



HAL
open science

Facteurs comportementaux et non-comportementaux associés au risque de cancer et de mortalité à partir des données de la cohorte de femmes françaises E3N

Laureen Dartois

► **To cite this version:**

Laureen Dartois. Facteurs comportementaux et non-comportementaux associés au risque de cancer et de mortalité à partir des données de la cohorte de femmes françaises E3N. Santé publique et épidémiologie. Université Paris Sud - Paris XI, 2014. Français. NNT : 2014PA11T081 . tel-01242552

HAL Id: tel-01242552

<https://theses.hal.science/tel-01242552>

Submitted on 13 Dec 2015

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

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

UNIVERSITÉ PARIS-SUD XI
FACULTÉ DE MÉDECINE PARIS-SUD
ÉCOLE DOCTORALE 420, SANTÉ PUBLIQUE PARIS-SUD XI, PARIS DESCARTES
DISCIPLINE : Épidémiologie et Santé Publique

THÈSE DE DOCTORAT

Présentée et soutenue publiquement le 12/12/2014

par

Laureen DARTOIS

Facteurs comportementaux et non-comportementaux associés
au risque de cancer et de mortalité à partir des données de la
cohorte de femmes françaises E3N

Directrice de thèse : Mme Françoise CLAVEL-CHAPELON

Composition du jury :

Président du jury : M. Alexis ELBAZ

Rapporteurs : M. Patrick ARVEUX

M. David COX

Examineurs : Mme Martine BELLANGER

M. Vivian VIALLO

Directrice de thèse : Mme Françoise CLAVEL-CHAPELON

AVANT-PROPOS

Le cancer est l'un des grands maux du XXI^{ème} siècle avec 355 000 nouveaux cas diagnostiqués (155 000 chez les femmes) et 148 000 décès (63 000 chez les femmes) en France en 2012. Le cancer est la première cause de mortalité en France, et la deuxième chez les femmes après les maladies de l'appareil circulatoire. Parmi les femmes âgées de moins de 84 ans, le cancer constitue la première cause de mortalité et le cancer du sein est à l'origine du plus grand nombre de décès par cancer.

Plusieurs plans gouvernementaux ont souligné l'importance d'améliorer la prévention, le traitement et la prise en charge des patients, afin de réduire l'incidence et la mortalité. Néanmoins, malgré une stabilisation, voire une diminution, de la mortalité par cancer, l'incidence de cancer est en constante augmentation en France.

Le cancer est une maladie multifactorielle, dont certains facteurs de risque sont modifiables, ce qui pourrait permettre d'améliorer la prévention. C'est dans ce contexte que se situe ce travail de thèse. À partir des données de la cohorte de femmes E3N (Étude épidémiologique auprès de femmes de la Mutuelle générale de l'Éducation nationale), ce travail vise à mieux comprendre l'influence qu'auraient les facteurs comportementaux et non-comportementaux sur le risque de cancer et de mortalité.

RÉSUMÉ

Titre : Facteurs comportementaux et non-comportementaux associés au risque de cancer et de mortalité à partir des données de la cohorte de femmes françaises E3N

Contexte : Le cancer est la seconde cause de mortalité chez la femme en France, et la première chez les femmes âgées de 35 à 84 ans. Le cancer du sein est le cancer le plus fréquemment diagnostiqué, représentant 35 % des cas chez les femmes en France en 2012. De multiples facteurs, comportementaux et non-comportementaux, augmentant le risque de cancer, tant en incidence qu'en mortalité, ont été identifiés dans la littérature, tandis que leur influence conjointe est très peu évaluée. Dans le cas du cancer du sein, certains facteurs diffèrent selon le statut ménopausique des femmes, suggérant une étiologie différente entre les cancers du sein diagnostiqués avant et après la ménopause.

Objectif : Les données de la cohorte prospective française E3N ont été utilisées pour évaluer l'influence des facteurs comportementaux et non-comportementaux sur le risque de cancer et de mortalité chez les femmes avant et après la ménopause. Nous avons également cherché à estimer leur impact relatif sur la population et à identifier les facteurs à forts pouvoirs prédictifs.

Résultats : Nos résultats suggèrent que le mode de vie a une influence modeste sur le risque de cancer et de mortalité lors de l'adhésion à une seule recommandation de santé publique. En revanche, elle est conséquente lors d'une adhésion conjointe à plusieurs recommandations. Les facteurs comportementaux jouent ainsi un rôle non négligeable dans la survenue de cancer et sur le risque de décès prématuré. Dans le cas du cancer du sein, ces facteurs influencent particulièrement le risque après la ménopause, tandis qu'avant la ménopause leur impact est plus faible que les facteurs qui ne relèvent pas du mode de vie ou de choix personnels. Ces observations sont retrouvées lorsque l'on cherche à prédire le risque de cancer du sein avant et après la ménopause. En effet, la prédiction du risque de cancer du sein en préménopause s'établit principalement à partir de facteurs non-comportementaux, alors que la prédiction du risque en postménopause est également déterminée par des facteurs comportementaux.

Conclusion : Nous avons montré que l'étiologie du cancer du sein diffère selon la nature de la tumeur, et en particulier selon le statut ménopausique des femmes. À tout âge, le mode de vie a une influence sur le risque de cancer et de mortalité prématurée, particulièrement après la ménopause lorsque leur impact est supérieur à celui des facteurs non-comportementaux. Ces résultats demandent, cependant, à être reproduits dans des études prospectives portant sur des femmes plus jeunes.

Mots-clés : Épidémiologie, femmes, ménopause, cohorte, mode de vie, comportement, obésité, tabagisme, alcool, traitements hormonaux, antécédent, cancer, cancer du sein, mortalité, prédiction.

ABSTRACT

Title: Behavioural and non-behavioural factors associated with cancer and mortality risk from the E3N cohort of French women

Background: Cancer is the second leading cause of mortality among women in France, and the leading cause of mortality among women aged between 35 and 84. Breast cancer is the most frequently diagnosed cancer, with 35% of cases among women in France in 2012. Multiple behavioural and non-behavioural factors have been associated with increases in cancer incidence and mortality. However, the literature about their combined impact is scarce. Regarding breast cancer, some risk factors differed according to the menopausal status, suggesting a different etiology between premenopausal and postmenopausal breast cancers.

Objectives: Data from the E3N prospective cohort of French women were used to evaluate the influence of behavioural and non-behavioural factors on cancer risk before and after the menopause and on mortality. In addition, we aimed at estimating their relative impact on the population and identifying factors with the highest predictive power.

Results: Our results suggest a modest influence of the lifestyle on cancer risk and mortality when adhering to only one public health recommendation. However, the influence is substantial with a combined adherence to several recommendations. Behavioural factors play a key role in the occurrence of cancer and mortality risk. Regarding breast cancer, these factors influence particularly the risk after the menopause, while before, their impact is lower than non-behavioural factors. These observations were retrieved when aiming at predicting breast cancer risk according to menopausal status. Prediction was established by non-behavioural factors in premenopause, while the prediction in postmenopause was driven by behavioural factors.

