	1.02 [0.96, 1.08]	1.08 [0.99, 1.16]
	M1 + Age at the first childbirth	0.99 [0.94, 1.05]	1.07 [1.00, 1.15]	1.07 [1.00, 1.15]	1.13 [1.04, 1.22]	1.01 [0.95, 1.07]	1.07 [0.98, 1.16]
	M1 + Breastfeeding	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.09 [1.02, 1.17]	1.21 [1.12, 1.30]	1.03 [0.97, 1.09]	1.08 [1.00, 1.17]
	M1 + Reproductive lifespan	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.01, 1.16]	1.19 [1.11, 1.29]	1.02 [0.96, 1.08]	1.07 [0.99, 1.16]
	M1 + BC screening	1.00 [0.94, 1.05]	1.09 [1.02, 1.17]	1.08 [1.00, 1.16]	1.19 [1.10, 1.28]	1.02 [0.96, 1.08]	1.08 [1.00, 1.17]
M1 + Postmenopausal status	1.00 [0.95, 1.06]	1.10 [1.02, 1.18]	1.08 [1.01, 1.16]	1.19 [1.11, 1.28]	1.02 [0.96, 1.08]	1.08 [1.00, 1.17]	
Model 1 + all reproductive factors	0.98 [0.93, 1.04]	1.05 [0.98, 1.13]	1.03 [0.96, 1.11]	1.06 [0.99, 1.15]	0.99 [0.93, 1.05]	1.06 [0.98, 1.15]	
M2	0.98 [0.93, 1.04]	1.04 [0.97, 1.12]	1.03 [0.96, 1.11]	1.06 [0.98, 1.14]	0.99 [0.93, 1.05]	1.05 [0.97, 1.14]	

M1 is adjusted for age

^aReferent group: "Disadvantaged"

^bReferent group: "Low education"

M2 is fully adjusted model

Hazard ratio (HR) and confidence interval are reported for (A) father's occupation (B) education and (C) occupation

(Model B, HR[95%CI] = 1.19 [1.11; 1.29], Table 3) or both childhood and adulthood SEP (Model C, HR[95%CI] = 1.23 [1.12; 1.35], Table 3). The association was attenuated after adjustment for age at first childbirth (HR[95%CI] = 1.14 [1.04; 1.25], data not shown) and HRs were weakened in the fully adjusted model (Model D, R[95%CI] = 1.06 [0.97; 1.17], Table 3).

Effect of social mobility on the risk of BC

Women who experienced a stable advantaged SEP had a higher risk of BC than those with a stable disadvantaged SEP (Model 1: HR[95%CI] = 1.24 [1.07; 1.43], Table 4). HR estimates were attenuated, especially after adjustment for age at first childbirth (HR[95%CI] = 1.19 [1.03; 1.37, data not shown), and after adjustment for all covariates (Fully adjusted model: HR[95%CI] = 1.13 [0.98; 1.31], Table 4).

Sensitivity analyses

Complete cases analyses showed similar results although associations with BC risk were slightly stronger for education and occupation (Additional file 7). HR estimates

for education were weakened but not entirely explained after accounting for all covariates. A similar pattern was observed with participants' occupation. When we restricted our analyses to women who developed an invasive BC, results were comparable (Additional file 8).

External validation in EPIC-Italy

Results in EPIC-Italy are provided in Additional file 9, 10, 11, 12. Briefly, HRs for the risk of BC in highly educated women were similar to those observed in E3N (Model 1: HR[95%CI] = 1.19 [0.96; 1.47], Additional file 10). Adjustment for age at first childbirth attenuated mostly the relation (HR[95%CI] = 1.05 [0.84; 1.31]). Association between education and the risk of BC was not affected by other SEP indicators as we observed for E3N (Model C: HR[95%CI] = 1.18 [0.94; 1.49], Additional file 11). Women with a stable medium or stable advantaged SEP had a higher risk of BC compared to stable disadvantaged SEP, and HRs were marginally affected after adjustment for covariates (Additional file 12).

Table 3 Lifecourse multiple regression analyses of SEP with the risk of BC in E3N using imputed data (N = 83,436)

			E3N			
Covariates	Reference	Modality	Model A	Model B	Model C	Model D
			HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Father's occupation	Disadvantaged	Medium	1.00 [0.95; 1.06]	0.98 [0.93; 1.04]	0.98 [0.93; 1.04]	0.98 [0.93; 1.04]
		Advantaged	1.10 [1.02; 1.18]	1.06 [0.98; 1.14]	1.05 [0.98; 1.13]	1.03 [0.96; 1.11]
Education	Low	Middle	–	1.09 [1.01; 1.17]	1.12 [1.02; 1.22]	1.04 [0.95; 1.14]
		High	–	1.19 [1.11; 1.29]	1.23 [1.12; 1.35]	1.06 [0.97; 1.17]
Occupation	Disadvantaged	Medium	–	–	0.95 [0.88; 1.02]	0.97 [0.90; 1.04]
		Advantaged	–	–	1.00 [0.91; 1.09]	1.03 [0.94; 1.13]

Model A is adjusted for age and father's occupation

Model B is adjusted for age, father's occupation and education

Model C is adjusted for age and both SEP

Model D is adjusted for age, both SEP and all covariates (i.e. alcohol consumption, smoking status, physical activity, Western diet pattern, height, weight, family history of ovarian cancer or BC, MHT use, breastfeeding, cancer screening, reproductive lifespan, age at first childbirth and menopausal status)

Discussion

In a large prospective cohort with available lifecourse SEP and a wide array of covariates, we found that women with an advantaged SEP had a higher risk of BC at each considered time point. Lifecourse analyses suggested that education had a stronger effect on BC risk. Analyses on social mobility indicated that women who stay in a stable advantaged SEP had a higher risk of BC than those remaining in stable disadvantaged SEP. Associations were weaker when adjusting for reproductive factors. Age at first childbirth was the strongest contributor to SEP-associated BC risk. Adjustment for BC screening participation or for health behaviours and anthropometric characteristics only marginally modified the association between education and BC risk. Results were robust to the sensitivity analyses we performed and strengthened by the external validation in EPIC-Italy.

Our study adds to the literature on the effect of lifecourse SEP on the risk of BC. In agreement with previous studies, education was the SEP indicator most strongly related to BC risk [9, 14, 17] and the association was mainly weakened by reproductive factors [9–15]. But we also show that this impact persists even after considering the main other potential mediators, which constitutes an original finding of our work.

Studies on the etiology of BC have highlighted the major role of BC risk factors such as weight at birth, early age to first menstruation, alcohol consumption, age at first childbirth [26], or late age at menopause [27]. All those factors are related, to various extents, to hormonal pathways. Our study supports the importance of age at first childbirth, corresponding to the time when maturation of the breast tissue ends [28], on BC risk.

We initially hypothesized that the remaining association after adjustment for reproductive factors could be explained by the fact that higher educated women are more likely to participate in mammographic screening [29]. But we observed only a modest effect on risk by adjusting for BC screening, in agreement with previous studies [13, 18]. Adjustment for health behaviours and anthropometric characteristics also had a modest effect on the association between education and BC risk.

According to the literature, some of the hormone-related risk factors for BC occur early in life, such as birth weight or age at menarche, and have been found to be associated with SEP [30–32]. This suggests that SEP in early life could be important. Our results are not in favour of the hypothesis of a socially differentiated early sensitive period in BC risk. Results on social mobility

Table 4 Association of social mobility with the risk of BC in E3N using imputed data (N = 83,436)

			E3N			
Covariates	Reference	Modality	Model 1	Model 1 + HB-A	Model 1 + RF	Fully adjusted model
			HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Social mobility	Stable disadvantaged SEP	Downward mobility	1.06 [0.95; 1.19]	1.04 [0.93; 1.17]	1.00 [0.90; 1.12]	1.00 [0.89; 1.12]
		Stable medium SEP	1.02 [0.92; 1.12]	1.01 [0.91; 1.11]	0.96 [0.87; 1.06]	0.96 [0.87; 1.06]
		Upward mobility	1.04 [0.94; 1.14]	1.03 [0.93; 1.13]	1.00 [0.91; 1.10]	0.99 [0.90; 1.09]
		Stable advantaged SEP	1.24 [1.07; 1.43]	1.20 [1.04; 1.39]	1.15 [0.99; 1.33]	1.13 [0.98; 1.31]

Model 1 is adjusted for age and social mobility

Model 1 + HB-A is adjusted for age, social mobility, alcohol consumption, smoking status, physical activity, Western diet pattern, height, weight

Model 1 + RF is adjusted for age, social mobility, family history of ovarian cancer or BC, MHT use, breastfeeding, cancer screening, reproductive lifespan, age at first childbirth and menopausal status

Fully adjusted model is adjusted for age, social mobility and all covariates

show that women in an advantaged SEP throughout their lives are the most at risk group, suggesting that these women would cumulate harmful effects over the lifecourse. Additional and more specific analyses are needed to better define the impact of risk factors accumulation across the lifecourse.

This work was conducted on a large prospective cohort of French women in which a very large number and breadth of potential mediators were available. The prospective design limits both recall and reverse causation biases. Women have been followed since the 1990s until now, which allows us to have a long view on the disease development. Self-reported cancer cases were validated avoiding misclassification. Additionally, the use of a second prospective and independent cohort, EPIC-Italy, enabled us to replicate our findings, providing an external validation.

The main limitation in E3N lies in its recruitment especially when focusing on social inequalities. Women volunteers included in this study were all affiliated with a national health system (MGEN) that mainly insures people working in the French education system and spouses. Although the cohort also includes administrative and cleaning staff, the average educational level of the cohort is higher compared to the one in the general population. However, by considering SEP at different life periods from childhood, we were able to observe a certain degree of variability. There is likely to be heterogeneity, measurement and misclassification errors in both cohorts regarding the 3 life course SEP indicators. However these individual-based measurements of SEP could capture individual factors (e.g. material, behavioural, or psychosocial factors), provide information about individuals' accesses to social and economic resources, and be related to macro-environmental features (e.g. geographical location). We cannot rule out selection bias due to attrition and loss of follow-up. To allow for uncertainty about the missing data, we ran multiple imputations and analyses on complete cases, which provided consistent results. Even if we considered one of the largest set of covariates, it is still possible that other factors may contribute in the relationship between SEP and BC. In particular, breast density appears to be an interesting risk factors of BC to consider [33, 34]. Several studies have reported an independent association between SEP, including education, and breast density after accounting for the potential mediators we considered in our study. Breast density is suspected to modulate estrogen level which could be at the origin of the risk of BC. We could think that breast density may reflect another path modifying level of estrogens that has not been taken into account here. Alternatively, other mechanisms could be investigated, including the impact of perceived stress on levels of biomarkers suspected to be involved in BC risk [35–38]. The approach we used in our study allowed us to identify potential mediators from a

large range of factors. A better understanding of the causal pathways through which educational processes operate is now needed using causal approaches.

Conclusion

In the E3N cohort, women with advantaged SEP along the life course had an increased risk of BC. Among the three investigated SEP time points, education was the factor most strongly related to subsequent BC risk. Accounting for the large set of mediators we studied, age at first childbirth explained an important part of the observed association between SEP and BC risk. Other potential mediators, including BC screening, had a lesser effect on the association. The association between life course SEP and BC risk was not fully explained suggesting that further research is needed to identify additional mediators. The association between a stable advantaged SEP and BC risk suggests a cumulative damaging effect of advantaged SEP across the life course. Our results from an independent cohort from EPIC-Italy were consistent in terms of direction and size of the effect compared to those obtained in E3N, although with less power due to reduced cohort size. Finally, studies need to be developed to identify the causal mechanisms of a higher BC risk in women with advantaged SEP over the lifecourse.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-020-07648-w>.

Additional file 1. Cohorts' description.

Additional file 2. Coding specificities.

Additional file 3. Description of all covariates selected according to the literature and tested in bivariate analyses.

Additional file 4. Scatterplots of Schoenfeld residues for variables that do not respect the proportional risk assumption.

Additional file 5. Characteristics of women with available data from E3N according to SEP.

Additional file 6 Forestplot of the association of the three time point SEP and each covariate used in the fully adjusted model in E3N (N = 83,436).

Additional file 7 Association between life course SEP on the risk of BC using complete cases in E3N [N = 39,122].

Additional file 8 Association between life course SEP on the risk of invasive BC only using imputed data in E3N (N = 82,458).

Additional file 9. Characteristics of women with available data from EPIC-Italy according to BC status and by SEP.

Additional file 10 Cox proportional hazard regression of BC risk using data from multiple imputation in EPIC-Italy (N = 20,530).

Additional file 11. Lifecourse multiple regression analyses of SEP with the future risk of BC in EPIC-Italy using imputed data [N = 20,530].

Additional file 12. Association of social mobility with the risk of BC in EPIC-Italy using imputed data [N = 20,530].

Abbreviations

BC: Breast cancer; HR: Hazard ratio; MHT: Menopausal hormone therapy; SEP: Socioeconomic position

Acknowledgements

This work was supported by La Ligue nationale contre le cancer [Equipe Labellisée LIGUE 2017/CD]. This study has been conducted by using data from the E3N cohort that is managed by Inserm and that has been created and is maintained thanks to the support of the MGEN, the Institute Gustave Roussy and « La Ligue contre le Cancer ».

Authors' contributions

CD, RC, PV, GS designed research; NM, FM, SS, SP, CS, RT, PV, MCB, GS collected and controlled the data; NM contributed to data acquisition and preparation. EB performed the statistical analysis and wrote the first draft of manuscript; CD, RC, EB, PV, CS, LB, VP, HC and GS discussed the results. All authors critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the French National Institute of Cancer [SHSESP 2017–130 to CD]. The funder had no role in the study design, analysis and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available and permission to use the data is restricted to the teams in charge of the cohorts, which can be extended to collaborators with a specific research agreement.

Ethics approval and consent to participate

The E3N study was approved by the French Commission for Data Protection and Privacy (NCT03285230, CNIL/Commission nationale informatique et libertés - no. 327346 V 13). For EPIC-Italy, the ethical review boards of the International Agency for Research on Cancer, and the ethics committee of the Local Health Authority of Florence in Italy approved the study. For both studies, all subjects signed an informed consent form at study entry.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing financial interests.

Author details

¹UMR LEASP, Université de Toulouse III, UPS, Inserm, Toulouse, France. ²CESP, Faculté de Médecine, Université Paris-Saclay, UVSQ, INSERM, Villejuif, France. ³Gustave Roussy, F-94805 Villejuif, France. ⁴Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. ⁵Université de Paris, CNRS, MAP5 UMR 8145, F-75006 Paris, France. ⁶Epidemiology Department, Toulouse Teaching Hospital, Toulouse, France. ⁷Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. ⁸Department of Clinical Medicine and Surgery, University of Naples Federico II, Naples, Italy. ⁹Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital and Center for Cancer Prevention (CPO), Turin, Italy. ¹⁰Cancer Registry and Histopathology Department, Provincial Health Authority (ASP) Ragusa, Ragusa, Italy. ¹¹Department of Epidemiology and Biostatistics, Imperial College London, MRC-PHE Centre for Environment and Health, School of Public Health, London, UK. ¹²Italian Institute for Genomic Medicine, Torino, Italy. ¹³Department of Statistics, Computer Science and Applications "G. Parenti" (DISIA), University of Florence, Florence, Italy.