Conclusion: We have shown that the etiology of breast cancer differs according to the nature of the tumour, and particularly according to the menopausal status of women. Whatever the age, lifestyle influence the risk of cancer and mortality, especially after the menopause when their impact is higher than the non-behavioural factors' one. New results from prospective study on younger women are warranted to confirm the results.

Key words: Epidemiology, women, menopause, cohort, lifestyle, behaviour, obesity, smoking, alcohol, hormonal treatments, family history, cancer, breast cancer, mortality, prediction.

REMERCIEMENTS

En premier lieu, je voudrais remercier chaleureusement ma directrice de thèse, Madame Françoise Clavel-Chapelon pour m'avoir fait confiance dès mon arrivée. Merci pour ta gentillesse, ta patience et ta disponibilité tout au long de ce travail de thèse. Merci également de m'offrir l'opportunité de continuer l'aventure E3N au-delà de mon doctorat.

J'adresse également mes chaleureux remerciements à M. Guy Fagherazzi, co-encadrant de ce travail doctoral, pour m'avoir accordé sa confiance pour mon stage de fin d'études en allant me chercher à l'Ensaï et m'avoir encadrée au quotidien dès mon arrivée dans l'équipe. Merci pour ta disponibilité, ton écoute, ton soutien et tes conseils toujours pertinents.

J'adresse un sincère remerciement à M. Patrick Arveux et à M. David Cox pour avoir accepté d'être rapporteurs de ce travail et pour leurs précieux commentaires ayant permis d'améliorer ce présent manuscrit. J'exprime ma gratitude à M. Alexis Elbaz pour avoir accepté de présider mon jury de thèse et à Mme Martine Bellanger et M. Vivian Viallon qui ont été en charge d'examiner ce travail. Je tiens également à remercier l'école doctorale de santé publique Paris-Sud (ED420), son directeur Jean Bouyer et Audrey Bourgeois pour le suivi rigoureux de leurs doctorants, leurs précieux conseils et leur disponibilité, surtout dans les derniers préparatifs administratifs en vue de la préparation de la soutenance. Je remercie également la Fondation de France qui m'a permis de bénéficier d'une allocation doctorale pendant ces trois années de thèse.

Je voudrais remercier toutes les personnes qui ont participé aux différents projets issus de ce travail de thèse, du soutien méthodologique à la relecture critique des articles scientifiques. Un grand merci à Sylvie Mesrine pour son soutien bibliographique sans faille et ses conseils médicaux donnés à la vraie néophyte des hormones endogènes que je suis, à Guy Fagherazzi, Agnès Fournier et Gaëlle Gusto pour leur soutien méthodologique, à Jessica Pontary pour son dévouement dans la génération des tables de mortalité, à l'ensemble de mes co-auteurs pour leur relecture attentive et leurs précieux conseils. Je double les remerciements à Gaëlle Gusto pour m'avoir accordé tant de temps pour me permettre d'y voir plus clair dans la méthodologie du troisième article, ton aide a été très précieuse. Merci également à Lucile Offredo pour son travail effectué dans le cadre de son stage, qui a été la base d'une partie de ce mémoire.

J'adresse également de chaleureux remerciements à tous les membres de l'équipe E3N qui permettent de mettre de la lumière dans notre sous-sol. À tous ceux avec qui j'ai partagé mon bureau durant ces trois dernières années : Aurélie (M), Anne, Alice, Émilien, Céline, Suvi, Jessica, Sarah et Théotime, à Marie, notre maman à tous, à Lyan, Marie-Christine, Maryvonne, Pascale et Rafika sans qui l'équipe E3N ne serait pas ce qu'elle est !, à tous les collègues pour leur enthousiasme et leur

REMERCIEMENTS

bonne humeur : Agnès, Aurélie, Camille, Céline, Claire, Emmanuelle, Fabienne, Gaëlle, Ghizlane, Laure, Marina, Mathilde, Maxime, Nadia, Roselyn, Siham, Sylvie, Théotime, Thierno et Wilna, à nos voisins Statlife Émilien, Magali et Sylvie. J'adresse tout mon courage aux futures docteurs de l'équipe Claire, Antoine, Mathilde, Aurélie (félicitations encore !) et Isabelle, et particulièrement à Claire pour les nombreux « oh, mais la prochaine c'est toi ! » lors de ma soutenance de thèse. Une pensée supplémentaire pour Nadia, pour toutes nos discussions, pour son écoute et sa bonne humeur, je te souhaite d'être heureuse grâce aux changements futurs !

Je remercie ma famille qui m'a soutenue du début à la fin de cette aventure. Merci de tout cœur à mes parents pour m'avoir offert la possibilité de continuer dans cette voie, merci à ma sœur, mes frères, ma belle-sœur et ma tite mamie pour m'avoir encouragée et épaulée tout du long, et surtout vers la fin... Merci de votre soutien au quotidien. Merci à mes beaux-parents, ma belle-sœur et mon beau-frère pour avoir été présents, m'avoir soutenue et acceptée au sein de votre famille. Je remercie également mes ami(e)s Audrey, Stéphanie, Fanny, Éric, Gwen, Laure, Pierrick, qui m'ont soutenue et qui, je l'espère, ont compris mon absence de plusieurs mois cette année...

Je double et triple les remerciements à toutes les personnes qui ont bien voulu m'accorder de leur temps pour m'aider dans la rédaction de cet ouvrage, en relisant tout ou partie de ce présent travail. Merci à Françoise, Guy, Sylvie, Nadia, Gaëlle, Agnès et Sarah, à Émilien pour l'usage de son Antidote, à ma famille et amis qui ont pris le courage de se plonger dans ce pavé, malgré l'éloignement de nos parcours professionnels, et m'avoir prodiguée de précieux conseils : mes parents, mes beaux-parents, Audrey et Cécile. Je remercie tout particulièrement Maxime qui a ouvert le bal par ses relectures toujours très attentives et ses remarques pertinentes, merci également d'avoir répondu à mes innombrables questions sur les règles typographiques françaises, l'utilisation des sigles, les accords grammaticaux ou encore la mise en forme des en-têtes et pieds de page...

Enfin, je remercie mon crapulous débile de ne pas s'être endormi derrière mon dos quand je travaillais à l'ordinateur, de ne pas avoir marché sur le clavier de l'ordinateur lorsque je rédigeais ou de ne pas m'avoir réveillée au beau milieu de la nuit... (ou pas...). Je remercie très tendrement Maxime pour sa patience et sa présence au quotidien. Merci de ton soutien sans faille, merci d'avoir compris ma fatigue et d'avoir supporté mes sautes d'humeur, merci d'avoir su, par moment, me faire oublier ces derniers mois en nous évadant, me permettant de reprendre mon souffle (mâste ladda), merci pour tout ce que tu m'apportes. Une aventure personnelle se termine, une nouvelle aventure commune débute...

Je suis également reconnaissante à toutes les personnes qui auront le courage de parcourir ce mémoire de doctorat.

VALORISATION SCIENTIFIQUE

Article publié en lien avec le travail doctoral

Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. *Association between five lifestyle habits and cancer risk: results from the E3N cohort*. Cancer Prev.Res.(Phila) 2014;7:516-25.

Article soumis en lien avec le travail doctoral

Dartois L, Fagherazzi G, Baglietto L, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. *Proportion of premenopausal and postmenopausal breast cancers attributable to established risk factors: Estimates from the French E3N cohort*.

Dartois L, Gauthier E, Heitzmann J, Baglietto L, Michiels S, Mesrine S, Boutron-Ruault MC, Ragusa S, Clavel-Chapelon F, Fagherazzi G. *A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort*.

Articles en préparation

Dartois L, Offredo L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. *Association between five lifestyle habits and mortality: results from the E3N cohort*.

Autres publications (relecture critique)

McKenzie F, Ferrari P, Freisling H, Chajes V, Rinaldi S, de BJ, Dahm CC, Overvad K, Baglietto L, **Dartois L**, Dossus L, Lagiou P, Trichopoulos D, Trichopoulou A, Krogh V, Panico S, Tumino R, Rosso S, Bueno-de-Mesquita HB, May A, Peeters PH, Weiderpass E, Buckland G, Sanchez MJ, Navarro C, Ardanaz E, Andersson A, Sund M, Ericson U, Wirfalt E, Key TJ, Travis RC, Gunter M, Riboli E, Vergnaud AC, Romieu I. *Healthy lifestyle and risk of breast cancer among postmenopausal women in the European Prospective Investigation into Cancer and Nutrition cohort study*. Int J Cancer 2014.

Engeset D, Braaten T, Teucher B, Kuhn T, Bueno-de-Mesquita HB, Leenders M, Agudo A, Bergmann MM, Valanou E, Naska A, Trichopoulou A, Key TJ, Crowe FL, Overvad K, Sonestedt E, Mattiello A, Peeters PH, Wennberg M, Jansson JH, Boutron-Ruault MC, Dossus L, **Dartois L**, Li K, Barricarte A, Ward H, Riboli E, Agnoli C, Huerta JM, Sanchez MJ, Tumino R, Altzibar JM, Vineis P, Masala G, Ferrari P, Muller DC, Johansson M, Luisa RM, Tjønneland A, Olsen A, Olsen KS, Brustad M, Skeie G, Lund E. *Fish consumption and mortality in the European Prospective Investigation into Cancer and Nutrition cohort*. Eur.J Epidemiol. 2014.

Aleksandrova K, Pischon T, Jenab M, Bueno-de-Mesquita H, Fedirko V, Norat T, Romaguera D, Knuppel S, Boutron-Ruault MC, Dossus L, **Dartois L**, Kaaks R, Li K, Tjønneland A, Overvad K, Quiros J, Buckland G, Sanchez M, Dorronsoro M, Chirlaque MD, Barricarte A, Khaw KT, Wareham NJ, Bradbury KE, Trichopoulou A, Lagiou P, Trichopoulos D, Palli D, Krogh V, Tumino R, Naccarati A, Panico S, Siersema PD, Peeters P, Ljuslinder I, Johansson I, Ericson U, Ohlsson B, Weiderpass E, Skeie G, Borch K, Rinaldi S, Romieu I, Kong J, Gunter MJ, Ward HA, Riboli E, Boeing H. *Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study*. BMC.Med. 2014; 12(1):168.

Leenders M, Boshuizen HC, Ferrari P, Siersema PD, Overvad K, Tjønneland A, Olsen A, Boutron-Ruault MC, Dossus L, **Dartois L**, Kaaks R, Li K, Boeing H, Bergmann MM, Trichopoulou A, Lagiou P, Trichopoulos D, Palli D, Krogh V, Panico S, Tumino R, Vineis P, Peeters PH, Weiderpass E, Engeset D, Braaten T, Redondo ML, Agudo A, Sánchez MJ, Amiano P, Huerta JM, Ardanaz E, Drake I, Sonestedt E, Johansson I, Winkvist A, Khaw KT, Wareham NJ, Key TJ, Bradbury KE, Johansson M, Licaj I, Gunter MJ, Murphy N, Riboli E, Bueno-de-Mesquita HB. *Fruit and vegetable intake and cause-specific mortality in the EPIC study*. Eur J Epidemiol 2014; 29(9):639-652.

Forouhi NG, Koulman A, Sharp SJ, Imamura F, Kröger J, Schulze MB, Crowe FL, Huerta JM, Guevara M, Beulens JW, van Woudenberg GJ, Wang L, Summerhill K, Griffin JL, Feskens EJ, Amiano P, Boeing H, Clavel-Chapelon F, **Dartois L**, Fagherazzi G, Franks PW, Gonzalez C, Jakobsen MU, Kaaks R, Key TJ, Khaw KT, Kühn T, Mattiello A, Nilsson PM, Overvad K, Pala V, Palli D, Quirós JR, Rolandsson O, Roswall N, Sacerdote C, Sánchez MJ, Slimani N, Spijkerman AM, Tjønneland A, Tormo MJ, Tumino R, van der A DL, van der Schouw YT, Langenberg C, Riboli E, Wareham NJ. *Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study*. Lancet Diabetes Endocrinol 2014 [Epub ahead of print].

Ferrari P, Licaj I, Muller DC, Kragh Andersen P, Johansson M, Boeing H, Weiderpass E, Dossus L, **Dartois L**, Fagherazzi G, Bradbury KE, Khaw KT, Wareham N, Duell EJ, Barricarte A, Molina-Montes E, Sanchez CN, Arriola L, Wallström P, Tjønneland A, Olsen A, Trichopoulou A, Benetou V, Trichopoulos D, Tumino R, Agnoli C, Sacerdote C, Palli D, Li K, Kaaks R, Peeters P, Beulens JW, Nunes L, Gunter M, Norat T, Overvad K, Brennan P, Riboli E, Romieu I. *Lifetime alcohol use and overall and cause-specific mortality in the European Prospective Investigation into Cancer and nutrition (EPIC) study*. *BMJ Open* 2014; 4(7):e005245.

Kengne AP, Beulens JW, Peelen LM, Moons KG, van der Schouw YT, Schulze MB, Spijkerman AM, Griffin SJ, Grobbee DE, Palla L, Tormo MJ, Arriola L, Barengo NC, Barricarte A, Boeing H, Bonet C, Clavel-Chapelon F, **Dartois L**, Fagherazzi G, Franks PW, Huerta JM, Kaaks R, Key TJ, Khaw KT, Li K, Mühlenbruch K, Nilsson PM, Overvad K, Overvad TF, Palli D, Panico S, Quirós JR, Rolandsson O, Roswall N, Sacerdote C, Sánchez MJ, Slimani N, Tagliabue G, Tjønneland A, Tumino R, van der A DL, Forouhi NG, Sharp SJ, Langenberg C, Riboli E, Wareham NJ. *Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models*. *Lancet Diabetes Endocrinol* 2014; 2(1):19-29.