Received: 22 June 2020 Accepted: 17 November 2020

Published online: 23 November 2020

References

- Hvidtfeldt UA. Mechanisms underlying social inequality in post-menopausal breast cancer. *Dan Med J*. 2014;61(10):B4922.
- Lundqvist A, Andersson E, Ahlberg I, Nilbert M, Gerdtham U. Socioeconomic inequalities in breast cancer incidence and mortality in Europe—a systematic review and meta-analysis. *Eur J Pub Health*. 2016;26(5):804–13.
- dos Santos Silva I, Beral V. Socioeconomic differences in reproductive behaviour. *IARC Sci Publ*. 1997;(138):285–308.
- Cavelaars AE, Kunst AE, Geurts JJ, Crialesi R, Grötvedt L, Helmert U, et al. Persistent variations in average height between countries and between socio-economic groups: an overview of 10 European countries. *Ann Hum Biol*. 2000;27(4):407–21.
- Barriuso L, Miqueleiz E, Albaladejo R, Villanueva R, Santos JM, Regidor E. Socioeconomic position and childhood-adolescent weight status in rich countries: a systematic review, 1990–2013. *BMC Pediatr*. 2015;15:129.
- Petrovic D, de Mestral C, Bochud M, Bartley M, Kivimäki M, Vineis P, et al. The contribution of health behaviors to socioeconomic inequalities in health: a systematic review. *Prev Med*. 2018;113:15–31.
- Martikainen P, Brunner E, Marmot M. Socioeconomic differences in dietary patterns among middle-aged men and women. *Soc Sci Med*. 2003;56(7):1397–410.
- Hussain SK, Altieri A, Sundquist J, Hemminki K. Influence of education level on breast cancer risk and survival in Sweden between 1990 and 2004. *Int J Cancer*. 2008;122(1):165–9.
- Pudrovska T, Anikputa B. The role of early-life socioeconomic status in breast cancer incidence and mortality: unraveling life course mechanisms. *J Aging Health*. 2012;24(2):323–44.
- Larsen SB, Olsen A, Lynch J, Christensen J, Overvad K, Tjønneland A, et al. Socioeconomic position and lifestyle in relation to breast cancer incidence among postmenopausal women: a prospective cohort study, Denmark, 1993–2006. *Cancer Epidemiol*. 2011;35(5):438–41.
- Villeneuve S, Févotte J, Anger A, Truong T, Lamkarkach F, Gaye O, et al. Breast cancer risk by occupation and industry: analysis of the CECILE study, a population-based case-control study in France. *Am J Ind Med*. 2011;54(7):499–509.
- Menvielle G, Kunst AE, van Gils CH, Peeters PH, Boshuizen H, Overvad K, et al. The contribution of risk factors to the higher incidence of invasive and in situ breast cancers in women with higher levels of education in the European prospective investigation into cancer and nutrition. *Am J Epidemiol*. 2011;173(1):26–37.
- Braaten T, Weiderpass E, Kumle M, Lund E. Explaining the socioeconomic variation in cancer risk in the Norwegian women and Cancer study. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2005;14(11 Pt 1):2591–7.
- Danø H, Hansen KD, Jensen P, Petersen JH, Jacobsen R, Ewertz M, et al. Fertility pattern does not explain social gradient in breast cancer in Denmark. *Int J Cancer*. 2004;111(3):451–6.
- Heck KE, Pamuk ER. Explaining the relation between education and postmenopausal breast cancer. *Am J Epidemiol*. 1997;145(4):366–72.
- Beiki O, Hall P, Ekbohm A, Moradi T. Breast cancer incidence and case fatality among 4.7 million women in relation to social and ethnic background: a population-based cohort study. *Breast Cancer Res BCR*. 2012;14(1):R5.
- Carlsen K, Høybye MT, Dalton SO, Tjønneland A. Social inequality and incidence of and survival from breast cancer in a population-based study in Denmark, 1994–2003. *Eur J Cancer Oxf Engl*. 1990;24(14):1996–2002.
- Meijer M, Bloomfield K, Engholm G. Neighbourhoods matter too: the association between neighbourhood socioeconomic position, population density and breast, prostate and lung cancer incidence in Denmark between 2004 and 2008. *J Epidemiol Community Health*. 2013;67(1):6–13.
- van der Linden BWA, Courvoisier DS, Cheval B, Sieber S, Bracke P, Guessous I, et al. Effect of childhood socioeconomic conditions on cancer onset in later life: an ambidirectional cohort study. *Int J Public Health*. 2018;63(7):799–810.
- Clavel-Chapelon F. Cohort profile: the French E3N cohort study. *Int J Epidemiol*. 2015;44(3):801–9.
- Palli D, Berrino F, Vineis P, Tumino R, Panico S, Masala G, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori*. 2003;89(6):586–93.
- d'Errico A, Ricceri F, Stringhini S, Carmeli C, Kivimäki M, Bartley M, et al. Socioeconomic indicators in epidemiological research: a practical example from the LIFEPAth study. *PLoS One*. 2017;12(5):e0178071.
- Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European prospective investigation into Cancer and nutrition (EPIC) study. *Public Health Nutr*. 2003;6(4):407–13.
- Agnoli C, Krogh V, Grioni S, Sieri S, Palli D, Masala G, et al. A priori-defined dietary patterns are associated with reduced risk of stroke in a large Italian cohort. *J Nutr*. 2011;141(8):1552–8.
- Buuren S van, Groothuis-Oudshoorn CGM. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* [Internet]. 2011 [cited 2019 Dec 18];45(3). Available from: <https://research.utwente.nl/en/publications/mice-multivariate-imputation-by-chained-equations-in-r>.

26. Colditz GA, Bohlke K, Berkey CS. Breast cancer risk accumulation starts early – prevention must also. *Breast Cancer Res Treat.* 2014;145(3):567–79.
27. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the nurses' health study. *Am J Epidemiol.* 2000;152(10):950–64.
28. Russo J, Tay LK, Russo IH. Differentiation of the mammary gland and susceptibility to carcinogenesis. *Breast Cancer Res Treat.* 1982;2(1):5–73.
29. Willems B, Bracke P. The education gradient in cancer screening participation: a consistent phenomenon across Europe? *Int J Public Health.* 2018;63(1):93–103.
30. Rafique M, Zia S, Ubaidullah, Sultan MA. Impact of socioeconomic status on birth weight and length of Newborns delivered at Services Hospital Lahore. *Pak Paed J.* 2008;32(2):94–100.
31. Khalid H, Khawar K, Fawad M, Farhat M, Imran M, Shah Nawaz M, et al. Age of Menarche in Relation to Socioeconomic Status, BMI, Physical Activity and Stress Among High School Girls.: 7.
32. Schoenaker DA, Jackson CA, Rowlands JV, Mishra GD. Socioeconomic position, lifestyle factors and age at natural menopause: a systematic review and meta-analysis of studies across six continents. *Int J Epidemiol.* 2014; 43(5):1542–62.
33. Samuels L, Harkness E, Astley SM, Maxwell A, Sergeant J, Morris J, et al. The Relationship of Volumetric Breast Density to Socio-Economic Status in a Screening Population. In: Fujita H, Hara T, Muramatsu C, editors. *Breast Imaging.* Cham: Springer International Publishing; 2014. p. 273–281. (Lecture Notes in Computer Science).
34. Aitken Z, Walker K, Stegeman BH, Wark PA, Moss SM, McCormack VA, et al. Mammographic density and markers of socioeconomic status: a cross-sectional study. *BMC Cancer.* 2010;10:35.
35. Lennartsson A-K, Theorell T, Rockwood AL, Kushnir MM, Jonsdottir IH. Perceived Stress at Work Is Associated with Lower Levels of DHEA-S. *PLoS ONE* [Internet]. 2013 Aug 28 [cited 2020 Jan 14];8(8). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3756071/>.
36. Cho S, Park W-J, Kang W, Lim H-M, Ahn J-S, Lim D-Y, et al. The association between serum dehydroepiandrosterone sulfate (DHEAS) levels and job-related stress among female nurses. *Ann Occup Environ Med* [Internet]. 2019 15 [cited 2020 Jan 14];31(1). Available from: <https://doi.org/10.35371/aoem.2019.31.e18>.
37. Kaaks R, Rinaldi S, Key TJ, Berrino F, Peeters PHM, Biessy C, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer.* 2005;12(4):1071–82.
38. Pudrovskaya T. Job authority and breast Cancer. *Soc Forces Sci Medium Soc Study Interpret.* 2013;92(1):1–24.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



ANNEXE 6 - DOCUMENTS SUPPLEMENTAIRES RELATIFS AU TRAVAIL 3

6.1 - Caractéristiques des femmes avec des données disponibles dans E3N selon la PSE.

Variables	Métier du père					Education					Profession				
	NA	Avantagée	Intermédiaire	Moins avantagée	p	NA	Elevé	Moyen	Faible	p	NA	Avantagée	Intermédiaire	Moins avantagée	p
Age*	0	n= 10369	n= 26111	n= 25243	<0.001	0	n= 29009	n= 40167	n= 11174	<0.001	0	n= 12212	n= 45011	n= 12203	<0.001
< 47.9 ans, n(%)		5327 (51.37)	13451 (51.51)	13798 (54.66)			17728 (61.11)	19712 (49.08)	4444 (39.77)			5107 (41.82)	25129 (55.83)	6467 (53)	
≥ 47.9 ans, n(%)		5042 (48.63)	12660 (48.49)	11445 (45.34)			11281 (38.89)	20455 (50.92)	6730 (60.23)			7105 (58.18)	19882 (44.17)	5736 (47)	
Consommation d'alcool	3294	n= 9783	n= 24807	n= 23839	<0.001	4419	n= 27578	n= 38193	n= 10160	<0.001	3966	n= 11562	n= 42651	n= 11247	<0.001
Aucune, n(%)		998 (10.2)	2673 (10.78)	2524 (10.59)			2426 (8.8)	3980 (10.42)	1040 (10.24)			1205 (10.42)	4537 (10.64)	1169 (10.39)	
Modérée, n(%)		5699 (58.25)	15003 (60.48)	14366 (60.26)			15150 (54.94)	21431 (56.11)	4838 (47.62)			6891 (59.6)	25597 (60.02)	6349 (56.45)	
Elevée, n(%)		1939 (19.82)	4142 (16.7)	3917 (16.43)			5088 (18.45)	5640 (14.77)	1193 (11.74)			2068 (17.89)	7346 (17.22)	1736 (15.44)	
Non répondante à Q3, n(%)		1147 (11.72)	2989 (12.05)	3032 (12.72)			4914 (17.82)	7142 (18.7)	3089 (30.4)			1398 (12.09)	5171 (12.12)	1993 (17.72)	
Tabagisme	103	n= 10340	n= 26073	n= 25207	<0.001	189	n= 28935	n= 40082	n= 11144	<0.001	121	n= 12193	n= 44943	n= 12169	<0.001
Jamais, n(%)		6337 (61.29)	17569 (67.38)	17378 (68.94)			17233 (59.56)	27911 (69.63)	8588 (77.06)			8259 (67.74)	29547 (65.74)	8360 (68.7)	
Ancienne, n(%)		2437 (23.57)	5391 (20.68)	4957 (19.67)			7134 (24.66)	7573 (18.89)	1482 (13.3)			2466 (20.22)	9722 (21.63)	2183 (17.94)	
Actuelle, n(%)		1566 (15.15)	3113 (11.94)	2872 (11.39)			4568 (15.79)	4598 (11.47)	1074 (9.64)			1468 (12.04)	5674 (12.62)	1626 (13.36)	

Régime occidental	3293	n= 9783	n= 24807	n= 23840	<0.001	4418	n=	n=	n=	<0.001	3965	n= 11562	n= 42652	n= 11247	<0.001
							27578	38194	10160						
[-2.93,-0.511], n(%)		2837 (29)	7266 (29.29)	6611 (27.73)			7315 (26.52)	10291 (26.94)	2412 (23.74)			3666 (31.71)	12007 (28.15)	2976 (26.46)	
(-0.511,0.307], n(%)		2925 (29.9)	7431 (29.96)	6849 (28.73)			7747 (28.09)	10280 (26.92)	2226 (21.91)			3386 (29.29)	12585 (29.51)	3040 (27.03)	
(0.307,7.78], n(%)		2874 (29.38)	7121 (28.71)	7347 (30.82)			7602 (27.57)	10480 (27.44)	2433 (23.95)			3112 (26.92)	12888 (30.22)	3238 (28.79)	
Non répondante à Q3,, n(%)		1147 (11.72)	2989 (12.05)	3033 (12.72)			4914 (17.82)	7143 (18.7)	3089 (30.4)			1398 (12.09)	5172 (12.13)	1993 (17.72)	
Activité physique	678	n= 10251	n= 25811	n= 24983	<0.001	893	n=	n=	n=	<0.001	781	n= 12062	n= 44513	n= 12070	<0.001
							28699	39731	11027						
(7.57,35.9], n(%)		3158 (30.81)	8572 (33.21)	8660 (34.66)			8373 (29.18)	13276 (33.41)	4869 (44.16)			3823 (31.69)	14210 (31.92)	4273 (35.4)	
(4.66,7.57], n(%)		3336 (32.54)	8665 (33.57)	8423 (33.71)			9561 (33.31)	13528 (34.05)	3220 (29.2)			3955 (32.79)	15248 (34.26)	3843 (31.84)	
[0,4.66], n(%)		3757 (36.65)	8574 (33.22)	7900 (31.62)			10765 (37.51)	12927 (32.54)	2938 (26.64)			4284 (35.52)	15055 (33.82)	3954 (32.76)	
Taille	815	n= 10243	n= 25738	n= 24927	<0.001	1079	n=	n=	n=	<0.001	963	n= 12043	n= 44439	n= 11981	<0.001
							28671	39670	10930						
[135,160], n(%)		4153 (40.54)	11663 (45.31)	11625 (46.64)			11691 (40.78)	18666 (47.05)	5726 (52.39)			5361 (44.52)	19876 (44.73)	5816 (48.54)	
(160,164], n(%)		2382 (23.25)	5805 (22.55)	5570 (22.35)			6569 (22.91)	8903 (22.44)	2201 (20.14)			2674 (22.2)	10074 (22.67)	2490 (20.78)	
(164,190], n(%)		3708 (36.2)	8270 (32.13)	7732 (31.02)			10411 (36.31)	12101 (30.5)	3003 (27.47)			4008 (33.28)	14489 (32.6)	3675 (30.67)	
Poids	1150	n= 10181	n= 25603	n= 24789	0,013	1563	n=	n=	n=	<0.001	1352	n= 11980	n= 44187	n= 11907	<0.001
							28445	39477	10865						
[29,55], n(%)		3422 (33.61)	8917 (34.83)	8328 (33.6)			10392 (36.53)	13350 (33.82)	3024 (27.83)			3679 (30.71)	15682 (35.49)	3917 (32.9)	
(55,62], n(%)		3485 (34.23)	8669 (33.86)	8402 (33.89)			9751 (34.28)	13237 (33.53)	3447 (31.73)			4069 (33.96)	14979 (33.9)	3888 (32.65)	
(62,163], n(%)		3274 (32.16)	8017 (31.31)	8059 (32.51)			8302 (29.19)	12890 (32.65)	4394 (40.44)			4232 (35.33)	13526 (30.61)	4102 (34.45)	

Antécédents familiaux d'un cancer de l'ovaire (1grade)	0	n= 10369	n= 26111	n= 25243	0,095	0	n= 29009	n= 40167	n= 11174	0,044	0	n= 12212	n= 45011	n= 12203	0,255
Non, n(%)		10228 (98.64)	25793 (98.78)	24967 (98.91)			28648 (98.76)	39707 (98.85)	11068 (99.05)			12054 (98.71)	44475 (98.81)	12073 (98.93)	
Oui, n(%)		141 (1.36)	318 (1.22)	276 (1.09)			361 (1.24)	460 (1.15)	106 (0.95)			158 (1.29)	536 (1.19)	130 (1.07)	
Antécédents familiaux d'un cancer du sein (1grade)	1232	n= 10156	n= 25574	n= 24761	<0.001	1792	n= 28343	n= 39337	n= 10878	<0.001	1440	n= 11945	n= 44107	n= 11934	<0.001
Non, n(%)		4438 (43.7)	11158 (43.63)	10675 (43.11)			11611 (40.97)	16985 (43.18)	4889 (44.94)			4844 (40.55)	18997 (43.07)	5210 (43.66)	
Oui, n(%)		911 (8.97)	1977 (7.73)	1692 (6.83)			2273 (8.02)	2863 (7.28)	788 (7.24)			885 (7.41)	3280 (7.44)	897 (7.52)	
Indéfini, n(%)		4807 (47.33)	12439 (48.64)	12394 (50.05)			14459 (51.01)	19489 (49.54)	5201 (47.81)			6216 (52.04)	21830 (49.49)	5827 (48.83)	
Traitement hormonal de la ménopause	3597	n= 9727	n= 24594	n= 23805	0,005	8666	n= 26384	n= 35817	n= 9483	<0.001	4000	n= 11389	n= 42484	n= 11553	<0.001
Non, n(%)		8285 (85.18)	21182 (86.13)	20632 (86.67)			23323 (88.4)	30945 (86.4)	8274 (87.25)			9484 (83.27)	36953 (86.98)	10183 (88.14)	
Oui, n(%)		1075 (11.05)	2489 (10.12)	2310 (9.7)			2273 (8.62)	3556 (9.93)	801 (8.45)			1415 (12.42)	4009 (9.44)	947 (8.2)	
Indéfini, n(%)		367 (3.77)	923 (3.75)	863 (3.63)			788 (2.99)	1316 (3.67)	408 (4.3)			490 (4.3)	1522 (3.58)	423 (3.66)	
Nombre de grossesses menées à terme	3	n= 10368	n= 26110	n= 25242	<0.001	7	n= 29007	n= 40165	n= 11171	<0.001	4	n= 12212	n= 45008	n= 12202	<0.001
3+, n(%)		3566 (34.39)	7520 (28.8)	6900 (27.34)			8039 (27.71)	11283 (28.09)	3983 (35.65)			3754 (30.74)	12230 (27.17)	3374 (27.65)	
1-2, n(%)		5427 (52.34)	15748 (60.31)	15788 (62.55)			16181 (55.78)	25048 (62.36)	6322 (56.59)			6863 (56.2)	27566 (61.25)	7287 (59.72)	