Steffen A, Sørensen TI, Knüppel S, Travier N, Sánchez MJ, Huerta JM, Quirós JR, Ardanaz E, Dorronsoro M, Teucher B, Li K, Bueno-de-Mesquita HB, van der A D, Mattiello A, Palli D, Tumino R, Krogh V, Vineis P, Trichopoulou A, Orfanos P, Trichopoulos D, Hedblad B, Wallström P, Overvad K, Halkjær J, Tjønneland A, Fagherazzi G, **Dartois L**, Crowe F, Khaw KT, Wareham N, Middleton L, May AM, Peeters PH, Boeing H. *Development and validation of a risk score predicting substantial weight gain over 5 years in middle-aged European men and women*. *PLoS One* 2013; 8(7):e67429.

Vergnaud AC, Romaguera D, Peeters PH, van Gils CH, Chan DS, Romieu I, Freisling H, Ferrari P, Clavel-Chapelon F, Fagherazzi G, **Dartois L**, Li K, Tikkanen K, Bergmann MM, Boeing H, Tjønneland A, Olsen A, Overvad K, Dahm CC, Redondo ML, Agudo A, Sánchez MJ, Amiano P, Chirlaque MD, Ardanaz E, Khaw KT, Wareham NJ, Crowe F, Trichopoulou A, Orfanos P, Trichopoulos D, Masala G, Sieri S, Tumino R, Vineis P, Panico S, Bueno-de-Mesquita HB, Ros MM, May A, Wirfält E, Sonestedt E, Johansson I, Hallmans G, Lund E, Weiderpass E, Parr CL, Riboli E, Norat T. *Adherence to the World Cancer Research Fund/American Institute for Cancer Research guidelines and risk of death in Europe: results from the European Prospective Investigation into Nutrition and Cancer cohort study*. *Am J Clin Nutr* 2013; 97(5):1107-1120.

May AM, Romaguera D, Travier N, Ekelund U, Bergmann MM, Kaaks R, Teucher B, Steffen A, Boeing H, Halkjaer J, Tjønneland A, Jakobsen MU, Overvad K, **Dartois L**, Fagherazzi G, Boutron-Ruault MC, Quirós JR, Agudo A, Gonzalez C, Sánchez MJ, Amiano P, Huerta JM, Ardanaz E, Wareham NJ, Crowe FL, Naska A, Orfanos P, Trichopoulou A, Palli D, Agnoli C, Tumino R, Vineis P, Panico S, Bueno-de-Mesquita HB, Verschuren M, Drake I, Sonestedt E, Braaten T, Rinaldi S, Romieu I, Slimani N, Norat T, Riboli E, Peeters PH. *Combined impact of lifestyle factors on prospective change in body weight and waist circumference in participants of the EPIC-PANACEA study*. PLoS One 2012; 7(11):e50712.

Chuang SC, Norat T, Murphy N, Olsen A, Tjønneland A, Overvad K, Boutron-Ruault MC, Perquier F, **Dartois L**, Kaaks R, Teucher B, Bergmann MM, Boeing H, Trichopoulou A, Lagiou P, Trichopoulos D, Grioni S, Sacerdote C, Panico S, Palli D, Tumino R, Peeters PH, Bueno-de-Mesquita B, Ros MM, Brustad M, Åsli LA, Skeie G, Quirós JR, González CA, Sánchez MJ, Navarro C, Ardanaz Aicua E, Dorransoro M, Drake I, Sonestedt E, Johansson I, Hallmans G, Key T, Crowe F, Khaw KT, Wareham N, Ferrari P, Slimani N, Romieu I, Gallo V, Riboli E, Vineis P. *Fiber intake and total and cause-specific mortality in the European Prospective Investigation into Cancer and Nutrition cohort*. Am J Clin Nutr 2012; 96(1):164-174.

Communications scientifiques sous forme de poster

Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. *Impact of five lifestyle habits on cancer risk: results from the E3N cohort*. Présenté à la conférence “European Congress of Epidemiology 2013”, Aarhus, Danemark, 11-14 août 2013.

Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. *Impact of physical activity, body mass index and three other lifestyle characteristics on cancer risk: results from the E3N cohort*. Présenté à la conférence “Obesity, Physical Activity and Cancer (IASO/WCRF)”, Londres, Royaume-Uni, 16-17 avril 2013.

Communications scientifiques sous forme de présentation orale

Dartois L. *E3N – Étude épidémiologique auprès de femmes de l'Éducation nationale. Les principaux résultats sur le risque de cancer chez les femmes*. Assemblée Générale de la Ligue contre le Cancer, comité du Gers. 24/04/2014.

Activités d'enseignement pendant le doctorat

Contrat doctoral – mission d'enseignement à l'Université de Paris Dauphine (2011-2013) : Cours et travaux dirigés de Probabilité/Statistique en deuxième année de licence, département LSO (Licence Sciences des Organisations).

Contrat doctoral – mission d'enseignement à l'Université de Paris Dauphine (2011-2012) : Cours et travaux dirigés de Probabilité en première année de licence, département LSO (Licence Sciences des Organisations).

Encadrement d'étudiants

Lucile Offredo : stage de Master 2 Modélisation en Pharmacologie Clinique et Épidémiologie « Association entre cinq facteurs de mode de vie et le risque de mortalité, toutes causes et par causes spécifiques » (Avril-Septembre 2014).

Pauline Jestin : stage de Master 1 Santé Publique « Observation des différences géographiques de l'incidence de cancer, de mortalité et de certaines caractéristiques individuelles dans la cohorte E3N » (Février-Avril 2012).

TABLE DES MATIÈRES

Avant-propos	3
Résumé	5
Abstract.....	6
Remerciements.....	7
Valorisation scientifique.....	9
Table des matières	15
Liste des tableaux.....	21
Liste des figures	25
Liste des équations.....	27
Liste des abréviations	29
Introduction générale.....	31
Partie 1 Épidémiologie des cancers	31
1.1. Processus biologique.....	31
1.1.a. D'où provient le mot « cancer » ?	31
1.1.b. La cancérogenèse.....	31
1.2. Épidémiologie de l'incidence de cancer en France et dans le monde	33
1.2.a. L'incidence du cancer en 2012	33
1.2.b. Évolution de l'incidence du cancer entre 1980 et 2012	35
1.2.c. Les différents sous-types de cancer du sein.....	37
1.2.d. Les facteurs de risque de cancer du sein	38
1.2.e. Les facteurs de risque des autres localisations de cancer.....	43
Partie 2 Épidémiologie de la mortalité en France	48
2.1. Les principales causes de mortalité chez la femme	48
2.1.a. Les principales causes de mortalité en 2011	48
2.1.b. Évolution des principales causes de mortalité de 1968 à 2011	50
2.2. La mortalité par maladie de l'appareil circulatoire chez la femme	52
2.2.a. Le taux de mortalité par maladie de l'appareil circulatoire en 2011	52
2.2.b. Évolution du taux de mortalité par maladie de l'appareil circulatoire entre 1979 et 2011	53
2.3. La mortalité par cancer chez la femme	53
2.3.a. Le taux de mortalité par cancer en 2012.....	54
2.3.b. Évolution du taux de mortalité par cancer entre 1980 et 2012	54
2.3.c. Cas particuliers des cancers du sein et du poumon.....	55
Partie 3 État de l'art.....	57
3.1. L'association entre le mode de vie et le risque global de cancer	57
3.1.a. Tabagisme actif et risque de cancer	57
3.1.b. Anthropométrie et risque de cancer	58
3.1.c. Consommation d'alcool et risque de cancer.....	58
3.1.d. Consommation de fruits et légumes et risque de cancer	59