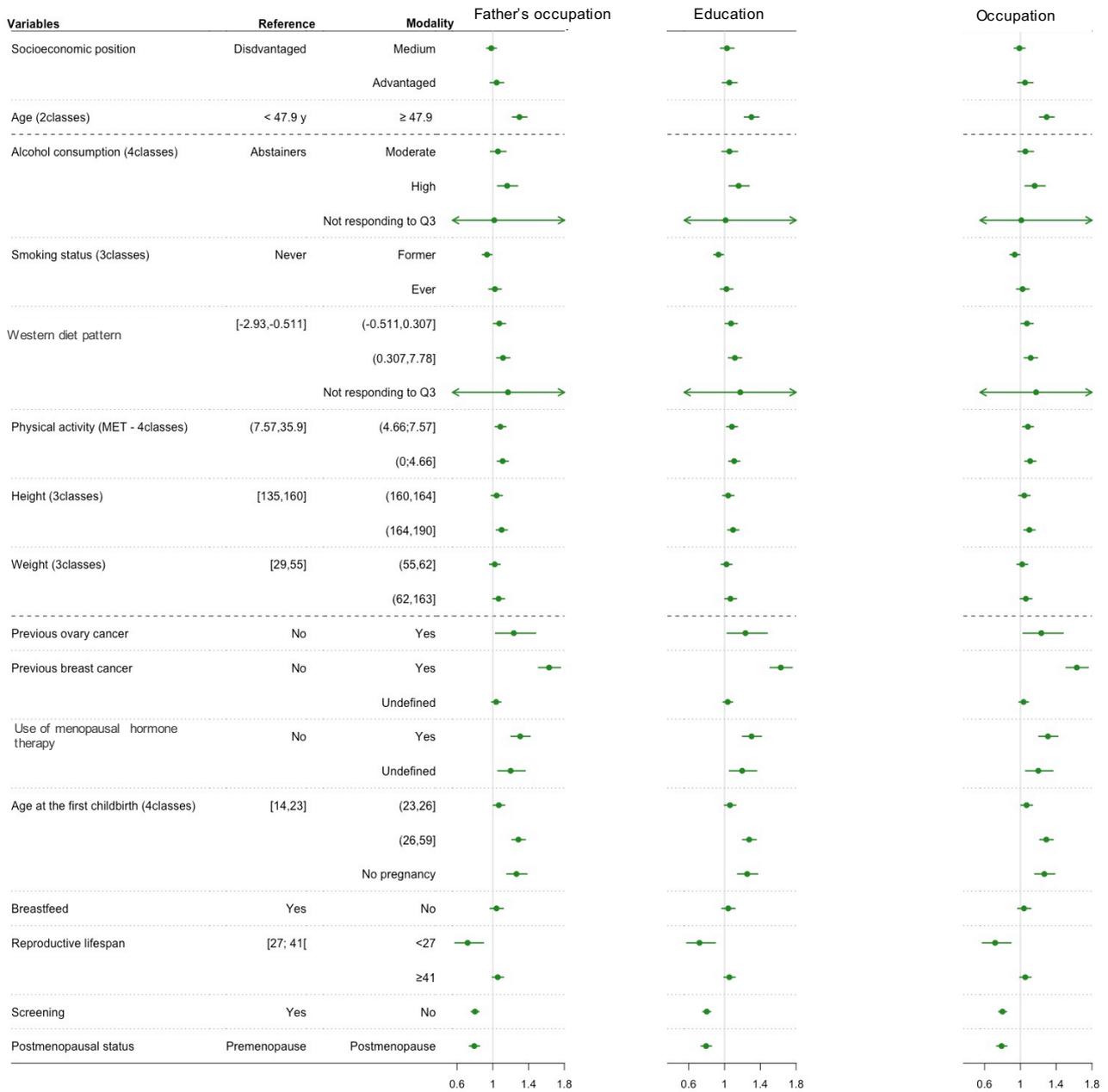
	0, n(%)	1375 (13.26)	2842 (10.88)	2554 (10.12)		4787 (16.5)	3834 (9.55)	866 (7.75)		1595 (13.06)	5212 (11.58)	1541 (12.63)			
Age à la première grossesse	161	n= 10349	n= 26035	n= 25178	<0.001	2636	n=	n=	n=	<0.001	234	n= 12177	n= 44877	n= 12138	<0.001
	[14,23], n(%)	3111 (30.06)	9031 (34.69)	9933 (39.45)		6088 (21.58)	16078 (41.3)	5625 (53.2)		4360 (35.81)	15066 (33.57)	5016 (41.32)			
	(23,26], n(%)	2911 (28.13)	7209 (27.69)	6661 (26.46)		7604 (26.95)	10787 (27.71)	2286 (21.62)		3247 (26.67)	12649 (28.19)	2799 (23.06)			
	(26,59], n(%)	2952 (28.52)	6953 (26.71)	6030 (23.95)		9736 (34.51)	8227 (21.13)	1796 (16.99)		2975 (24.43)	11950 (26.63)	2782 (22.92)			
	Pas de grossesse n(%)	1375 (13.29)	2842 (10.92)	2554 (10.14)		4787 (16.97)	3834 (9.85)	866 (8.19)		1595 (13.1)	5212 (11.61)	1541 (12.7)			
Allaitement	375	n= 10310	n= 25951	n= 25087	0,002	8172	n=	n=	n=	<0.001	545	n= 12111	n= 44693	n= 12077	<0.001
	Oui, n(%)	6456 (62.62)	15740 (60.65)	15340 (61.15)		26455 (60.14)	36235 (59.19)	9488 (57.16)		7422 (61.28)	27053 (60.53)	6759 (55.97)			
	Non, n(%)	3854 (37.38)	10211 (39.35)	9747 (38.85)		10545 (39.86)	14787 (40.81)	4065 (42.84)		4689 (38.72)	17640 (39.47)	5318 (44.03)			
Durée de vie reproductive	3821	n= 9742	n= 24463	n= 23697	<0.001	5916	n=	n=	n=	<0.001	4392	n= 11529	n= 42201	n= 11304	<0.001
	[27; 41[, n(%)	7743 (79.48)	19584 (80.06)	19169 (80.89)		26608 (80.05)	37543 (80.19)	10283 (81.95)		9102 (78.95)	33933 (80.41)	9226 (81.62)			
	<27, n(%)	124 (1.27)	342 (1.4)	385 (1.62)		284 (1.07)	571 (1.52)	337 (3.28)		196 (1.7)	556 (1.32)	264 (2.34)			
	≥41, n(%)	1875 (19.25)	4537 (18.55)	4143 (17.48)		5025 (18.89)	6868 (18.29)	1519 (14.77)		2231 (19.35)	7712 (18.27)	1814 (16.05)			
Dépistage du cancer du sein	0	n= 10369	n= 26111	n= 25243	<0.001	0	n=	n=	n=	<0.001	0	n= 12212	n= 45011	n= 12203	0,058
	Oui, n(%)	7534 (72.66)	18626 (71.33)	17772 (70.4)		29009 (71.84)	40167 (70.71)	11174 (67.05)		8739 (71.56)	32016 (71.13)	8570 (70.23)			
	Non, n(%)	2835 (27.34)	7485 (28.67)	7471 (29.6)		8169 (28.16)	11765 (29.29)	3682 (32.95)		3473 (28.44)	12995 (28.87)	3633 (29.77)			
Statut ménopausique	2666	n= 9940	n= 24955	n= 24162	<0.001	4240	n=	n=	n=	<0.001	3037	n= 11776	n= 43017	n= 11596	<0.001
						27171	38307	10632							

Pré-menopause, n(%)	5874 (59.09)	14713 (58.96)	14702 (60.85)	18208 (67.01)	21780 (56.86)	4680 (44.02)	5933 (50.38)	27028 (62.83)	6752 (58.23)
Post-menopause, n(%)	4066 (40.91)	10242 (41.04)	9460 (39.15)	8963 (32.99)	16527 (43.14)	5952 (55.98)	5843 (49.62)	15989 (37.17)	4844 (41.77)

P-values are estimated with log rank tests.

Defined according to the median.

6.2 - Représentation graphique de l'association entre les trois PSE et chaque covariable utilisée dans le modèle entièrement ajusté dans E3N (N = 83 436).



6.3 - Caractéristiques de la population des cas complets et comparaison avec la population exclue des analyses

Variable	Manquants	Population avec manquants	Population cas-complets	p
Métier du père	21713	n= 22601	n= 39122	<0,001
Avantagé, n(%)		4018 (17.78)	6351 (16.23)	
Intermédiaire, n(%)		9610 (42.52)	16501 (42.18)	
Moins avantagé, n(%)		8973 (39.7)	16270 (41.59)	
Education	3086	n= 41228	n= 39122	<0,001
Elevé, n(%)		14370 (34.85)	14639 (37.42)	
Moyen, n(%)		19010 (46.11)	21157 (54.08)	
Faible, n(%)		7848 (19.04)	3326 (8.5)	
Profession	14010	n= 30304	n= 39122	<0,001
Avantagée, n(%)		5413 (17.86)	6799 (17.38)	
Intermédiaire, n(%)		18822 (62.11)	26189 (66.94)	
Moins avantagée, n(%)		6069 (20.03)	6134 (15.68)	
Diagnostic cancer du sein	0	n= 44314	n= 39122	<0,001
Non, n(%)		40310 (90.96)	35249 (90.1)	
Oui, n(%)		4004 (9.04)	3873 (9.9)	
Temps de suivi	0	n= 44314	n= 39122	
Médiane[IQR]		24.4[16.0 - 24.5]	24.4[21.5-24.5]	<0,001
Age	0	n= 44314	n= 39122	<0,001
< 47.9, n(%)		21627 (48.8)	21701 (55.47)	
≥ 47.9, n(%)		22687 (51.2)	17421 (44.53)	

6.4 - Association entre la PSE au cours de la vie et le risque de cancer du sein à partir de la population de cas complets dans E3N [N = 39 122].

	E3N					
	A. Métier du père ^a		B. Education ^b		C. Profession ^a	
	Intermédiaire	Avantagée	Moyen	Elevé	Intermédiaire	Avantagée
	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
M1	1.03 [0.96; 1.10]	1.05 [0.96; 1.15]	1.19 [1.05; 1.35]	1.30 [1.14; 1.48]	1.04 [0.95; 1.14]	1.13 [1.01; 1.26]
M1 + tous les comportements de santé et caractéristiques anthropométriques	1.03 [0.96; 1.10]	1.04 [0.95; 1.14]	1.17 [1.03; 1.33]	1.26 [1.11; 1.44]	1.04 [0.95; 1.14]	1.12 [1.00; 1.25]
M1 + Age à la première grossesse	1.02 [0.95; 1.10]	1.03 [0.94; 1.13]	1.18 [1.04; 1.34]	1.22 [1.07; 1.40]	1.04 [0.95; 1.14]	1.13 [1.01; 1.26]
M1 + tous les facteurs reproductifs	1.01 [0.95; 1.09]	1.01 [0.92; 1.11]	1.15 [1.01; 1.31]	1.17 [1.03; 1.34]	1.03 [0.94; 1.13]	1.13 [1.01; 1.26]
M2	1.02 [0.95; 1.09]	1.01 [0.92; 1.10]	1.15 [1.01; 1.30]	1.16 [1.01; 1.32]	1.03 [0.94; 1.13]	1.13 [1.01; 1.26]

M1 est ajusté sur l'âge

^aRéférence: "Moins avantagée"

^bRéférence: "Faible niveau d'éducation"

M2 est ajusté sur l'ensemble des covariables

6.5 - Association entre la PSE au cours de la vie et le risque de cancer du sein invasif à partir des données imputées dans E3N (N = 82 458).

	E3N					
	Cancers du sein invasifs [87.6 % de tous les cas]					
	A. Métier du père ^a		B. Education ^b		C. Profession ^a	
	Intermédiaire	Avantagée	Moyen	Elevé	Intermédiaire	Avantagée
	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
M1	1.03 (0.97; 1.09)	1.11 (1.02; 1.20)	1.09 (1.01; 1.17)	1.20 (1.11; 1.30)	1.02 (0.95; 1.08)	1.09 (1.00; 1.18)
M1 + tous les comportements de santé et caractéristiques anthropométriques	1.02 (0.97; 1.09)	1.09 (1.01; 1.18)	1.08 (1.00; 1.17)	1.18 (1.09; 1.28)	1.01 (0.95; 1.08)	1.08 (0.99; 1.17)
M1 + tous les facteurs reproductifs	1.01 (0.95; 1.07)	1.06 (0.98; 1.15)	1.03 (0.95; 1.11)	1.07 (0.98; 1.16)	0.98 (0.92; 1.05)	1.07 (0.98; 1.16)
M2	1.01 (0.95; 1.07)	1.05 (0.97; 1.14)	1.03 (0.95; 1.11)	1.06 (0.98; 1.15)	0.98 (0.92; 1.05)	1.06 (0.98; 1.15)

M1 est ajusté sur l'âge.

^aRéférence: "Moins avantagée".

^bRéférence: "Faible niveau d'éducation".

M2 est ajusté sur l'ensemble des covariables.

6.6 - Caractéristiques des femmes avec des données disponibles dans EPIC-Italie selon le statut vis-à-vis du cancer du sein et par PSE.

Variables	BC		p	Métier du père					Education					Profession la plus élevée du ménage					
	No	Yes		NA	Avant- agé	Inter- médiaire	Moins avantagé	p	NA	Elevé	Moyen	Faible	p	NA	Avant- agée	Inter- médiaire	Moins avantagée	p	
Métier du père (3cl)	n= 14804	n= 723																	
Avantagé, n(%)	797 (5.38)	41 (5.67)	0.942			-				-					-				
Intermédiaire, n(%)	4229 (28.57)	207 (28.63)																	
Moins avantagé, n(%)	9778 (66.05)	475 (65.7)																	
Education (3cl)	n= 19297	n= 881																	
Elevé, n(%)	2256 (11.69)	106 (12.03)	0.567			-				-					-				
Moyen, n(%)	6467 (33.51)	280 (31.78)																	
Faible, n(%)	10574 (54.8)	495 (56.19)																	
Profession la plus élevée du ménage (3cl)	n= 19401	n= 884																	
Avantagée, n(%)	2518 (12.98)	120 (13.57)	0.156			-				-					-				
Intermédiaire, n(%)	4971 (25.62)	201 (22.74)																	
Moins avantagée, n(%)	11912 (61.4)	563 (63.69)																	
Age	n= 19637	n= 893		0	n= 838	n= 4436	n= 10253	<0.001	0	n= 2362	n= 6747	n= 11069	<0.001	0	n= 2638	n= 5172	n= 12475	<0.001	
<50, n(%)	10478 (53.36)	403 (45.13)	<0.001		436 (52.03)	2120 (47.79)	5760 (56.18)			1552 (65.71)	4300 (63.73)	4835 (43.68)			1485 (56.29)	2649 (51.22)	6632 (53.16)		
>50, n(%)	9159 (46.64)	490 (54.87)			402 (47.97)	2316 (52.21)	4493 (43.82)			810 (34.29)	2447 (36.27)	6234 (56.32)			1153 (43.71)	2523 (48.78)	5843 (46.84)		
Centre	n= 19637	n= 893		0	n= 838	n= 4436	n= 10253	<0.001	0	n= 2362	n= 6747	n= 11069	<0.001	0	n= 2638	n= 5172	n= 12475	<0.001	
Sud, n(%)	7611 (38.76)	260 (29.12)	<0.001		123 (14.68)	1329 (29.96)	1612 (15.72)			1315 (55.67)	2720 (40.31)	3752 (33.9)			831 (31.5)	3847 (74.38)	3066 (24.58)		
Nord, n(%)	12026 (61.24)	633 (70.88)			715 (85.32)	3107 (70.04)	8641 (84.28)			1047 (44.33)	4027 (59.69)	7317 (66.1)			1807 (68.5)	1325 (25.62)	9409 (75.42)		

Consommation d'alcool	n= 19310	n= 881		315	n= 822	n= 4326	n= 10064	<0.001	17	n= 2358	n= 6742	n= 11061	<0.001	318	n= 2596	n= 5123	n= 12248	<0.001
Aucune, n(%)	3509 (18.17)	170 (19.3)	0.450	125 (15.21)	894 (20.67)	1829 (18.17)			363 (15.39)	1087 (16.12)	2219 (20.06)			413 (15.91)	906 (17.68)	2319 (18.93)		
Modérée, n(%)	12890 (66.75)	570 (64.7)		554 (67.4)	2863 (66.18)	6700 (66.57)			1654 (70.14)	4643 (68.87)	7146 (64.61)			1731 (66.68)	3437 (67.09)	8144 (66.49)		
Elevée, n(%)	2911 (15.08)	141 (16)		143 (17.4)	569 (13.15)	1535 (15.25)			341 (14.46)	1012 (15.01)	1696 (15.33)			452 (17.41)	780 (15.23)	1785 (14.57)		
Tabagisme	n= 19319	n= 881		313	n= 823	n= 4329	n= 10062	<0.001	0	n= 2362	n= 6747	n= 11069	<0.001	312	n= 2598	n= 5128	n= 12247	<0.001
Jamais n(%)	10820 (56.01)	518 (58.8)	0.253	417 (50.67)	2786 (64.36)	6015 (59.78)			1117 (47.29)	3224 (47.78)	6983 (63.09)			1333 (51.31)	2420 (47.19)	7465 (60.95)		
Ancienne, n(%)	3561 (18.43)	155 (17.59)		197 (23.94)	729 (16.84)	1876 (18.64)			577 (24.43)	1534 (22.74)	1600 (14.45)			548 (21.09)	965 (18.82)	2159 (17.63)		
Actuelle, n(%)	4938 (25.56)	208 (23.61)		209 (25.39)	814 (18.8)	2171 (21.58)			668 (28.28)	1989 (29.48)	2486 (22.46)			717 (27.6)	1743 (33.99)	2623 (21.42)		
Activité physique	n= 19637	n= 893		0	n= 838	n= 4436	n= 10253	<0.001	0	n= 2362	n= 6747	n= 11069	<0.001	0	n= 2638	n= 5172	n= 12475	<0.001
Elevée, n(%)	4455 (22.69)	222 (24.86)	0.295	185 (22.08)	1512 (34.08)	2829 (27.59)			193 (8.17)	1109 (16.44)	3370 (30.45)			525 (19.9)	505 (9.76)	3580 (28.7)		
Modérée, n(%)	7711 (39.27)	346 (38.75)		341 (40.69)	1772 (39.95)	4277 (41.71)			730 (30.91)	2468 (36.58)	4848 (43.8)			1078 (40.86)	1788 (34.57)	5086 (40.77)		
Faible/inactive, n(%)	7471 (38.05)	325 (36.39)		312 (37.23)	1152 (25.97)	3147 (30.69)			1439 (60.92)	3170 (46.98)	2851 (25.76)			1035 (39.23)	2879 (55.67)	3809 (30.53)		
Régime méditerranéen	n= 19310	n= 881		315	n= 822	n= 4326	n= 10064	0,427	17	n= 2358	n= 6742	n= 11061	<0.001	318	n= 2596	n= 5123	n= 12248	<0.001
Adhérence élevée, n(%)	4419 (22.88)	199 (22.59)	0.090	146 (17.76)	760 (17.57)	1722 (17.11)			615 (26.08)	1591 (23.6)	2405 (21.74)			595 (22.92)	1671 (32.62)	2293 (18.72)		
Adhérence intermédiaire, n(%)	7535 (39.02)	316 (35.87)		317 (38.56)	1644 (38)	3718 (36.94)			943 (39.99)	2556 (37.91)	4341 (39.25)			957 (36.86)	2122 (41.42)	4696 (38.34)		
Faible adhérence, n(%)	7356 (38.09)	366 (41.54)		359 (43.67)	1922 (44.43)	4624 (45.95)			800 (33.93)	2595 (38.49)	4315 (39.01)			1044 (40.22)	1330 (25.96)	5259 (42.94)		
Poids	n= 19491	n= 888		132	n= 830	n= 4389	n= 10176	<0.001	95	n= 2352	n= 6717	n= 11014	<0.001	139	n= 2617	n= 5150	n= 12379	<0.001
[37,59], n(%)	6747 (34.62)	266 (29.95)	0.005	368 (44.34)	1498 (34.13)	3792 (37.26)			1007 (42.81)	2662 (39.63)	3263 (29.63)			1014 (38.75)	1584 (30.76)	4343 (35.08)		
(59,67.5), n(%)	6285 (32.25)	288 (32.43)		254 (30.6)	1442 (32.85)	3191 (31.36)			732 (31.12)	2137 (31.81)	3609 (32.77)			835 (31.91)	1764 (34.25)	3904 (31.54)		
(67.5,156], n(%)	6459 (33.14)	334 (37.61)		208 (25.06)	1449 (33.01)	3193 (31.38)			613 (26.06)	1918 (28.55)	4142 (37.61)			768 (29.35)	1802 (34.99)	4132 (33.38)		
Taille	n= 19495	n= 888	0.210	137	n= 830	n= 4384	n= 10176	<0.001	91	n= 2352	n= 6724	n= 11011	<0.001	134	n= 2619	n= 5154	n= 12378	<0.001