TABLE DES MATIÈRES

3.1.e. Niveau d'activité physique et risque de cancer.....	59
3.1.f. Combinaison de ces cinq facteurs de mode de vie et risque de cancer.....	60
3.2. Le poids relatif des différents facteurs de risque de cancer du sein.....	61
3.2.a. Les facteurs comportementaux	62
3.2.b. Les facteurs non-comportementaux.....	62
3.2.c. Prise en considération conjointe des facteurs comportementaux et non-comportementaux	62
3.3. Les scores prédictifs de cancer du sein	63
3.3.a. Origine des scores prédictifs en santé	63
3.3.b. Les scores de prédiction du risque de cancer du sein	63
3.4. L'association entre le mode de vie et la mortalité.....	64
3.4.a. Tabagisme et risque de mortalité	64
3.4.b. Anthropométrie et risque de mortalité.....	65
3.4.c. Consommation d'alcool et risque de mortalité	65
3.4.d. Consommation de fruits et légumes et risque de mortalité.....	66
3.4.e. Niveau d'activité physique et risque de mortalité	66
3.4.f. Combinaison de ces cinq facteurs de mode de vie et risque de mortalité	67
Partie 4 Justification du travail doctoral.....	68
Matériel.....	71
Partie 1 La cohorte E3N-EPIC	71
1.1. Présentation de la cohorte E3N et objectifs de l'étude	71
1.2. Le recueil des données.....	71
1.2.a. Le commencement de l'étude.....	71
1.2.b. Le suivi des participantes.....	72
1.2.c. Les données alimentaires.....	74
1.2.d. Les données biologiques.....	76
1.2.e. Les données de la MGEN.....	76
1.3. EPIC : une dimension européenne.....	77
1.4. Perspectives : d'E3N à E4N.....	78
Partie 2 Recueil de l'information sur le cancer	79
2.1. Les cas de cancer	79
2.2. Classification des cancers	79
2.2.a. La CIM (Classification internationale des maladies)	79
2.2.b. La CIM-O (Classification internationale des maladies oncologiques).....	80
2.2.c. La classification TNM (<i>Tumor, Node, Metastasis</i>) des tumeurs malignes.....	82
2.2.d. Classification des localisations de cancer utilisée dans le présent travail.....	82
2.3. Codification des récepteurs hormonaux des cancers du sein	84
2.3.a. Le grade Scarff-Bloom-Richardson (SBR).....	84
2.3.b. Les récepteurs hormonaux.....	84
2.3.c. La protéine HER2.....	84
Partie 3 Recueil de l'information sur la mortalité	86
3.1. La connaissance des décès	86
3.2. Demande d'informations complémentaires sur les décès	86

3.3.	Recueil des causes de décès auprès du CépiDc.....	86
3.3.a.	Présentation du CépiDc	86
3.3.b.	Le certificat médical de décès	86
3.3.c.	La qualité de la certification des décès	88
3.3.d.	Le circuit administratif du certificat de décès.....	89
3.3.e.	Du certificat papier au certificat électronique	90
3.3.f.	Codification des décès et extraction de la cause initiale de décès.....	90
3.3.g.	Appariement des données E3N et des données du CépiDc.....	92
3.4.	Classification des causes de décès utilisée dans le présent travail	92
Partie 4 Recueil des données d'exposition.....		94
4.1.	Les données alimentaires.....	94
4.1.a.	La consommation d'alcool.....	94
4.1.b.	La consommation de fruits et légumes	95
4.1.c.	L'apport en vitamine D alimentaire.....	96
4.1.d.	Les profils alimentaires	97
4.1.e.	L'énergie consommée totale	97
4.2.	Les données anthropométriques	98
4.2.a.	Anthropométrie à la naissance et à la puberté	98
4.2.b.	Anthropométrie au cours du suivi	98
4.3.	Autres données de mode de vie	99
4.3.a.	Le statut tabagique.....	99
4.3.b.	Le niveau d'activité physique	99
4.4.	Les données des facteurs reproductifs endogènes	100
4.4.a.	L'âge à la puberté.....	100
4.4.b.	Le statut ménopausique.....	100
4.4.c.	L'âge à la ménopause	100
4.4.d.	Le nombre d'enfants, l'âge à la première grossesse à terme et la durée d'allaitement cumulée.....	101
4.5.	Les traitements hormonaux exogènes.....	101
4.5.a.	Les contraceptifs oraux œstro-progestatifs.....	101
4.5.b.	Les progestatifs seuls	102
4.5.c.	Les traitements hormonaux de la ménopause.....	103
4.6.	Autres caractéristiques personnelles ou familiales.....	104
4.6.a.	L'âge	104
4.6.b.	Le niveau d'étude.....	105
4.6.c.	Les antécédents personnels de maladie bénigne du sein	105
4.6.d.	Les antécédents personnels de maladie chronique (cancer, maladie cardiovasculaire et diabète)	105
4.6.e.	Les antécédents familiaux de cancer.....	105
4.6.f.	Le statut marital	106
4.6.g.	La zone de résidence.....	106
4.6.h.	Les antécédents de mammographie.....	106
Méthodes statistiques.....		107
Partie 1 Analyses descriptives		107
1.1.	Les statistiques descriptives utilisées	107

1.2. Comparaison des statistiques descriptives entre deux échantillons	107
1.2.a. Comparaison de deux moyennes	107
1.2.b. Comparaison de deux proportions	108
Partie 2 Analyses statistiques : le modèle de régression de Cox.....	109
2.1. Les analyses de survie	109
2.2. Le modèle de régression semi-paramétrique de Cox	109
2.3. Les hypothèses d'application du modèle de régression de Cox	111
2.3.a. L'hypothèse des risques proportionnels.....	111
2.3.b. L'hypothèse de log-linéarité des covariables	113
2.4. Le rapport de risques ou <i>hazard ratio</i>	113
2.5. La fonction de risque instantané de base.....	114
2.6. Les variables dépendantes du temps	115
2.7. L'âge comme échelle de temps.....	116
2.8. Le test de tendance.....	116
2.9. L'effet cohorte	116
2.10. L'analyse en risques compétitifs	118
2.11. La prédiction à cinq ans du risque	119
2.11.a. Estimation du risque prédit à cinq ans	119
2.11.b. Validation d'un modèle prédictif.....	120
2.11.c. Méthodologie employée pour l'obtention du modèle de prédiction final.....	125
2.11.d. Comparaison de deux scores prédictifs.....	126
2.11.e. Deux autres méthodes pour la construction d'un score prédictif : le modèle des plus proches voisins et le modèle de Gail.....	128
Partie 3 Estimation des fractions attribuables	130
3.1. Définition	130
3.2. Formalisation mathématique	130
3.3. Erreurs à éviter	131
3.4. Interprétation	132
Partie 4 Gestion des valeurs manquantes	132
4.1. Gestion des valeurs manquantes de l'exposition principale.....	132
4.2. Gestion des valeurs manquantes des variables d'ajustement	132
Résultats	133
Partie 1 Influence du mode de vie sur le risque de cancer	133
1.1. Introduction	133
1.2. Matériel et méthodes.....	133
1.2.a. Population d'étude	133
1.2.b. Collecte des données et construction de l'indice de santé.....	134
1.2.c. Définition des cas de cancer	135
1.2.d. Méthodologie statistique	135
1.3. Résultats.....	136
1.3.a. Caractéristiques de la population d'analyse.....	136