[1.3,1.56], n(%)	7804 (40.03)	332 (37.39)		206 (24.82)	1847 (42.13)	3768 (37.03)		683 (29.04)	2270 (33.76)	5069 (46.04)		785 (29.97)	2142 (41.56)	5097 (41.18)				
(1.56,1.61], n(%)	6214 (31.87)	305 (34.35)		273 (32.89)	1389 (31.68)	3304 (32.47)		765 (32.53)	2204 (32.78)	3453 (31.36)		854 (32.61)	1649 (31.99)	3952 (31.93)				
(1.61,1.84], n(%)	5477 (28.09)	251 (28.27)		351 (42.29)	1148 (26.19)	3104 (30.5)		904 (38.44)	2250 (33.46)	2489 (22.6)		980 (37.42)	1363 (26.45)	3329 (26.89)				
Statut ménopausique	n= 17017	n= 798		508	n= 815	n= 4256	n= 9948	<0.001	238 4	n= 2064	n= 5993	n= 9737	<0.001	2678	n= 2377	n= 3610	n= 11620	<0.001
Pré-ménopause, n(%)	11166 (65.62)	481 (60.28)	0.002	501 (61.47)	2282 (53.62)	6175 (62.07)			1663 (80.57)	4590 (76.59)	5380 (55.25)		1651 (69.46)	2857 (79.14)	7023 (60.44)			
Post-menopause, n(%)	5851 (34.38)	317 (39.72)		314 (38.53)	1974 (46.38)	3773 (37.93)			401 (19.43)	1403 (23.41)	4357 (44.75)		726 (30.54)	753 (20.86)	4597 (39.56)			
Nombre de grossesses menées à terme	n= 19318	n= 881		314	n= 823	n= 4328	n= 10062	<0.001	1	n= 2362	n= 6747	n= 11068	<0.001	313	n= 2598	n= 5128	n= 12246	<0.001
0, n(%)	2283 (11.82)	119 (13.51)	0.151	144 (17.5)	477 (11.02)	1270 (12.62)			491 (20.79)	954 (14.14)	955 (8.63)		264 (10.16)	519 (10.12)	1543 (12.6)			
1-2, n(%)	12147 (62.88)	559 (63.45)		530 (64.4)	2752 (63.59)	6944 (69.01)			1453 (61.52)	4536 (67.23)	6701 (60.54)		1690 (65.05)	2991 (58.33)	7916 (64.64)			
3+, n(%)	4888 (25.3)	203 (23.04)		149 (18.1)	1099 (25.39)	1848 (18.37)			418 (17.7)	1257 (18.63)	3412 (30.83)		644 (24.79)	1618 (31.55)	2787 (22.76)			
Age à la première grossesse	n= 19301	n= 881		317	n= 823	n= 4325	n= 10062	<0.001	18	n= 2359	n= 6741	n= 11060	<0.001	330	n= 2596	n= 5118	n= 12241	<0.001
[14,24], n(%)	7732 (40.06)	280 (31.78)		211 (25.64)	1960 (45.32)	3863 (38.39)			294 (12.46)	2186 (32.43)	5522 (49.93)		854 (32.9)	1998 (39.04)	5075 (41.46)			
(24,27], n(%)	4689 (24.29)	236 (26.79)	<0.001	180 (21.87)	972 (22.47)	2545 (25.29)			581 (24.63)	1730 (25.66)	2608 (23.58)		711 (27.39)	1271 (24.83)	2912 (23.79)			
(27,47], n(%)	4598 (23.82)	246 (27.92)		288 (34.99)	916 (21.18)	2385 (23.7)			993 (42.09)	1871 (27.76)	1976 (17.87)		767 (29.55)	1330 (25.99)	2712 (22.16)			
Pas de grossesse, n(%)	2282 (11.82)	119 (13.51)		144 (17.5)	477 (11.03)	1269 (12.61)			491 (20.81)	954 (14.15)	954 (8.63)		264 (10.17)	519 (10.14)	1542 (12.6)			

6.7 - Résultats des modèles de Cox du risque de cancer du sein en utilisant les données de l'imputation multiple dans EPIC-Italie (N= 20530).

	A. Métier du père ^a		B. Education ^b		C. Profession la plus élevée de ménage ^a	
	Intermédiaire	Avantagé	Moyen	Elevé	Intermédiaire	Avantagée
	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
M1	1.02 [0.87; 1.19]	1.06 [0.78; 1.45]	1.03 [0.88; 1.19]	1.19 [0.96; 1.47]	1.03 [0.86; 1.23]	1.05 [0.86; 1.28]
M1 + Consommation d'alcool	1.02 [0.87; 1.19]	1.06 [0.78; 1.45]	1.03 [0.89; 1.20]	1.19 [0.96; 1.48]	1.03 [0.86; 1.23]	1.05 [0.86; 1.28]
M1 + Tabagisme	1.02 [0.87; 1.19]	1.06 [0.78; 1.45]	1.03 [0.89; 1.20]	1.19 [0.96; 1.48]	1.03 [0.86; 1.23]	1.05 [0.86; 1.28]
HB-A* M1 + Régime méditerranéen	1.02 [0.87; 1.19]	1.07 [0.78; 1.45]	1.03 [0.88; 1.19]	1.19 [0.96; 1.48]	1.03 [0.86; 1.24]	1.05 [0.86; 1.28]
M1 + Activité physique	1.02 [0.87; 1.20]	1.05 [0.77; 1.44]	1.01 [0.87; 1.18]	1.16 [0.93; 1.45]	1.01 [0.85; 1.22]	1.04 [0.86; 1.27]
M1 + Taille	1.02 [0.87; 1.19]	1.05 [0.77; 1.43]	1.02 [0.88; 1.18]	1.17 [0.95; 1.46]	1.02 [0.86; 1.22]	1.04 [0.85; 1.27]
M1 + Poids	1.01 [0.87; 1.19]	1.08 [0.79; 1.48]	1.05 [0.90; 1.22]	1.23 [0.99; 1.52]	1.03 [0.86; 1.24]	1.06 [0.87; 1.29]
M1 + HB-A	1.02 [0.87; 1.20]	1.08 [0.79; 1.47]	1.04 [0.89; 1.21]	1.21 [0.97; 1.52]	1.02 [0.85; 1.23]	1.05 [0.86; 1.29]
RF* M1 + Age à la première grossesse	1.03 [0.88; 1.21]	1.02 [0.75; 1.39]	0.96 [0.83; 1.12]	1.05 [0.84; 1.31]	1.01 [0.85; 1.21]	1.02 [0.84; 1.25]
M1 + Statut ménopausique	1.02 [0.87; 1.20]	1.06 [0.78; 1.44]	1.02 [0.88; 1.18]	1.18 [0.95; 1.46]	1.02 [0.85; 1.22]	1.04 [0.86; 1.27]
Model 1 + FR	1.04 [0.88; 1.21]	1.02 [0.75; 1.39]	0.95 [0.82; 1.11]	1.04 [0.83; 1.30]	1.00 [0.84; 1.20]	1.02 [0.83; 1.24]
M2	1.04 [0.89; 1.22]	1.04 [0.76; 1.42]	0.97 [0.83; 1.14]	1.06 [0.85; 1.34]	1.01 [0.84; 1.22]	1.03 [0.84; 1.26]

M1 est ajusté sur l'âge et le centre

^aRéférence: "Moins avantagée"

^bRéférence: "Faible niveau d'éducation"

* HB-A correspond à l'ajustement sur les comportements de santé et caractéristiques anthropométriques

RF correspond à l'ajustement sur les facteurs reproductifs

M2 est ajusté sur l'ensemble des variables

Les Hazard Ratio et l'intervalle de confiance sont rapportés pour (A) la profession du père, (B) l'éducation et (C) la profession la plus élevée du ménage..

6.8 - Résultats des régressions après ajustement chronologique des PSE dans EPIC-Italie à partir des données imputées [n = 20530].

			EPIC-Italie			
			Modèle A	Modèle B	Modèle C	Modèle D
Covariables	Référence	Modalité	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Métier du père	Moins avantageé	Intermédiaire	1.01 [0.87; 1.19]	1.02 [0.87; 1.20]	1.02 [0.87; 1.20]	1.04 [0.88; 1.22]
		Avantageé	1.06 [0.78; 1.45]	1.01 [0.73; 1.39]	1.00 [0.72; 1.39]	1.02 [0.74; 1.41]
Education	Faible	Moyen	-	1.03 [0.88; 1.19]	1.02 [0.88; 1.19]	0.97 [0.83; 1.14]
		Elevé	-	1.19 [0.95; 1.49]	1.18 [0.94; 1.49]	1.06 [0.83; 1.35]
Profession la plus élevée du ménage	Moins avantageée	Intermédiaire	-	-	1.02 [0.85; 1.22]	1.01 [0.84; 1.21]
		Avantageée	-	-	1.02 [0.83; 1.25]	1.02 [0.83; 1.26]

Le modèle A est ajusté sur l'âge, le centre et la profession du père.

Le modèle B est ajusté sur l'âge, le centre, la profession du père et l'éducation.

Le modèle C est ajusté sur l'âge, le centre et les trois PSE.

Le modèle D est ajusté sur l'âge, le centre, les trois PSE et toutes les covariables (consommation d'alcool, tabagisme, activité physique, régime méditerranéen, taille, poids, âge à la première grossesse, statut ménopausique).

6.9 - Association entre la mobilité sociale et le risque de cancer du sein à partir des données imputées de EPIC-Italie [n = 20530].

			EPIC-Italie			
			Modèle 1	Modèle 1 + HB-A	Modèle 1 + RF	Modèle totalement ajusté
Covariables	Référence	Modalité	HR [95%CI]	HR [95%CI]	HR [95%CI]	HR [95%CI]
Mobilité sociale	PSE stable moins avantagée	Mobilité descendante	0.99 [0.83; 1.19]	1.01 [0.84; 1.21]	1.00 [0.84; 1.20]	1.01 [0.85; 1.22]
		PSE stable intermédiaire	1.12 [0.84; 1.50]	1.12 [0.84; 1.50]	1.12 [0.84; 1.50]	1.13 [0.85; 1.51]
		Mobilité ascendante	1.00 [0.84; 1.20]	1.01 [0.84; 1.21]	0.98 [0.82; 1.18]	1.00 [0.83; 1.19]
		PSE stable avantagée	1.32 [0.86; 2.03]	1.34 [0.87; 2.06]	1.27 [0.83; 1.94]	1.30 [0.85; 2.00]

Le modèle 1 est ajusté sur l'âge, le centre et la mobilité sociale.

Le modèle 1 + HB-A est ajusté sur l'âge, le centre, la mobilité sociale, la consommation d'alcool, le tabagisme, l'activité physique, le régime méditerranéen, la taille et le poids.

Le modèle 1 + RF est ajusté sur l'âge, le centre, la mobilité sociale, l'âge à la première grossesse et le statut ménopausique.

Le modèle totalement ajusté est ajusté sur l'âge, le centre, la mobilité sociale et toutes les covariables.

ANNEXE 7 - DOCUMENTS SUPPLEMENTAIRES RELATIFS AU TRAVAIL 4

7.1 - Caractéristiques sociodémographiques, comportementales, anthropométriques et reproductives des femmes ayant des données disponibles dans E3N (N= 2603) selon (A) l'éducation et (B) le statut vis-à-vis du cancer du sein.

Variables	(A) Education							(B) Cancer du sein			
	NA	Description	NA	Moins éduquées	Moyennement éduquées	Plus éduquées	p	NA	Non	Oui	p
Education	72	n= 2531						72	n= 2024	n= 507	0,300
		Faible, n(%)							257 (12.7)	61 (12.03)	
		Moyen, n(%)							997 (49.26)	232 (45.76)	
		Elevé, n(%)							770 (38.04)	214 (42.21)	
Cancer du sein	0	n= 2603	0	n= 318	n= 1229	n= 984	0,226				
		Non, n(%)		257 (80.82)	997 (81.12)	770 (78.25)				-	
		Oui, n(%)		61 (19.18)	232 (18.88)	214 (21.75)					
Temps de suivi	0	n= 2603	0	n= 318	n= 1229	n= 984		0	n= 2084	n= 519	
		Médiane [IQR]		16,3 (10,7 - 17,5)	16,3 (12,2 - 17,5)	16,4 (10,5 - 17,7)	0,908		16.1 (3.1)	5 (3.7)	
Age	0	n= 2603	0	n= 318	n= 1229	n= 984	0,010	0	n= 2084	n= 519	0,300
		>56.4, n(%)		215 (67.61)	798 (64.93)	588 (59.76)			1313 (63)	336 (64.74)	
		≤56.4,, n(%)		103 (32.39)	431 (35.07)	396 (40.24)			771 (37)	183 (35.26)	
Tabac	0	n= 2603	0	n= 318	n= 1229	n= 984	<0,001	0	n= 2084	n= 519	0,800
		Jamais, n(%)		247 (77.67)	909 (73.96)	585 (59.45)			1432 (68.71)	352 (67.82)	
		Ancien, n(%)		54 (16.98)	237 (19.28)	290 (29.47)			477 (22.89)	125 (24.08)	
		Actuelle, n(%)		17 (5.35)	83 (6.75)	109 (11.08)			175 (8.4)	42 (8.09)	
Activité physique	8	n= 2595	8	n= 316	n= 1227	n= 980	<0,001	8	n= 2078	n= 517	0,400
		(13.5,67.5], n(%)		131 (41.46)	460 (37.49)	302 (30.82)			725 (34.89)	195 (37.72)	

	(7.75,13.5], n(%)	836 (32.22)	74 (23.42)	393 (32.03)	348 (35.51)			671 (32.29)	165 (31.91)		
	[0,7.75], n(%)	839 (32.33)	111 (35.13)	374 (30.48)	330 (33.67)			682 (32.82)	157 (30.37)		
Consommation d'alcool		60 n= 2543	57 n= 302	n= 1203	n= 969	0,294	60	n= 2036	n= 507	0,090	
	Modérée, n(%)	1736 (68.27)	210 (69.54)	822 (68.33)	659 (68.01)			1399 (68.71)	337 (66.47)		
	Aucune, n(%)	285 (11.21)	40 (13.25)	139 (11.55)	98 (10.11)			236 (11.59)	49 (9.66)		
	Elevée, n(%)	522 (20.53)	52 (17.22)	242 (20.12)	212 (21.88)			401 (19.7)	121 (23.87)		
Régime occidental		60 n= 2543	57 n= 302	n= 1203	n= 969	0,028	60	n= 2036	n= 507	0,300	
	[-2.93,-0.511], n(%)	924 (36.34)	122 (40.4)	416 (34.58)	362 (37.36)			750 (36.84)	174 (34.32)		
	(-0.511,0.307], n(%)	886 (34.84)	83 (27.48)	434 (36.08)	348 (35.91)			713 (35.02)	173 (34.12)		
	(0.307,7.78], n(%)	733 (28.82)	97 (32.12)	353 (29.34)	259 (26.73)			573 (28.14)	160 (31.56)		
Taille		4 n= 2599	4 n= 318	n= 1226	n= 983	<0,001	4	n= 2082	n= 517	0,090	
	[143,158], n(%)	844 (32.47)	141 (44.34)	410 (33.44)	271 (27.57)			695 (33.38)	149 (28.82)		
	(158,161], n(%)	526 (20.24)	61 (19.18)	252 (20.55)	198 (20.14)			430 (20.65)	96 (18.57)		
	(161,165], n(%)	677 (26.05)	58 (18.24)	337 (27.49)	267 (27.16)			531 (25.5)	146 (28.24)		
	(165,187], n(%)	552 (21.24)	58 (18.24)	227 (18.52)	247 (25.13)			426 (20.46)	126 (24.37)		
Poids		0 n= 2603	0 n= 318	n= 1229	n= 984	0,006	0	n= 2084	n= 519	0,800	
	[35,55], n(%)	649 (24.93)	69 (21.7)	314 (25.55)	249 (25.3)			517 (24.81)	132 (25.43)		
	(55,60], n(%)	660 (25.36)	74 (23.27)	309 (25.14)	262 (26.63)			533 (25.58)	127 (24.47)		
	(60,67], n(%)	651 (25.01)	75 (23.58)	291 (23.68)	268 (27.24)			515 (24.71)	136 (26.2)		
	(67,150], n(%)	643 (24.7)	100 (31.45)	315 (25.63)	205 (20.83)			519 (24.9)	124 (23.89)		
Antécédents familiaux de cancers du sein		61 n= 2542	56 n= 315	n= 1205	n= 955	0,467	61	n= 2033	n= 509	<0,001	
	Non, n(%)	1046 (41.15)	135 (42.86)	513 (42.57)	372 (38.95)			843 (41.47)	203 (39.88)		
	Oui, n(%)	215 (8.46)	26 (8.25)	96 (7.97)	87 (9.11)			149 (7.33)	66 (12.97)		
	Indéfini, n(%)	1281 (50.39)	154 (48.89)	596 (49.46)	496 (51.94)			1041 (51.21)	240 (47.15)		
Antécédents familiaux de cancers ovariens/utérus		0 n= 2603	0 n= 318	n= 1229	n= 984	0,090	0	n= 2084	n= 519	0,500	

	Non, n(%)	2444 (93.89)	304 (95.6)	1142 (92.92)	932 (94.72)		1960 (94.05)	484 (93.26)
	Oui, n(%)	159 (6.11)	14 (4.4)	87 (7.08)	52 (5.28)		124 (5.95)	35 (6.74)
Utilisation d'un THM		5 n= 2598	5 n= 316	n= 1227	n= 983	0,020	5 n= 2080	n= 518 0,400
	Non, n(%)	766 (29.48)	111 (35.13)	370 (30.15)	266 (27.06)		622 (29.9)	144 (27.8)
	Déjà, n(%)	1832 (70.52)	205 (64.87)	857 (69.85)	717 (72.94)		1458 (70.1)	374 (72.2)
Utilisation d'un contraceptif oral		31 n= 2572	31 n= 315	n= 1211	n= 974	<0,001	31 n= 2058	n= 514 0,200
	Non, n(%)	1179 (45.84)	182 (57.78)	566 (46.74)	407 (41.79)		933 (45.34)	246 (47.86)
	Oui, n(%)	1393 (54.16)	133 (42.22)	645 (53.26)	567 (58.21)		1125 (54.66)	268 (52.14)
Age à la première grossesse		20 n= 2583	19 n= 316	n= 1220	n= 976	<0,001	20 n= 2069	n= 514 0,300
	[14,23], n(%)	879 (34.03)	164 (51.9)	480 (39.34)	213 (21.82)		713 (34.46)	166 (32.3)
	(23,26], n(%)	682 (26.4)	67 (21.2)	345 (28.28)	256 (26.23)		551 (26.63)	131 (25.49)
	(26,59], n(%)	696 (26.95)	58 (18.35)	272 (22.3)	340 (34.84)		557 (26.92)	139 (27.04)
	Pas de grossesse, n(%)	326 (12.62)	27 (8.54)	123 (10.08)	167 (17.11)		248 (11.99)	78 (15.18)
Allaitement		76 n= 2527	74 n= 309	n= 1196	n= 952	0,577	76 n= 2020	n= 507 0,900
	Oui, n(%)	1583 (62.64)	192 (62.14)	762 (63.71)	586 (61.55)		1265 (62.62)	318 (62.72)
	Non, n(%)	944 (37.36)	117 (37.86)	434 (36.29)	366 (38.45)		755 (37.38)	189 (37.28)
Durée de vie reproductive		51 n= 2552	49 n= 311	n= 1207	n= 964	0,152	51 n= 2043	n= 509 0,300
	[27; 41[, n(%)	2168 (84.95)	275 (88.42)	1014 (84.01)	816 (84.65)		1743 (85.32)	425 (83.5)
	≥41, n(%)	384 (15.05)	36 (11.58)	193 (15.99)	148 (15.35)		300 (14.68)	84 (16.5)

7.2 - Analyses de régression multinomiale (OR [intervalle de confiance à 95%]) pour l'association entre l'éducation et chacun des biomarqueurs (A. T1 par rapport à T2 et B. T1 par rapport à T3) précédemment associés au cancer du sein, sur des données imputées de E3N (N = 2603). Les résultats sont présentés dans le modèle sans ajustement (modèle 1), après ajustement pour les comportements de santé et les mesures anthropométriques (modèle 1a) ou pour les facteurs reproductifs (modèle 1b) ou les deux (modèle 2).

		(A) Biomarqueur Tercile 2				(B) Biomarqueur Tercile 3			
		Modèle 1	Modèle 1a	Modèle 1b	Modèle 2	Modèle 1	Modèle 1a	Modèle 1b	Modèle 2
Education		OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)
CRP (mg/L)	Moyennement éduquées	0,74 [0,54 ; 1,03]	0,77 [0,56 ; 1,07]	0,79 [0,56 ; 1,09]	0,82 [0,59 ; 1,14]	0,57 [0,42 ; 0,78]	0,66 [0,48 ; 0,91]	0,60 [0,44 ; 0,82]	0,70 [0,50 ; 0,97]
	Plus éduquées	0,85 [0,61 ; 1,18]	0,90 [0,64 ; 1,27]	0,92 [0,66 ; 1,30]	0,98 [0,69 ; 1,39]	0,59 [0,43 ; 0,81]	0,76 [0,55 ; 1,07]	0,66 [0,48 ; 0,92]	0,83 [0,59 ; 1,18]
HDL (mmol/L)	Moyennement éduquées	1,70 [1,25 ; 2,33]	1,60 [1,16 ; 2,20]	1,69 [1,23 ; 2,32]	1,60 [1,16 ; 2,22]	1,38 [1,04 ; 1,84]	1,25 [0,93 ; 1,69]	1,34 [0,99 ; 1,80]	1,22 [0,90 ; 1,66]
	Plus éduquées	1,63 [1,18 ; 2,25]	1,50 [1,08 ; 2,09]	1,70 [1,22 ; 2,37]	1,61 [1,14 ; 2,28]	1,34 [1,00 ; 1,80]	1,18 [0,86 ; 1,61]	1,28 [0,94 ; 1,74]	1,16 [0,84 ; 1,60]
LDL (mmol/L)	Moyennement éduquées	0,95 [0,70 ; 1,28]	0,97 [0,71 ; 1,31]	0,96 [0,70 ; 1,31]	0,98 [0,72 ; 1,34]	1,08 [0,81 ; 1,46]	1,17 [0,87 ; 1,59]	1,09 [0,81 ; 1,48]	1,16 [0,85 ; 1,59]
	Plus éduquées	1,12 [0,82 ; 1,53]	1,17 [0,85 ; 1,60]	1,17 [0,85 ; 1,61]	1,21 [0,87 ; 1,68]	1,04 [0,76 ; 1,41]	1,21 [0,88 ; 1,66]	1,08 [0,79 ; 1,49]	1,22 [0,88 ; 1,70]
Cholestérol (mmol/L)	Moyennement éduquées	1,10 [0,81 ; 1,50]	1,12 [0,82 ; 1,53]	1,09 [0,79 ; 1,49]	1,10 [0,80 ; 1,51]	0,96 [0,72 ; 1,29]	1,02 [0,75 ; 1,38]	0,98 [0,72 ; 1,33]	1,03 [0,76 ; 1,40]
	Plus éduquées	1,18 [0,86 ; 1,62]	1,23 [0,89 ; 1,71]	1,12 [0,81 ; 1,56]	1,16 [0,83 ; 1,63]	0,94 [0,69 ; 1,27]	1,08 [0,79 ; 1,49]	0,96 [0,70 ; 1,33]	1,09 [0,78 ; 1,52]
Fréquence cardiaque (bpm)	Moyennement éduquées	0,97 [0,72 ; 1,32]	0,96 [0,71 ; 1,31]	0,98 [0,72 ; 1,34]	0,97 [0,71 ; 1,33]	0,86 [0,63 ; 1,16]	0,84 [0,62 ; 1,14]	0,85 [0,62 ; 1,16]	0,84 [0,61 ; 1,15]
	Plus éduquées	0,85 [0,62 ; 1,16]	0,84 [0,61 ; 1,15]	0,86 [0,62 ; 1,20]	0,86 [0,62 ; 1,20]	0,82 [0,60 ; 1,12]	0,81 [0,59 ; 1,12]	0,80 [0,58 ; 1,11]	0,80 [0,57 ; 1,12]

Oestradiol (pmol/L)	Moyennement éduquées	1,01 [0,76 ; 1,35]	0,96 [0,72 ; 1,28]	0,92 [0,68 ; 1,25]	0,88 [0,65 ; 1,20]	1,20 [0,87 ; 1,67]	1,12 [0,80 ; 1,56]	1,08 [0,76 ; 1,54]	1,03 [0,72 ; 1,47]
	Plus éduquées	1,20 [0,89 ; 1,61]	1,08 [0,80 ; 1,46]	1,06 [0,77 ; 1,47]	0,98 [0,70 ; 1,36]	1,55 [1,11 ; 2,16]	1,37 [0,97 ; 1,93]	1,40 [0,97 ; 2,02]	1,29 [0,88 ; 1,88]
Testostérone (nmol/L)	Moyennement éduquées	0,75 [0,56 ; 1,02]	0,78 [0,57 ; 1,05]	0,78 [0,58 ; 1,06]	0,81 [0,59 ; 1,10]	0,74 [0,55 ; 1,00]	0,74 [0,55 ; 1,01]	0,76 [0,56 ; 1,03]	0,75 [0,55 ; 1,02]
	Plus éduquées	0,82 [0,60 ; 1,11]	0,85 [0,62 ; 1,17]	0,88 [0,64 ; 1,22]	0,91 [0,65 ; 1,26]	0,76 [0,56 ; 1,03]	0,75 [0,55 ; 1,03]	0,78 [0,56 ; 1,07]	0,76 [0,55 ; 1,05]
SHBG (nmol/L)	Moyennement éduquées	1,14 [0,85 ; 1,53]	1,03 [0,76 ; 1,40]	1,08 [0,80 ; 1,45]	0,96 [0,70 ; 1,32]	1,51 [1,11 ; 2,04]	1,27 [0,92 ; 1,76]	1,44 [1,05 ; 1,97]	1,21 [0,87 ; 1,69]
	Plus éduquées	1,47 [1,08 ; 1,99]	1,27 [0,92 ; 1,75]	1,37 [0,99 ; 1,88]	1,19 [0,85 ; 1,66]	1,93 [1,41 ; 2,63]	1,51 [1,08 ; 2,13]	1,71 [1,23 ; 2,38]	1,37 [0,96 ; 1,95]

Modèle 1 : éducation + âge

Modèle 1a : **Modèle 1** + alcool, tabac, régime occidental, activité physique, poids, taille

Modèle 1b : **Modèle 1** + antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement

Modèle 2 : **Modèle 1** + alcool, tabac, régime occidental, activité physique, poids, taille, antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement

Tercile 1 (ref) CRP : [0 ; 0,91]; HDL : [0,46 ; 1,61]; LDL: [0,36 ; 2,53]; Cholesterol : [0,01 ; 5]; FC : [37 ; 65]; Oestradiol : [18,4 ; 69.1]; Testostérone : [0,069 ; 0,46]; SHBG : [0,35 ; 49,5]

Tercile 2 CRP : (0,91 ; 1,40); HDL : (1,61 ; 2,10); LDL: (2,53 ; 3,47); Cholesterol : (5,00 ; 5,81); FC : (65 ; 73); Oestradiol : (69,1 ; 238); Testostérone : (0,46 ; 0,898); SHBG : (49,5 ; 75,5)

Tercile 3 CRP : (1,40 ; 56); HDL : (2,10 ; 5,11); LDL: (3,47 ; 8,30); Cholesterol : (5,81 ; 11,9); FC : (73 ; 138); Oestradiol : (238 ; 4970); Testostérone : (0,898 ; 47,7); SHBG : (75,5 ; 305)

7.3 - Caractéristiques de la population des cas complets et comparaison avec la population exclue des analyses

Variable	NA	Population avec manquants	Population cas-complets	p
Education	72	n= 696	n= 1835	0,481
Faible, n(%)		80 (11.49)	238 (12.97)	
Moyen, n(%)		335 (48.13)	894 (48.72)	
Élevé, n(%)		281 (40.37)	703 (38.31)	
Diagnostic de cancer du sein	0	n= 768	n= 1835	0,058
Non, n(%)		633 (82.42)	1451 (79.07)	
Oui, n(%)		135 (17.58)	384 (20.93)	
Temps de suivi	0	n= 768	n= 1835	
Médiane[IQR]		16.3[10.9 - 17.5]	16.3[11.0 - 17.6]	0,782
Age	0	n= 768	n= 1835	0,016
>56.4, n(%)		514 (66.93)	1135 (61.85)	
≤56.4, n(%)		254 (33.07)	700 (38.15)	
CRP	0	n= 768	n= 1835	<0,001
[0,0.91], n(%)		314 (40.89)	569 (31.01)	
(0.91,1.4], n(%)		218 (28.39)	617 (33.62)	
(1.4,56], n(%)		236 (30.73)	649 (35.37)	
Triglycérides	0	n= 768	n= 1835	0,006
[0.06,0.77], n(%)		299 (38.93)	595 (32.43)	
(0.77,1.09], n(%)		227 (29.56)	602 (32.81)	
(1.09,5.53], n(%)		242 (31.51)	638 (34.77)	
HDL	0	n= 768	n= 1835	<0,001
[0.46,1.61], n(%)		204 (26.56)	648 (35.31)	
(1.61,2.1], n(%)		236 (30.73)	570 (31.06)	
(2.1,5.11], n(%)		328 (42.71)	617 (33.62)	
LDL	0	n= 768	n= 1835	0,001
[0.36,2.53], n(%)		286 (37.24)	615 (33.51)	
(2.53,3.47], n(%)		198 (25.78)	612 (33.35)	
(3.47,8.3], n(%)		284 (36.98)	608 (33.13)	
Cholestérol	0	n= 768	n= 1835	0,003
[0.01,5], n(%)		248 (32.29)	587 (31.99)	
(5,5.81], n(%)		219 (28.52)	636 (34.66)	
(5.81,11.9], n(%)		301 (39.19)	612 (33.35)	
Fréquence cardiaque	0	n= 768	n= 1835	0,450
[37,65], n(%)		264 (34.38)	639 (34.82)	
(65,73], n(%)		272 (35.42)	606 (33.02)	
(73,138], n(%)		232 (30.21)	590 (32.15)	
Pression artérielle systolique	0	n= 768	n= 1835	0,958
[81.5,116], n(%)		239 (31.12)	573 (31.23)	
(116,132], n(%)		248 (32.29)	601 (32.75)	
(132,220], n(%)		281 (36.59)	661 (36.02)	
Pression artérielle diastolique	0	n= 768	n= 1835	0,117
[42.5,72], n(%)		282 (36.72)	602 (32.81)	
(72,80.5], n(%)		231 (30.08)	612 (33.35)	
(80.5,135], n(%)		255 (33.2)	621 (33.84)	
Oestradiol	0	n= 768	n= 1835	<0,001
[18.4,69.1], n(%)		387 (50.39)	678 (36.95)	
(69.1,238], n(%)		206 (26.82)	643 (35.04)	
(238,4970], n(%)		175 (22.79)	514 (28.01)	

Testostérone	0	n= 768	n= 1835	<0,001
[0.069,0.46], n(%)		402 (52.34)	642 (34.99)	
(0.46,0.898], n(%)		186 (24.22)	606 (33.02)	
(0.898,47.7], n(%)		180 (23.44)	587 (31.99)	
SHBG	0	n= 768	n= 1835	<0,001
[0.35,49.5], n(%)		227 (29.56)	640 (34.88)	
(49.5,75.5], n(%)		228 (29.69)	596 (32.48)	
(75.5,305], n(%)		313 (40.76)	599 (32.64)	

7.4 - Résultats des modèles de Cox (HR [intervalle de confiance à 95 %]) de l'association entre chaque biomarqueur et du risque de cancer du sein sur une population ayant l'ensemble des données renseignées dans E3N (N = 1835).

		M1	M2
		HR(95%CI)	HR(95%CI)
CRP (mg/L)	(0,91 ; 1,40]	2,74 [2,07 ; 3,63]	2,79 [2,10 ; 3,70]
	(1,40 ; 56]	1,96 [1,46 ; 2,62]	2,18 [1,60 ; 2,95]
Triglycérides (mmol/L)	(0,77 ; 1,09]	1,05 [0,82 ; 1,34]	1,08 [0,84 ; 1,38]
	(1,09 ; 5,53]	0,99 [0,78 ; 1,27]	1,05 [0,81 ; 1,36]
HDL (mmol/L)	(1,61 ; 2,10]	0,96 [0,74 ; 1,26]	0,96 [0,73 ; 1,27]
	(2,10 ; 5,11]	1,70 [1,34 ; 2,15]	1,68 [1,32 ; 2,15]
LDL (mmol/L)	(2,53 ; 3,47]	0,57 [0,46 ; 0,72]	0,57 [0,45 ; 0,72]
	(3,47 ; 8,30]	0,30 [0,23 ; 0,40]	0,29 [0,22 ; 0,39]
Cholestérol (mmol/L)	(5,00 ; 5,81]	0,85 [0,67 ; 1,07]	0,87 [0,69 ; 1,10]
	(5,81 ; 11,9]	0,61 [0,47 ; 0,79]	0,61 [0,47 ; 0,80]
Fréquence cardiaque (bpm)	(65 ; 73]	1,12 [0,87 ; 1,43]	1,10 [0,86 ; 1,42]
	(73 ; 138]	1,24 [0,97 ; 1,59]	1,24 [0,97 ; 1,59]
PAS (mmHg)	(116 ; 132]	1,05 [0,81 ; 1,35]	1,05 [0,81 ; 1,35]
	(132 ; 220]	1,06 [0,83 ; 1,36]	1,09 [0,84 ; 1,41]
PAD (mmHg)	(72 ; 80,5]	1,07 [0,84 ; 1,36]	1,07 [0,84 ; 1,37]
	(80,5 ; 135]	0,96 [0,75 ; 1,23]	1,00 [0,78 ; 1,29]
Oestradiol (pmol/L)	(69,1 ; 238]	1,33 [1,05 ; 1,69]	1,38 [1,07 ; 1,78]
	(238 ; 4970]	1,21 [0,93 ; 1,58]	1,25 [0,94 ; 1,67]
Testostérone (nmol/L)	(0,46 ; 0,898]	1,28 [0,97 ; 1,69]	1,33 [1,01 ; 1,76]
	(0,898 ; 47,7]	2,43 [1,89 ; 3,12]	2,47 [1,92 ; 3,18]
SHBG (nmol/L)	(49,5 ; 75,5]	0,92 [0,72 ; 1,17]	0,87 [0,68 ; 1,12]
	(75,5 ; 305]	0,86 [0,68 ; 1,10]	0,79 [0,60 ; 1,03]

Tercile 1 (ref) CRP : [0 ; 0,91]; TG : [0,06 ; 0,77]; HDL : [0,46 ; 1,61]; LDL: [0,36 ; 2,53]; Cholestérol : [0,01 ; 5]; FC : [37 ; 65]; PAS : [81,5 ; 116]; PAD : [42,5 ; 72]; Oestradiol : [18,4 ; 69,1]; Testostérone : [0,069 ; 0,46]; SHBG : [0,35 ; 49,5]

M1 : biomarqueurs + âge

Modèle 2 : Modèle 1 + alcool, tabac, régime occidental, activité physique, poids, taille, antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement.