1.3.b.	Association entre les comportements individuels, leur combinaison et le risque de cancer, toutes localisations confondues et par site spécifique	138
1.3.c.	Proportions de cas de cancer attribuables aux comportements individuels de mode de vie ou à leur combinaison.....	145
1.4.	Discussion	147
1.4.a.	Confrontation à la littérature	148
1.4.b.	Limites de l'étude.....	150
1.4.c.	Conclusion de l'étude.....	150
Partie 2	Estimation du poids relatif de différents facteurs de risque établis de cancer du sein	151
2.1.	Introduction	151
2.2.	Matériel et méthodes.....	151
2.2.a.	Population d'étude	151
2.2.b.	Facteurs de risque comportementaux et non-comportementaux	152
2.2.c.	Méthodologie statistique	152
2.3.	Résultats.....	153
2.3.a.	Caractéristiques de la population d'analyse.....	153
2.3.b.	Proportions de cas de cancer du sein attribuables à un facteur de risque.....	154
2.3.c.	Proportions de cas de cancer du sein attribuables à la combinaison des facteurs de risque.....	156
2.3.d.	Proportions de cas de cancer du sein en préménopause attribuables aux facteurs de risque selon le statut RO	157
2.3.e.	Proportions de cas de cancer du sein en postménopause attribuables aux facteurs de risque selon le statut RO et l'expression de la protéine HER2	157
2.4.	Discussion	158
2.4.a.	Confrontation à la littérature	158
2.4.b.	Limites de l'étude.....	161
2.4.c.	Conclusion.....	161
Partie 3	Prédiction à cinq ans du risque de cancer du sein	162
3.1.	Introduction	162
3.2.	Matériel et méthodes.....	162
3.2.a.	Population d'étude	162
3.2.b.	Collecte des données.....	163
3.2.c.	Méthodologie statistique	163
3.3.	Résultats.....	164
3.3.a.	Caractéristiques de la population d'analyse.....	164
3.3.b.	Sélection des modèles de prédiction du risque de cancer du sein en préménopause et en postménopause.....	165
3.3.c.	Validation des meilleurs modèles de prédiction du risque de cancer du sein en préménopause et en postménopause	168
3.3.d.	Comparaison avec d'autres modèles de prédiction	174
3.4.	Discussion	178
3.4.a.	Confrontation à la littérature	178
3.4.b.	Limites de l'étude.....	179
3.4.c.	Conclusion.....	180

Partie 4 Influence du mode de vie sur le risque de mortalité	181
4.1. Introduction	181
4.2. Matériel et méthodes.....	181
4.2.a. Population d'étude	181
4.2.b. Collecte des données et construction de l'indice de santé.....	182
4.2.c. Méthodologie statistique	182
4.3. Résultats.....	183
4.3.a. Caractéristiques de la population d'analyse.....	183
4.3.b. Association entre les comportements individuels, leur combinaison et le risque de décès, toutes causes confondues et par cause initiale spécifique.....	185
4.3.c. Association entre les comportements individuels, leur combinaison et le risque de décès par localisations spécifiques de cancer.....	189
4.3.d. Proportions de décès attribuables aux comportements individuels de mode de vie et à leur combinaison	194
4.4. Discussion	196
4.4.a. Confrontation à la littérature	197
4.4.b. Conclusion.....	197
Conclusion générale	199
Partie 1 Principaux résultats	199
Partie 2 Perspectives de recherche.....	200
Partie 3 Impact en terme de santé publique	202
Glossaire	205
Annexes	217
Annexe A : Groupes alimentaires considérés dans l'analyse en composantes principales pour la création des profils alimentaires	217
Annexe B : Jeu de données simulées pour l'illustration des performances d'un modèle prédictif.....	219
Annexe C : Caractéristiques des populations d'analyse.....	220
Annexe D : Résultats complémentaires pour l'analyse de l'association entre le mode de vie et le risque de cancer	231
Annexe E : Résultats complémentaires pour l'analyse des poids relatifs des facteurs de risque de cancer du sein en préménopause et en postménopause.....	233
Annexe F : « <i>Association between five lifestyle habits and cancer risk : Results from the E3N cohort</i> » - Article publié dans le journal « Cancer Prevention Research »	236
Annexe G : « <i>Proportion of premenopausal and postmenopausal breast cancers attributable to established risk factors: Estimates from the E3N-EPIC cohort study</i> » - Article en revue au journal « The BMJ »	246
Annexe H : « <i>A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort</i> » - Article en revue au journal « Breast Cancer Research and Treatment »	277
Références	309

LISTE DES TABLEAUX

Tableau 1 : Interprétation de l'IMC par l'OMS.....	42
Tableau 2 : Taux de réponse et de participation aux questionnaires E3N, cohorte E3N (N = 98 995).....	74
Tableau 3 : Évolution de la codification des tumeurs malignes du sein de la CIM 8 à la CIM 10.....	80
Tableau 4 : Groupes histopathologiques des tumeurs selon la CIM-O.....	81
Tableau 5 : Code de comportement évolutif selon la CIM-O.....	81
Tableau 6 : Code pour le grade histopathologique et la différenciation selon la CIM-O.....	81
Tableau 7 : Classification anatomique d'une tumeur selon la TNM.....	82
Tableau 8 : Stade d'une tumeur selon la TNM.....	82
Tableau 9 : Classification des localisations de cancer utilisée dans le présent travail.....	83
Tableau 10 : Méthode de calcul du grade SBR.....	84
Tableau 11 : Calcul du score HER2.....	85
Tableau 12 : Classification des causes de décès utilisée dans le présent travail.....	93
Tableau 13 : Définition du placement des nœuds pour la modélisation par spline restreint.....	115
Tableau 14 : Tableau de contingence entre la classe prédite et la classe réelle.....	121
Tableau 15 : Exemple de calcul des rapports E/O par intervalles de quartiles du score prédit pour un jeu de données simulées.....	124
Tableau 16 : Tableau de concordance pour n individus classés par tertiles de valeurs de deux modèles prédictifs.....	127
Tableau 17 : Construction de l'indice de santé à partir de cinq facteurs de mode de vie.....	134
Tableau 18 : Localisations des cancers diagnostiqués dans la population d'analyse, cohorte E3N (N = 64 732).....	136
Tableau 19 : Coefficients de corrélation de Pearson entre les scores individuels, deux à deux, cohorte E3N (N = 64 732).....	138
Tableau 20 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de cancer, toutes localisations confondues et par site spécifique, cohorte E3N (N = 64 732).....	139
Tableau 21 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de cancer, toutes localisations confondues et par site spécifique, cohorte E3N (N = 64 732).....	143
Tableau 22 : Proportions [IC 95 %] de cas de cancer, toutes localisations confondues et par site spécifique, attribuables à la non-adhésion aux recommandations individuelles de santé publique et à leur combinaison, cohorte E3N (N = 64 732).....	146
Tableau 23 : Types de cancer du sein diagnostiqués dans la population d'analyse, cohorte E3N (N = 67 634).....	153
Tableau 24 : Différences de moyennes et de pourcentages, pour les variables analysées, observées entre les échantillons d'apprentissage et de validation, cohorte E3N (N = 64 995).....	165
Tableau 25 : Performances des modèles à un seul prédicteur pour la prédiction du risque de cancer du sein invasif en préménopause et en postménopause, cohorte E3N (N = 51 821).....	166