7.5 - Analyses de régression multinomiale (OR [intervalle de confiance à 95 %]) de l'association entre l'éducation et chacun des biomarqueurs précédemment associés au cancer du sein sur une population ayant l'ensemble des données renseignées dans E3N (N = 1835).

		T1 vs T2		T2 vs T3	
		Modèle 1	Modèle 2	Modèle 1	Modèle 2
Education		OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)
CRP (mg/L)	Moyennement éduquées	0,64 [0,44 ; 0,94]	0,68 [0,46 ; 1,00]	0,58 [0,40 ; 0,83]	0,73 [0,49 ; 1,09]
	Plus éduquées	0,79 [0,54 ; 1,16]	0,84 [0,56 ; 1,26]	0,64 [0,44 ; 0,94]	0,89 [0,59 ; 1,36]
HDL (mmol/L)	Moyennement éduquées	1,67 [1,17 ; 2,40]	1,58 [1,08 ; 2,30]	1,32 [0,95 ; 1,85]	1,16 [0,82 ; 1,66]
	Plus éduquées	1,60 [1,11 ; 2,32]	1,59 [1,07 ; 2,37]	1,18 [0,84 ; 1,67]	1,02 [0,70 ; 1,48]
LDL (mmol/L)	Moyennement éduquées	0,98 [0,69 ; 1,38]	1,03 [0,72 ; 1,48]	1,12 [0,79 ; 1,59]	1,31 [0,91 ; 1,88]
	Plus éduquées	1,30 [0,91 ; 1,85]	1,45 [0,99 ; 2,11]	1,15 [0,80 ; 1,65]	1,48 [1,00 ; 2,18]
Cholestérol (mmol/L)	Moyennement éduquées	1,15 [0,81 ; 1,64]	1,15 [0,80 ; 1,66]	1,02 [0,72 ; 1,45]	1,18 [0,82 ; 1,69]
	Plus éduquées	1,16 [0,81 ; 1,68]	1,13 [0,77 ; 1,66]	0,95 [0,67 ; 1,37]	1,17 [0,79 ; 1,72]
Fréquence cardiaque (bpm)	Moyennement éduquées	0,93 [0,66 ; 1,33]	0,91 [0,64 ; 1,31]	0,78 [0,55 ; 1,10]	0,74 [0,52 ; 1,07]
	Plus éduquées	0,85 [0,59 ; 1,22]	0,84 [0,57 ; 1,24]	0,78 [0,54 ; 1,11]	0,73 [0,50 ; 1,07]
Oestradiol (pmol/L)	Moyennement éduquées	0,99 [0,71 ; 1,37]	0,81 [0,57 ; 1,16]	1,20 [0,82 ; 1,75]	0,96 [0,63 ; 1,47]
	Plus éduquées	1,11 [0,79 ; 1,57]	0,87 [0,59 ; 1,27]	1,57 [1,07 ; 2,32]	1,24 [0,80 ; 1,93]
Testostérone (nmol/L)	Moyennement éduquées	0,80 [0,56 ; 1,15]	0,86 [0,60 ; 1,24]	0,75 [0,53 ; 1,06]	0,75 [0,52 ; 1,07]
	Plus éduquées	0,89 [0,61 ; 1,28]	0,98 [0,66 ; 1,44]	0,68 [0,48 ; 0,98]	0,67 [0,46 ; 0,99]
SHBG (nmol/L)	Moyennement éduquées	1,27 [0,91 ; 1,79]	1,06 [0,74 ; 1,53]	1,46 [1,03 ; 2,08]	1,10 [0,74 ; 1,62]
	Plus éduquées	1,68 [1,18 ; 2,39]	1,33 [0,90 ; 1,96]	1,85 [1,29 ; 2,66]	1,23 [0,81 ; 1,86]

Modèle 1 : éducation + âge

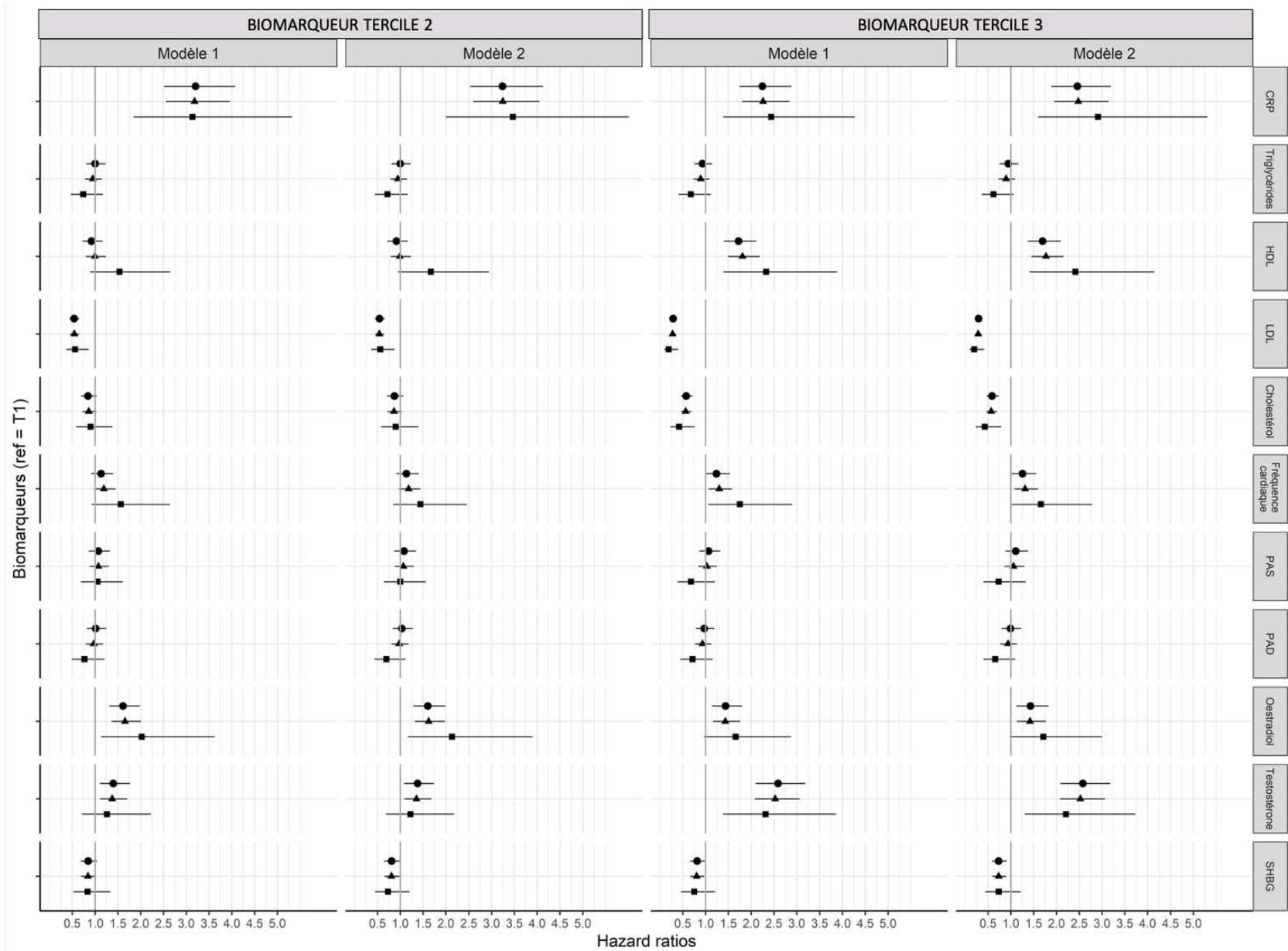
Modèle 2 : Modèle 1 + alcool, tabac, régime occidental, activité physique, poids, taille, antécédents familiaux de cancer du sein ou de cancers hormonaux dépendants, THM, contraceptif oral, âge à la première grossesse, durée de vie reproductive et allaitement.

Tercile 1 (ref) CRP : [0 ; 0,91]; HDL : [0,46 ; 1,61]; LDL: [0,36 ; 2,53]; Cholestérol : [0,01 ; 5]; FC : [37 ; 65]; Oestradiol : [18,4 ; 69.1]; Testostérone : [0,069 ; 0,46]

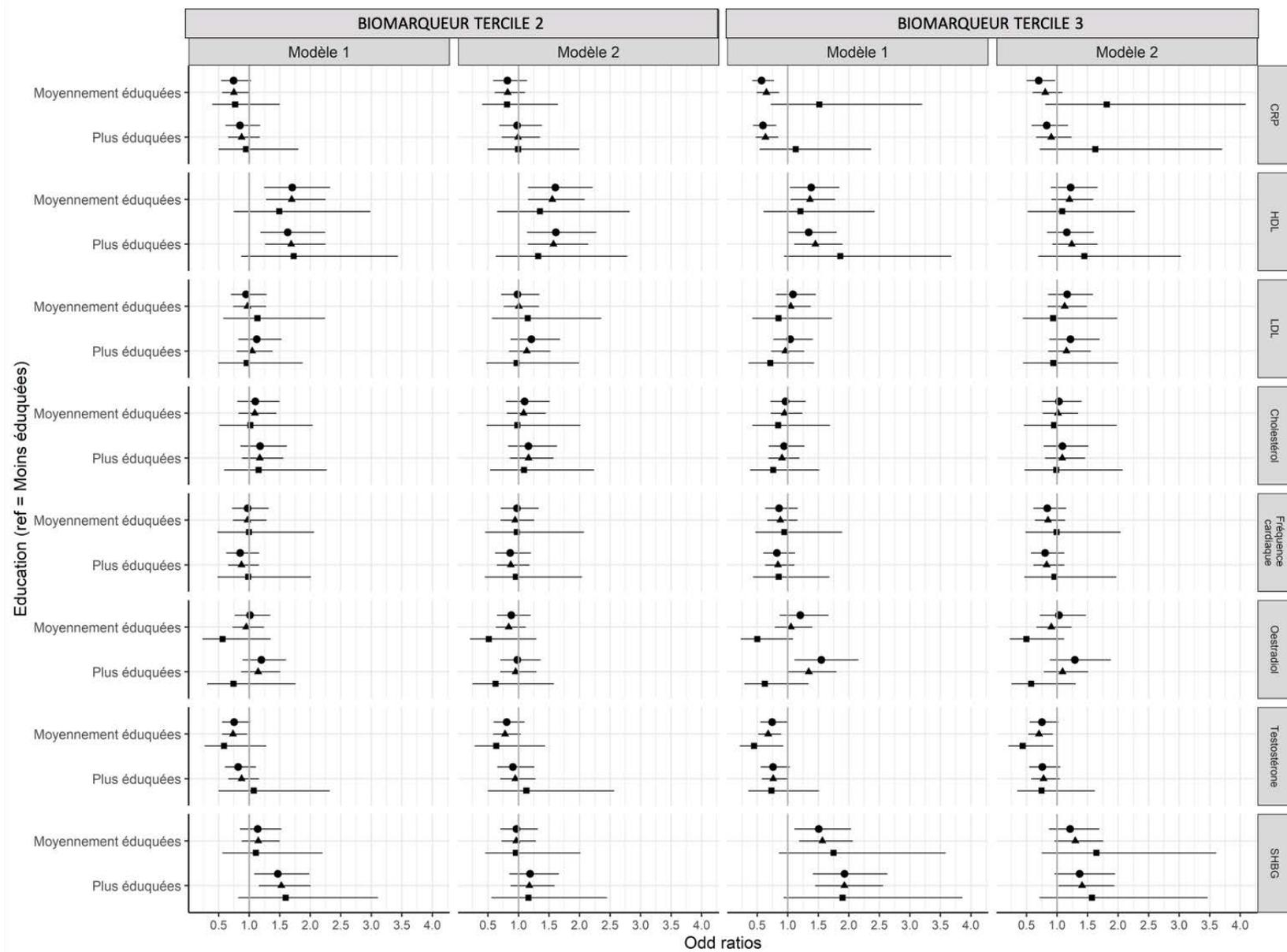
Tercile 2 CRP : (0,91 ; 1,40); HDL : (1,61 ; 2,10); LDL: (2,53 ; 3,47); Cholestérol : (5,00 ; 5,81); FC : (65 ; 73); Oestradiol : (69,1 ; 238); Testostérone : (0,46 ; 0,898]

Tercile 3 CRP : (1,40 ; 56); HDL : (2,10 ; 5,11); LDL: (3,47 ; 8,30); Cholestérol : (5,81 ; 11,9); FC : (73 ; 138); Oestradiol : (238 ; 4970); Testostérone : (0,898 ; 47,7]

7.6 - Représentation graphique des résultats des modèles de Cox (HR [intervalle de confiance à 95%]) pour l'association entre chaque marqueur biologique et le risque de cancer du sein post-ménopausique sur les données imputées de E3N. Les résultats sont reportés en fonction du statut ménopausique au moment du prélèvement sanguin : (●) femmes ménopausées (N = 2603), (▲) femmes en pré et post-ménopause (N = 3251) et (■) femmes pré-ménopausées (N = 634).



7.7 - Représentation graphique des analyses de régression multinomiale (OR [intervalle de confiance à 95%]) pour l'association entre l'éducation et chaque marqueur biologique précédemment identifié comme étant associé au risque de cancer du sein post-ménopausique sur les données imputées de E3N. Les résultats sont reportés en fonction du statut ménopausique au moment du prélèvement sanguin : (●) femmes ménopausées (N = 2603), (▲) femmes en pré et post-ménopause (N = 3251) et (■) femmes en pré-ménopause (N = 634).



7.8 - Comparaison des caractéristiques des femmes avec des données biologiques par rapport aux femmes qui n'avaient pas d'information sur les biomarqueurs dans la population utilisée dans le cadre du travail 5 (N = 83 436).

Variable	NA	Population sans marqueur biologique	Population avec biomarqueurs	p
Cancer du sein	0	n= 80132	n= 3304	<0,001
Non, n(%)		72930 (91.01)	2629 (79.57)	
Oui, n(%)		7202 (8.99)	675 (20.43)	
Age	0	n= 80132	n= 3304	<0,001
< 47.9, n(%)		41857 (52.24)	1471 (44.52)	
≥ 47.9, n(%)		38275 (47.76)	1833 (55.48)	
Education	3086	n= 77132	n= 3218	<0,001
Elevé, n(%)		27680 (35.89)	1329 (41.3)	
Moyen, n(%)		38659 (50.12)	1508 (46.86)	
Faible, n(%)		10793 (13.99)	381 (11.84)	
Tabagisme	32	n= 80100	n= 3304	<0,001
Jamais, n(%)		53518 (66.81)	2221 (67.22)	
Ancien, n(%)		17794 (22.21)	796 (24.09)	
Actuel, n(%)		8788 (10.97)	287 (8.69)	
Activité physique	8358	n= 71784	n= 3294	0,018
(13.2,92.1], n(%)		23805 (33.16)	1131 (34.34)	
(7.38,13.2], n(%)		23973 (33.4)	1140 (34.61)	
[0,7.38], n(%)		24006 (33.44)	1023 (31.06)	
Consommation d'alcool	20500	n= 59706	n= 3230	0,005
Modérée, n(%)		40669 (68.12)	2218 (68.67)	
Aucune, n(%)		7377 (12.36)	341 (10.56)	
Elevée, n(%)		11660 (19.53)	671 (20.77)	
Régime occidental	20500	n= 59706	n= 3230	0,059
[-2.93,-0.511], n(%)		19702 (33)	1095 (33.9)	
(-0.511,0.307], n(%)		19847 (33.24)	1110 (34.37)	
(0.307,7.78], n(%)		20157 (33.76)	1025 (31.73)	
Taille	596	n= 79540	n= 3300	0,913
[132,158], n(%)		24063 (30.25)	1001 (30.33)	
(158,161], n(%)		16241 (20.42)	657 (19.91)	
(161,165], n(%)		20774 (26.12)	869 (26.33)	
(165,189], n(%)		18462 (23.21)	773 (23.42)	
Poids	169	n= 79963	n= 3304	0,644
[29,55], n(%)		20520 (25.66)	818 (24.76)	
(55,60], n(%)		19515 (24.41)	803 (24.3)	
(60,67], n(%)		19955 (24.96)	841 (25.45)	
(67,170], n(%)		19973 (24.98)	842 (25.48)	

Antécédents familiaux de cancers du sein	2070	n= 78133	n= 3233	0,037
Non, n(%)		33296 (42.61)	1365 (42.22)	
Oui, n(%)		5845 (7.48)	281 (8.69)	
Indéfini, n(%)		38992 (49.9)	1587 (49.09)	
Antécédents familiaux de cancers ovaires/utérus	0	n= 80132	n= 3304	0,008
Non, n(%)		76521 (95.49)	3122 (94.49)	
Oui, n(%)		3611 (4.51)	182 (5.51)	
Utilisation d'un THM	7616	n= 72523	n= 3297	<0,001
Non, n(%)		38672 (53.32)	1439 (43.65)	
Déjà, n(%)		33851 (46.68)	1858 (56.35)	
Utilisation d'un contraceptif oral	4068	n= 76099	n= 3269	0,001
Non, n(%)		27634 (36.31)	1278 (39.09)	
Oui, n(%)		48465 (63.69)	1991 (60.91)	
Age à la première grossesse	2758	n= 77399	n= 3279	0,018
[14,23], n(%)		27741 (35.84)	1094 (33.36)	
(23,26], n(%)		20569 (26.58)	877 (26.75)	
(26,59], n(%)		19609 (25.33)	890 (27.14)	
Pas d'enfants, n(%)		9480 (12.25)	418 (12.75)	
Allaitement	8539	n= 71689	n= 3208	<0,001
Oui, n(%)		42237 (58.92)	2032 (63.34)	
Non, n(%)		29452 (41.08)	1176 (36.66)	
Durée de vie reproductive	6231	n= 73977	n= 3228	0,213
<41, n(%)		60669 (82.01)	2619 (81.13)	
≥41, n(%)		13308 (17.99)	609 (18.87)	
Participation au dépistage	3130	n= 77024	n= 3282	<0,001
Non, n(%)		13001 (16.88)	395 (12.04)	
Oui, n(%)		64023 (83.12)	2887 (87.96)	

ANNEXE 8 - DOCUMENTS SUPPLEMENTAIRES RELATIFS AU TRAVAIL 5

8.1 - Caractéristiques des femmes avec les données disponibles dans E3N selon la PSE

Variables	NA	Description	Education			p
			NA	Elevé	Faible	
Récepteurs aux oestrogènes	879	n= 5822	842	n= 2134	n= 3465	0,183
Positif, n(%)		4775 (82.02)		1773 (83.08)	2829 (81.65)	
Négatif, n(%)		1047 (17.98)		361 (16.92)	636 (18.35)	
Récepteurs à la progestérone	1045	n= 5656	1001	n= 2049	n= 3391	0,324
Positif, n(%)		3665 (64.8)		1310 (63.93)	2214 (65.29)	
Négatif, n(%)		1991 (35.2)		739 (36.07)	1177 (34.71)	
Surexpression de la protéine HER2	3810	n= 2891	3657	n= 1118	n= 1666	0,473
Surexpression, n(%)		384 (13.28)		141 (12.61)	227 (13.63)	
Pas de surexpression, n(%)		2507 (86.72)		977 (87.39)	1439 (86.37)	
Stade	206	n= 6495	199	n= 2404	n= 3838	0,903
1, n(%)		3998 (61.56)		1491 (62.02)	2360 (61.49)	
2, n(%)		1898 (29.22)		691 (28.74)	1123 (29.26)	
3-4, n(%)		599 (9.22)		222 (9.23)	355 (9.25)	
Grade	976	n= 5725	942	n= 2099	n= 3400	0,059
1, n(%)		782 (13.66)		315 (15.01)	435 (12.79)	
2, n(%)		1969 (34.39)		700 (33.35)	1186 (34.88)	
3-4, n(%)		2974 (51.95)		1084 (51.64)	1779 (52.32)	
Sous type récepteurs hormonaux	883	n= 5818	845	n= 2131	n= 3465	0,106
RH+, n(%)		4955 (85.17)		1840 (86.34)	2936 (84.73)	
RH-, n(%)		863 (14.83)		291 (13.66)	529 (15.27)	
Sous type triples négatifs	3856	n= 2845	3700	n= 1103	n= 1638	0,882
Non, n(%)		2589 (91)		1006 (91.21)	1490 (90.96)	
Oui, n(%)		256 (9)		97 (8.79)	148 (9.04)	
Age	0	n= 6701	0	n= 2474	n= 3967	<0,001
< 47.9 ans, n(%)		3245 (48.43)		1423 (57.52)	1707 (43.03)	
≥ 47.9 ans, n(%)		3456 (51.57)		1051 (42.48)	2260 (56.97)	
Alcool	1438	n= 5263	1361	n= 2000	n= 3080	<0,001
Aucune, n(%)		564 (10.72)		199 (9.95)	339 (11.01)	
Modérée, n(%)		3535 (67.17)		1302 (65.1)	2120 (68.83)	
Élevée, n(%)		1164 (22.12)		499 (24.95)	621 (20.16)	
Tabac	24	n= 6677	17	n= 2466	n= 3958	<0,001
Jamais, n(%)		4496 (67.34)		1490 (60.42)	2832 (71.55)	
Anciennement, n(%)		1309 (19.6)		577 (23.4)	687 (17.36)	
Actuellement, n(%)		872 (13.06)		399 (16.18)	439 (11.09)	
Régime occidental	1438	n= 5263	1361	n= 2000	n= 3080	0,682
[-2.93,-0.511], n(%)		1612 (30.63)		602 (30.1)	951 (30.88)	
(-0.511,0.307], n(%)		1765 (33.54)		668 (33.4)	1041 (33.8)	
(0.307,7.78], n(%)		1886 (35.84)		730 (36.5)	1088 (35.32)	

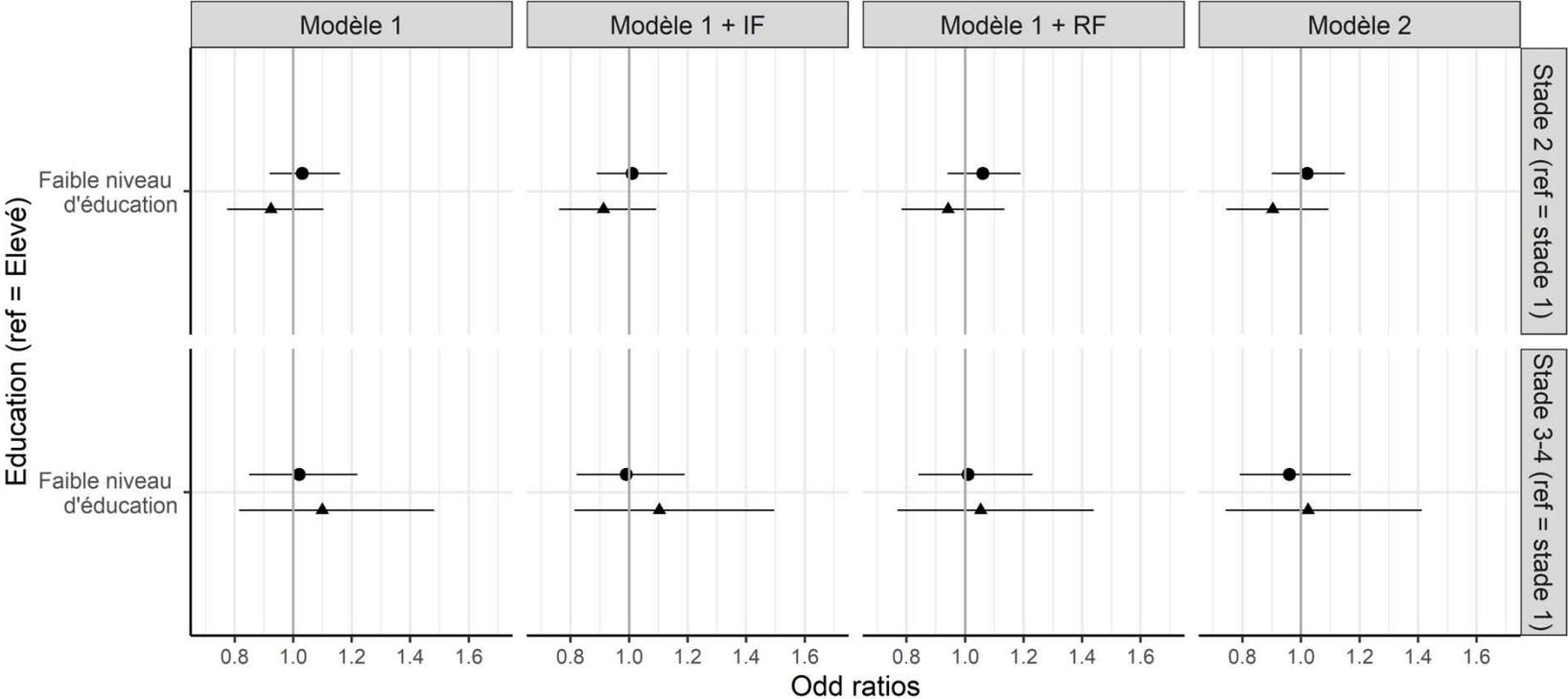
Activité physique	89	n= 6612	67	n= 2448	n= 3926	<0,001
(7.57,35.9], n(%)		2023 (30.6)		640 (26.14)	1313 (33.44)	
(4.66,7.57], n(%)		2252 (34.06)		828 (33.82)	1345 (34.26)	
[0,4.66], n(%)		2337 (35.34)		980 (40.03)	1268 (32.3)	
Taille	96	n= 6605	73	n= 2440	n= 3928	<0,001
[135,160], n(%)		2841 (43.01)		931 (38.16)	1810 (46.08)	
(160,164], n(%)		1487 (22.51)		568 (23.28)	871 (22.17)	
(164,190], n(%)		2277 (34.47)		941 (38.57)	1247 (31.75)	
Poids	130	n= 6571	106	n= 2432	n= 3903	<0,001
[29,55], n(%)		2119 (32.25)		843 (34.66)	1211 (31.03)	
(55,62], n(%)		2240 (34.09)		853 (35.07)	1311 (33.59)	
(62,163], n(%)		2212 (33.66)		736 (30.26)	1381 (35.38)	
Age aux premières règles	147	n= 6554	127	n= 2419	n= 3895	0,153
≥12, n(%)		5123 (78.17)		1867 (77.18)	3067 (78.74)	
<12, n(%)		1431 (21.83)		552 (22.82)	828 (21.26)	
Antécédents familiaux de cancers ovaires/utérus	0	n= 6701	0	n= 2474	n= 3967	0,971
Non, n(%)		6333 (94.51)		2340 (94.58)	3750 (94.53)	
Oui, n(%)		368 (5.49)		134 (5.42)	217 (5.47)	
Antécédents familiaux de cancers du sein	168	n= 6533	145	n= 2409	n= 3887	0,730
Non, n(%)		2581 (39.51)		964 (40.02)	1518 (39.05)	
Oui, n(%)		747 (11.43)		277 (11.5)	447 (11.5)	
Non définissable, n(%)		3205 (49.06)		1168 (48.48)	1922 (49.45)	
Utilisation d'un contraceptif oral	0	n= 6701	0	n= 2474	n= 3967	<0,001
Non, n(%)		3020 (45.07)		941 (38.04)	1943 (48.98)	
Oui, n(%)		3681 (54.93)		1533 (61.96)	2024 (51.02)	
Utilisation d'un THM	652	n= 6049	614	n= 2276	n= 3551	0,018
Non, n(%)		5141 (84.99)		1974 (86.73)	2984 (84.03)	
Oui, n(%)		671 (11.09)		226 (9.93)	419 (11.8)	
Non définissable, n(%)		237 (3.92)		76 (3.34)	148 (4.17)	
Nombre d'enfants	0	n= 6701	0	n= 2474	n= 3967	<0,001
3+, n(%)		1727 (25.77)		618 (24.98)	1042 (26.27)	
1-2, n(%)		4092 (61.07)		1427 (57.68)	2518 (63.47)	
0, n(%)		882 (13.16)		429 (17.34)	407 (10.26)	
Age à la première grossesse	180	n= 6521	176	n= 2408	n= 3857	<0,001
[14,23], n(%)		2091 (32.07)		468 (19.44)	1551 (40.21)	
(23,26], n(%)		1673 (25.66)		613 (25.46)	994 (25.77)	
(26,59], n(%)		1875 (28.75)		898 (37.29)	905 (23.46)	
Pas d'enfants, n(%)		882 (13.53)		429 (17.82)	407 (10.55)	
Durée d'allaitement	631	n= 6070	607	n= 2262	n= 3572	0,002
(5.5,120], n(%)		1052 (17.33)		411 (18.17)	603 (16.88)	
(2.5,5.5], n(%)		1007 (16.59)		411 (18.17)	545 (15.26)	
[1,2.5], n(%)		1441 (23.74)		496 (21.93)	903 (25.28)	
0, n(%)		2570 (42.34)		944 (41.73)	1521 (42.58)	
Durée de vie reproductive	311	n= 6390	284	n= 2347	n= 3810	0,143
[27; 41[, n(%)		5156 (80.69)		1872 (79.76)	3098 (81.31)	

	≥41, n(%)	1234 (19.31)	475 (20.24)	712 (18.69)	
Participation au dépistage	0	n= 6701	0	n= 2474	n= 3967 0,006
Oui, n(%)		5093 (76)	1937 (78.29)	2987 (75.3)	
Non, n(%)		1608 (24)	537 (21.71)	980 (24.7)	
Statut ménopausique	51	n= 6650	44	n= 2455	n= 3942 <0,001
Pre-ménopause, n(%)		1061 (15.95)	457 (18.62)	565 (14.33)	
Post-ménopause, n(%)		5589 (84.05)	1998 (81.38)	3377 (85.67)	

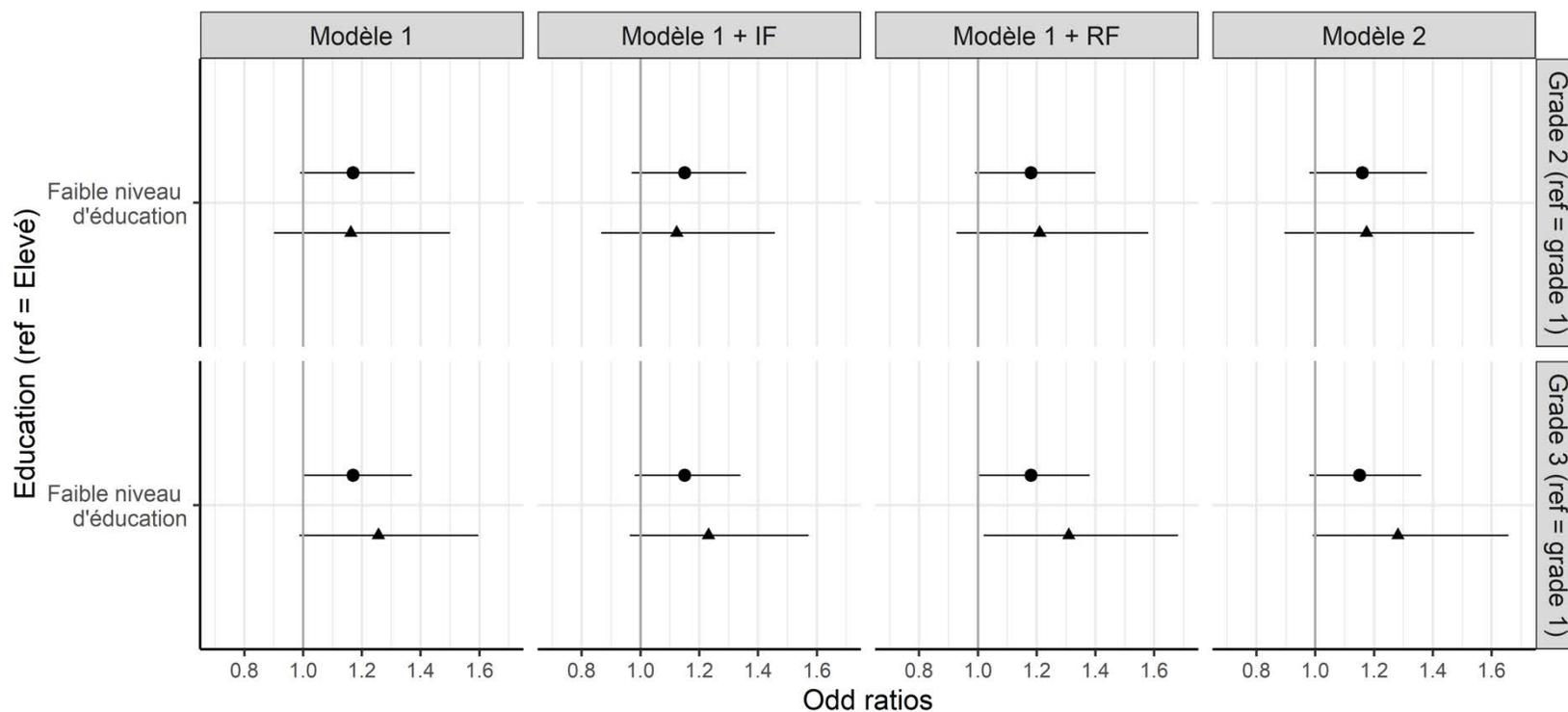
8.2 - Caractéristiques de la population des cas complets et comparaison avec la population exclue des analyses