LISTE DES TABLEAUX

Tableau 26 : Meilleures combinaisons optimisant l’AUC et le coefficient de régression associé au PI, échantillon d’apprentissage, cohorte E3N (N = 51 821).....	167
Tableau 27 : Performances des modèles sélectionnés, échantillon de validation, cohorte E3N (N = 13 174)...	168
Tableau 28 : Associations entre les scores prédictifs préménopausique et postménopausique, catégorisés selon les quartiles de distribution, et le risque de cancer du sein invasif, cohorte E3N (N = 64 995).....	169
Tableau 29 : Sélection des prédicteurs entrant dans la construction du score de prédiction du risque de cancer du sein pour le modèle de Cox, l’algorithme des plus proches voisins et le modèle logistique de Gail.....	174
Tableau 30 : Prédiction à cinq ans du risque de cancer du sein invasif, à partir du modèle de Cox, de l’algorithme des plus proches voisins et du modèle logistique de Gail, pour dix femmes hypothétiques.	176
Tableau 31 : Comparaisons des scores de prédiction à cinq ans du risque de cancer du sein invasif, construits à partir du modèle de Cox, de l’algorithme des plus proches voisins et du modèle logistique de Gail, échantillon de validation, cohorte E3N (N = 13 174).....	177
Tableau 32 : Performances des scores de prédiction à cinq ans du risque de cancer du sein invasif, construits à partir du modèle de Cox, de l’algorithme des plus proches voisins et du modèle logistique de Gail, échantillon de validation, cohorte E3N (N = 13 174).....	177
Tableau 33 : Causes initiales des décès survenus dans la population d’analyse, cohorte E3N (N = 66 721).	184
Tableau 34 : HR [IC 95 %] de l’association entre les cinq comportements de mode de vie et le risque de décès, toutes causes confondues (N = 66 721) et par cause initiale spécifique (N = 66 560), cohorte E3N.	186
Tableau 35 : HR [IC 95 %] de l’association entre l’indice de santé et le risque de décès, toutes causes confondues (N = 66 721) et par cause spécifique (N = 66 560), cohorte E3N.....	188
Tableau 36 : HR [IC 95 %] de l’association entre les cinq comportements de mode de vie et le risque de décès par cancer, selon la localisation de la tumeur, cohorte E3N (N = 66 560).	190
Tableau 37 : HR [IC 95 %] de l’association entre l’indice de santé et le risque de décès par cancer, selon la localisation de la tumeur, cohorte E3N (N = 66 560).	193
Tableau 38 : Proportions [IC 95 %] de décès, toutes causes confondues (N = 66 721) et par cause spécifique (N = 66 560), attribuables à la non-adhésion aux recommandations individuels de santé publique et à leur combinaison, cohorte E3N.	195
Tableau 39 : Groupes alimentaires considérés dans l’analyse en composantes principales pour la création des profils alimentaires. [Lien vers la première citation du tableau dans le texte]	217
Tableau 40 : Exemple de calcul de la sensibilité et de la spécificité à partir d’un jeu de données simulées. [Lien vers la première citation du tableau dans le texte]	219
Tableau 41 : Caractéristiques, au début de l’étude sur l’association entre l’indice de santé et le risque de cancer, de la population d’analyse, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]	220

Tableau 42 : Caractéristiques, au début de l'étude sur le poids relatif des facteurs de risque de cancer du sein en préménopause et en postménopause, de la population d'analyse et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]	222
Tableau 43 : Caractéristiques, au début de l'étude sur la prédiction du risque de cancer du sein en préménopause et en postménopause, de l'échantillon d'apprentissage et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 51 821). [Lien vers la première citation du tableau dans le texte]	225
Tableau 44 : Caractéristiques, au début de l'étude sur la prédiction du risque de cancer du sein en préménopause et en postménopause, de l'échantillon de validation et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 13 174). [Lien vers la première citation du tableau dans le texte]	227
Tableau 45 : Caractéristiques, au début de l'étude sur l'association entre l'indice de santé et le risque de mortalité, de la population d'analyse, cohorte E3N (N = 66 721). [Lien vers la première citation dans le texte].....	229
Tableau 46 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de mélanome et de cancer cutané non basal non mélanome, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]	231
Tableau 47 : Proportions [IC 95 %] de cas de mélanome et de cancer cutané non basal et non mélanome attribuables à la non-adhésion aux recommandations individuelles de santé publique et à leur combinaison, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]	231
Tableau 48 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de mélanome et de cancer cutané non basal non mélanome, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]	232
Tableau 49 : Proportions [IC 95 %] de cas de cancer du sein en préménopause RO+ et RO- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]	233
Tableau 50 : Proportions [IC 95 %] de cas de cancer du sein en postménopause HER2+, HER2-, RO+ et RO- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]	234
Tableau 51 : Proportions [IC 95 %] de cas de cancer du sein en postménopause RO+/HER2+, RO+/HER2-, RO-/HER2+ et RO-/HER2- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]	235