Variable	Manquants	Population avec manquants	Population cas-complets	p
Education	260	n= 3880	n= 2561	<0,001
Elevé, n(%)		1419 (36.57)	1055 (41.19)	
Faible, n(%)		2461 (63.43)	1506 (58.81)	
Stade	206	n= 3934	n= 2561	0,063
1, n(%)		2423 (61.59)	1575 (61.5)	
2, n(%)		1124 (28.57)	774 (30.22)	
3-4, n(%)		387 (9.84)	212 (8.28)	
Grade	976	n= 3164	n= 2561	0,012
1, n(%)		429 (13.56)	353 (13.78)	
2, n(%)		1140 (36.03)	829 (32.37)	
3-4, n(%)		1595 (50.41)	1379 (53.85)	
Sous type RH	883	n= 3257	n= 2561	0,082
RH+, n(%)		2750 (84.43)	2205 (86.1)	
RH-, n(%)		507 (15.57)	356 (13.9)	
Age	0	n= 4140	n= 2561	<0,001
< 47.9, n(%)		1899 (45.87)	1346 (52.56)	
≥ 47.9, n(%)		2241 (54.13)	1215 (47.44)	

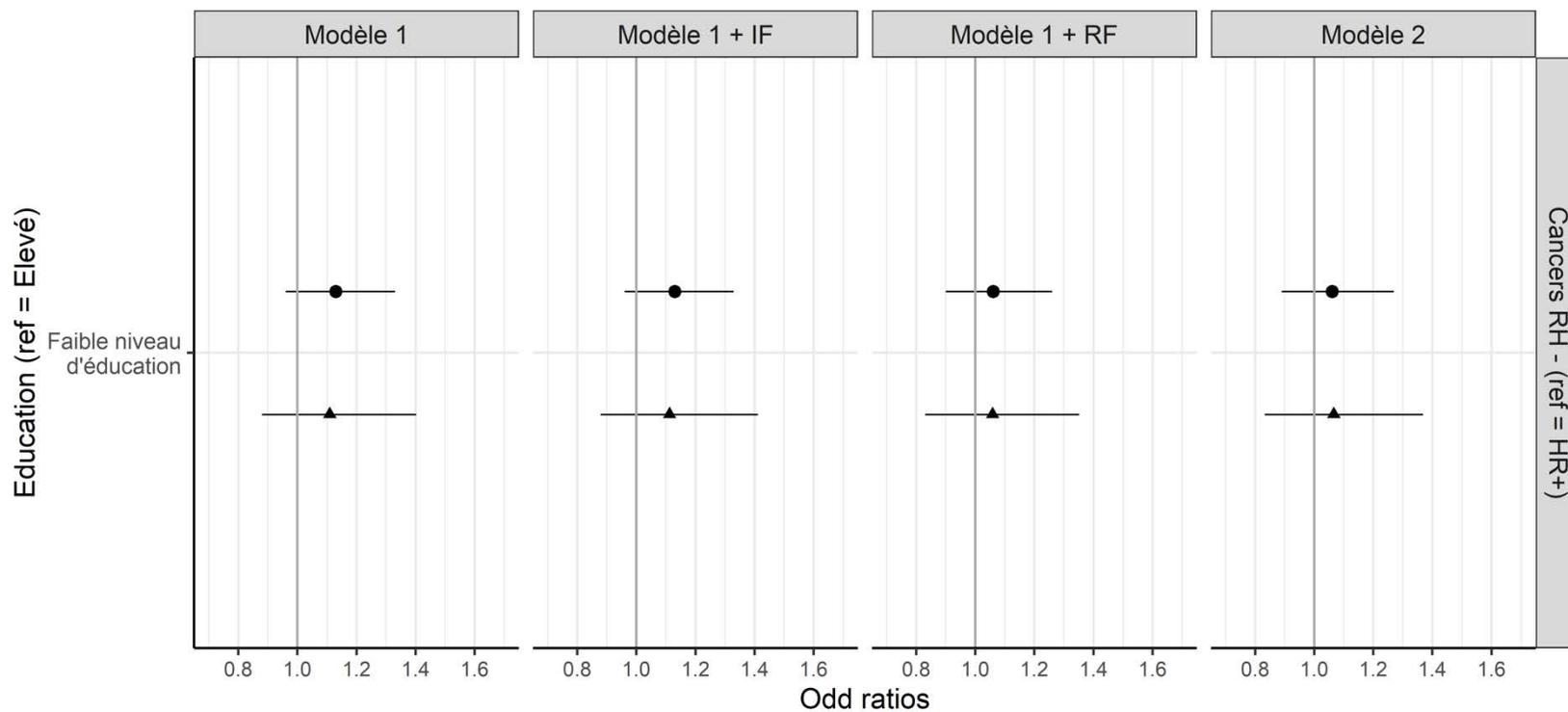
8.3 - Résultats des modèles de régression logistique multivariés de l'association entre l'éducation et le stade du cancer du sein à partir (●) des données imputées de E3N (N = 6,701) et (▲) des données en cas complets (N = 2,561). Modèle 1 + IF est ajusté sur les comportements de santé et les mesures anthropométriques et le modèle 1 + RF est ajusté sur les facteurs reproductifs.



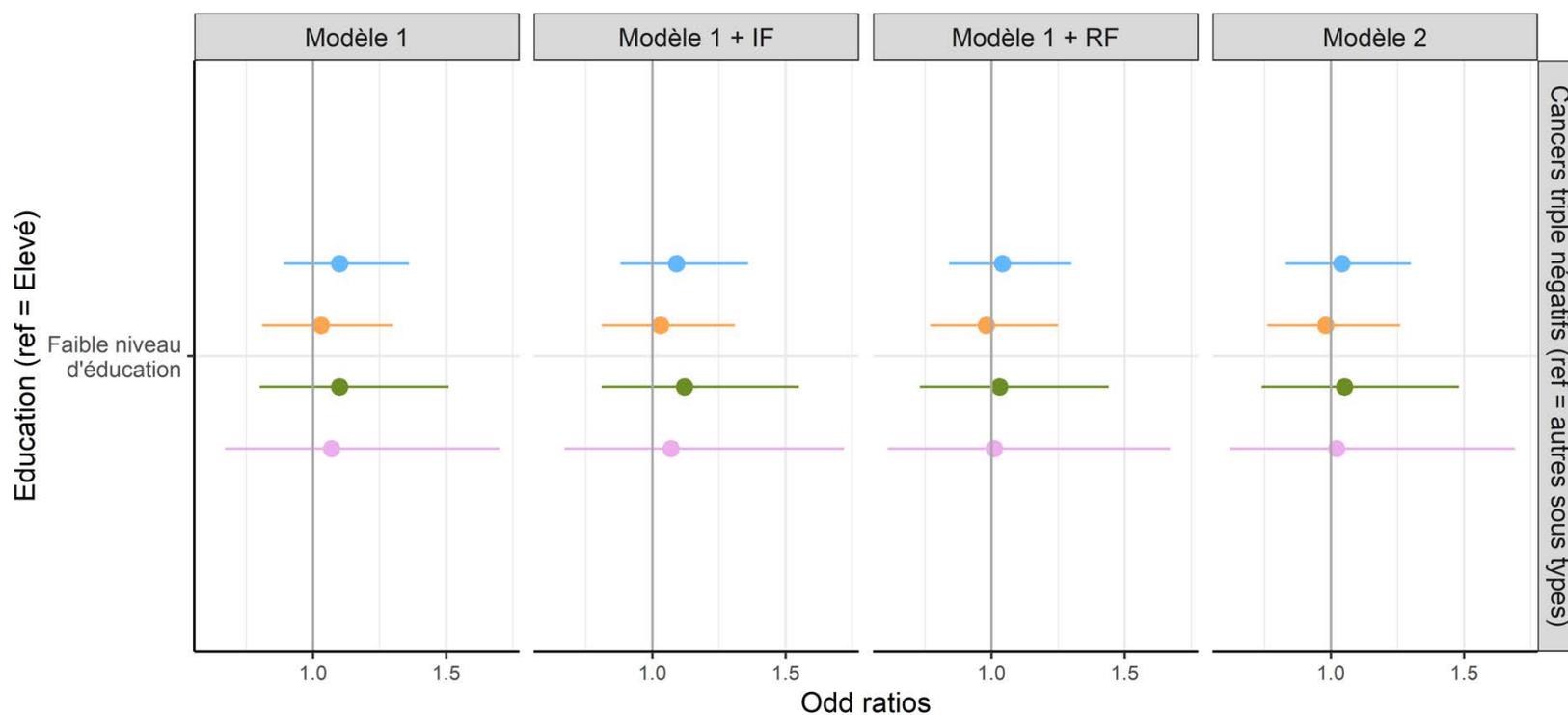
8.4 - Résultats des modèles de régression logistique multivariés de l'association entre l'éducation et le grade du cancer du sein à partir (●) des données imputées de E3N (N = 6,701) et (▲) des données en cas complets (N = 2,561). Modèle 1 + IF est ajusté sur les comportements de santé et les mesures anthropométriques et le modèle 1 + RF est ajusté sur les facteurs reproductifs.



8.5 - Résultats des modèles des régressions logistiques multivariées de l'association entre l'éducation et le sous type RH du cancer du sein à partir (●) des données imputées de E3N (N = 6,701) et (▲) des données en cas complets (N = 2,561). Modèle 1 + IF est ajusté sur les comportements de santé et les mesures anthropométriques et le modèle 1 + RF est ajusté sur les facteurs reproductifs.



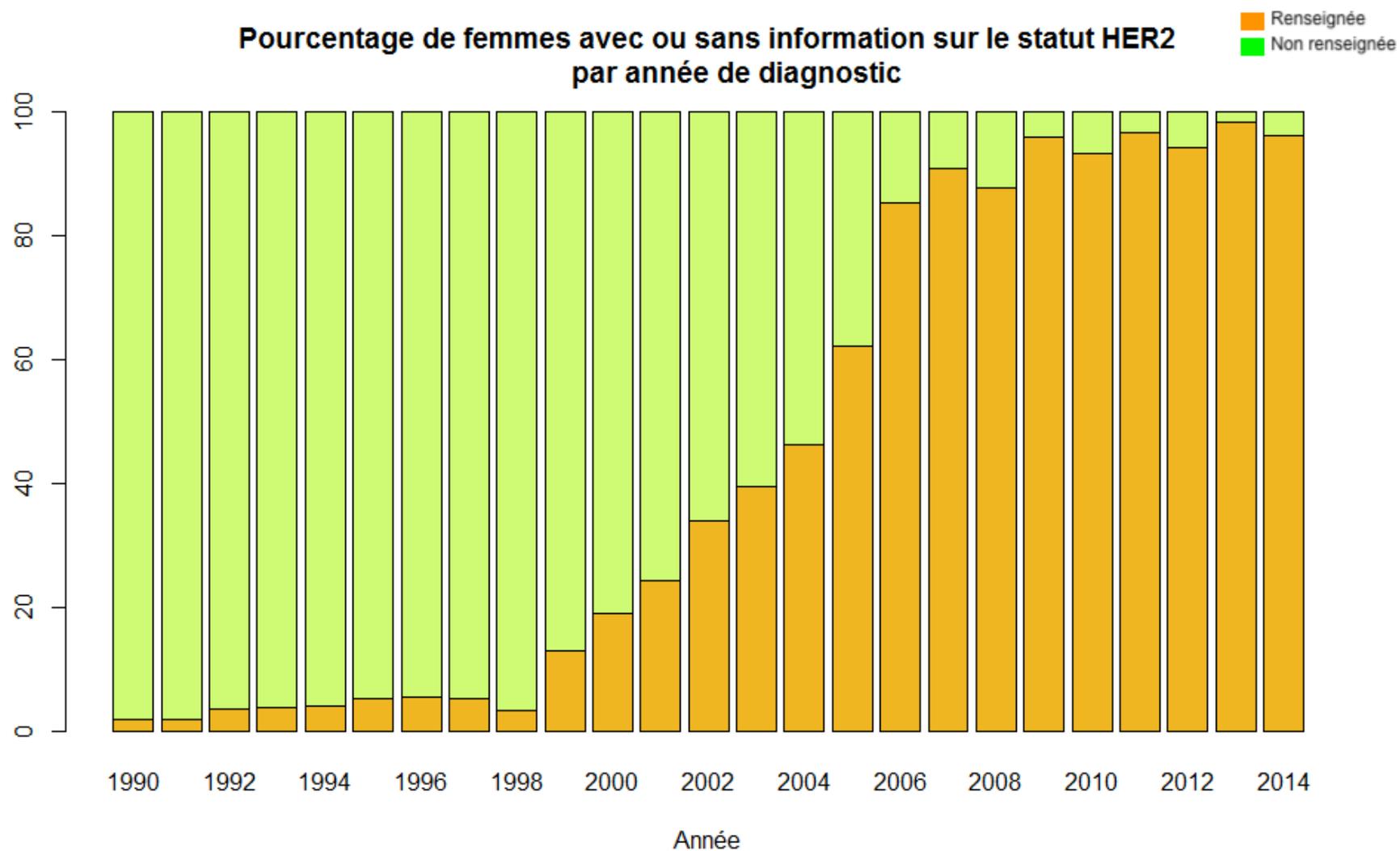
8.6 - Résultats des modèles de régression logistique multivariés de l'association entre l'éducation et le sous type (triple négatif vs autres) du cancer du sein à partir des données imputées de E3N suivant les périodes d'étude. Le point bleu représente tous les cas (N = 6,701), le point orange les cas à partir de 1999 (N = 4,540), le point vert les cas à partir de 2005 (N = 2,465) et le point rose à partir de 2009 (N = 1,286). Modèle 1 + IF est ajusté sur les comportements de santé et les mesures anthropométriques et le modèle 1 + RF est ajusté sur les facteurs reproductifs.



8.7 - Résultats des modèles de régression logistique de l'association entre l'éducation et le sous type RH ajusté sur le nombre de grossesses

	Education
	OR (95%CI)
M1	1,13 [0,96 ; 1,33]
M1 + Age à la première grossesse	1,07 [0,91 ; 1,26]
M1 + Nombre de grossesses	1,11 [0,94 ; 1,30]

8.8 - Proportion de femmes avec ou sans information sur le statut HER2 par année de diagnostic dans E3N



ANNEXE 9 - ACTIVITES PENDANT LA THESE

ENSEIGNEMENTS

Septembre 2017/2018/2019/2020 :

Université Toulouse III - Paul Sabatier
M2R Épidémiologie clinique
Remise à niveau logiciel STATA (12h)

Mars 2018 :

Institut Catholique de Toulouse
M1 Psychologie clinique
Recherche: outils statistiques - Utilisation du logiciel STATISTICA' (12h)

ACTIVITES ANNEXES

Juin 2020 : Rédaction d'un article vulgarisé du travail de recherche dans le cadre d'une formation doctorale : Inégalités sociales et cancer du sein : une relation atypique | Exploreur [Internet]. Available from: <https://exploreur.univ-toulouse.fr/inegalites-sociales-et-cancer-du-sein-une-relation-atypique>

2018-2021 : Membre du groupe de doctorants Santé-Société de Institut Fédératif d'Études et de Recherches Interdisciplinaires Santé-Société [IFERISS - FED 4142]. En mars 2019, nous avons organisé un séminaire autour des inégalités sociales de santé en collaboration avec des doctorants de l'Université de Québec à Montréal (UQAM) au cours duquel j'ai pu présenter mon travail de thèse.

Titre : Construction des inégalités sociales de cancer du sein : mécanismes impliqués dans l'incidence et la présentation initiale de la maladie

Mots clés : Cancer du sein ; inégalités sociales ; incorporation biologique ; déterminants sociaux ; facteurs pronostics ; sous-types moléculaires ; approche life course

Des inégalités sociales de mortalité par cancer du sein ont été mises en évidence dans la littérature. Nous faisons l'hypothèse que les inégalités sociales observées dans la mortalité par cancer du sein sont la conséquence d'une distribution socialement différenciée de l'incidence de la maladie mais aussi de ses caractéristiques initiales, et de l'accès au système de soin. Les mécanismes par lesquels la position socio-économique (PSE) influence l'incidence, les caractéristiques initiales et la survie ne sont pas pleinement compris. A partir des données de la cohorte prospective française E3N regroupant 98 995 femmes, ce travail de thèse cherche à étudier la façon dont la PSE peut favoriser la survenue des cancers du sein et influencer leurs caractéristiques au moment du diagnostic, via des mécanismes comportementaux, anthropométriques et reproductifs mais aussi biologiques. Les résultats de ce travail mettent en évidence un lien entre la PSE et l'incidence/les caractéristiques initiales du cancer. Les facteurs reproductifs, notamment l'âge à la première grossesse, joue un rôle important sur le risque de cancer du sein mais d'autres mécanismes d'incorporation biologique de l'environnement social en plus des facteurs de risque classiques, pourraient être impliqués dans les inégalités sociales d'incidence et de caractéristiques des tumeurs.

Title : Construction of social inequalities in breast cancer: mechanisms involved in the incidence and initial presentation of the disease

Keywords : Breast cancer; social inequalities; biological embodiment; social determinants; prognostic factors; molecular subtypes; life course approach

Social inequalities in breast cancer mortality have been identified in the literature. We hypothesise that the observed social inequalities in breast cancer mortality are the consequence of a socially differentiated distribution of the incidence of the disease but also of its initial characteristics, and of access to the health care system. The mechanisms by which socio-economic position (SEP) influences incidence, baseline characteristics and survival are not fully understood. Using data from the French prospective E3N cohort of 98,995 women, this thesis aims to investigate how SEP may promote the occurrence of breast cancer and influence its characteristics at diagnosis, via behavioural, anthropometric, reproductive and biological mechanisms. The results of this work show an association between SEP and the incidence/initial characteristics of cancer. Reproductive factors, especially age at first childbirth, play an important role in breast cancer risk, but other mechanisms of biological embodiment of the social environment, in addition to the classical risk factors, may be involved in social inequalities in tumour incidence and characteristics.