LISTE DES FIGURES

Figure 1 : Représentation des différentes mutations génétiques les plus fréquemment rencontrées.	32
Figure 2 : Représentation de l'évolution d'une tumeur : de la mutation jusqu'au cancer invasif.	33
Figure 3 : Taux d'incidence et de mortalité par cancer, standardisés sur la structure d'âge de la population mondiale, pour 100 000 femmes en 2012 en France métropolitaine.	34
Figure 4 : Évolution du taux d'incidence de cancer pour 100 000 femmes, standardisé sur la structure d'âge de la population mondiale, pour certains pays d'Europe (à gauche), d'Asie (au milieu) et d'Amérique et d'Océanie (à droite) de 1975 à 2010.	35
Figure 5 : Évolution du taux d'incidence pour 100 000 femmes en France métropolitaine, standardisé sur la structure d'âge de la population mondiale, pour le cancer, toutes localisations confondues (en haut), du sein (en haut) et d'autres localisations (en bas) entre 1980 et 2012.	36
Figure 6 : Action biologique de l'hormonothérapie sur les cellules mammaires cancéreuses.	37
Figure 7 : Représentation de mammographies pour des seins de densité faible (gauche) à élevée (droite).	43
Figure 8 : Représentation d'une coupe transversale de l'épiderme.	46
Figure 9 : Taux de mortalité brut pour 100 000 femmes en 2011 en France pour les principales causes de mortalité.	49
Figure 10 : Part des principales causes de mortalité des femmes, en France, en 2011 parmi l'ensemble des décès en fonction de l'âge au décès (<75, [75 ; 84], >84).	50
Figure 11 : Évolution de la part des principales causes de décès chez les femmes en France de 1979 à 2011.	51
Figure 12 : Évolution de la part des décès par maladie de l'appareil circulatoire et par cancer chez les femmes en France de 1979 à 2011 selon la classe d'âge.	51
Figure 13 : Évolution de la part des décès des différentes formes de maladie de l'appareil circulatoire parmi l'ensemble des décès par maladie de l'appareil circulatoire, chez les femmes en France, de 1979 à 2011.	53
Figure 14 : Évolution du taux de mortalité par cancer pour 100 000 femmes, standardisé sur la structure d'âge de la population mondiale, pour certains pays d'Europe (à gauche), d'Asie (au milieu) et d'Amérique et d'Océanie (à droite) de 1975 à 2010.	54
Figure 15 : Évolution du taux de mortalité pour 100 000 femmes en France métropolitaine, standardisé sur la structure d'âge de la population mondiale, pour le cancer, toutes localisations confondues (en haut), du sein (en haut) et d'autres localisations de cancer (en bas) entre 1980 et 2012.	55
Figure 16 : Chronologie des questionnaires E3N et recueil des différentes données.	73
Figure 17 : Première page des questionnaires Q1 (à gauche) et Q6 (à droite).	74
Figure 18 : Exemple de page des parties quantitative (à gauche) et qualitative (à droite) du questionnaire alimentaire Q3 (1993).	75
Figure 19 : Pays participants à l'étude EPIC (<i>European Prospective Investigation into Cancer and Nutrition</i>).	77
Figure 20 : E4N, étude des enfants d'E3N, une cohorte familiale.	78
Figure 21 : Certificat de décès traditionnel (format papier).	88

LISTE DES FIGURES

Figure 22 : Circuit d'un certificat de décès traditionnel.....	89
Figure 23 : Questionnaire alimentaire de fréquence et de quantité de boissons alcoolisées consommées.....	94
Figure 24 : Photographies d'aide au remplissage du questionnaire alimentaire. Volumes standards d'eau, de vin, de jus de fruits et de sodas (en haut), de bière (au milieu), d'apéritifs et de digestifs (en bas).....	95
Figure 25 : Questionnaires quantitatifs (à gauche) et qualitatifs (à droite) utilisés pour estimer la consommation de fruits et légumes.....	95
Figure 26 : Photographies d'aide au remplissage du questionnaire alimentaire. De gauche à droite : portions standards de salade verte, de crudités et légumes à la vinaigrette, de légumes cuits (hors légumes secs) et de fruits.....	96
Figure 27 : Dose journalière moyenne de rayonnement solaire perçu par département français.....	97
Figure 28 : Silhouettes de Sørensen (féminines).....	98
Figure 29 : Photographies des conditionnements des contraceptifs oraux œstro-progestatifs accompagnants le questionnaire Q2 (1992).....	102
Figure 30 : Photographies des conditionnements des progestatifs seuls accompagnants le questionnaire Q2 (1992).....	103
Figure 31 : Photographies des conditionnements des THM accompagnants le questionnaire Q2 (1992).....	104
Figure 32 : Répartition en zones de résidence des départements français.....	106
Figure 33 : Risques cumulés d'incidence (en haut) et de mortalité (en bas) chez les femmes âgées de 0 à 74 ans en France métropolitaine entre 1980 et 2012 selon la cohorte de naissance (1920 à 1950).....	117
Figure 34 : Représentation de la discrimination et de la calibration d'un score prédictif.....	120
Figure 35 : Exemple de représentation d'une courbe ROC à partir d'un jeu de données simulées.....	123
Figure 36 : Exemple d'un diagramme de calibration pour un jeu de données simulées.....	124
Figure 37 : Distribution des cinq scores individuels dans la population d'analyse, cohorte E3N (N = 64 732)...	137
Figure 38 : Proportions de cas de cancer du sein en préménopause et en postménopause attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634).....	155
Figure 39 : Diagrammes de calibration des scores préménopausique et postménopausique dans l'échantillon d'apprentissage, cohorte E3N (N = 51 821).....	170
Figure 40 : Diagrammes de calibration des scores préménopausique et postménopausique dans l'échantillon de validation, cohorte E3N (N = 13 174).....	171
Figure 41 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en préménopause dans l'échantillon d'apprentissage, cohorte E3N (N = 23 362).....	172
Figure 42 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en postménopause dans l'échantillon d'apprentissage, cohorte E3N (N = 28 459).....	172
Figure 43 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en préménopause dans l'échantillon de validation, cohorte E3N (N = 5 843).....	173
Figure 44 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en postménopause dans l'échantillon de validation, cohorte E3N (N = 7 331).....	173

LISTE DES ÉQUATIONS

Équation 1 : Statistique de test associée à la comparaison de la moyenne d'une variable au sein de deux échantillons indépendants.	108
Équation 2 : Statistique de test associée à la comparaison de deux proportions de deux échantillons indépendants.	108
Équation 3 : Définition de la fonction de survie.	109
Équation 4 : Définition de la fonction de risque instantané.	109
Équation 5 : Relations directes entre les fonctions de survie, de risque instantané et de risque cumulé.	109
Équation 6 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox.	110
Équation 7 : Formalisation de la transformée logarithmique du risque instantané par le modèle de régression de Cox.	110
Équation 8 : Fonction de vraisemblance partielle d'un modèle de régression de Cox.	110
Équation 9 : Rapport de risques entre deux individus de profils différents.	111
Équation 10 : Expressions de $S(t)$ et de $\ln(-\ln(S(t)))$	111
Équation 11 : Équation sous-jacente à l'hypothèse de log-linéarité d'une covariable.	113
Équation 12 : Définition de l' <i>hazard ratio</i> entre deux modalités d'une covariable.	114
Équation 13 : Estimateur de Breslow pour la fonction de risque cumulée de base.	114
Équation 14 : Modélisation par spline de la fonction de risque cumulée de base.	114
Équation 15 : Formulation du critère d'information d'Akaike.	115
Équation 16 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox en présence de variables explicatives dépendantes du temps.	116
Équation 17 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox dans un modèle stratifié.	117
Équation 18 : Définition de la fonction de risque instantané en présence de risques compétitifs.	118
Équation 19 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox en présence de risques compétitifs.	119
Équation 20 : Rapport de risques entre deux individus de profils différents en présence de risques compétitifs.	119
Équation 21 : Estimation de la survie et du risque à cinq ans.	119
Équation 22 : Définition de la sensibilité d'un score prédictif.	122
Équation 23 : Définition de la spécificité d'un score prédictif.	122
Équation 24 : Définition du coefficient de corrélation de Pearson entre deux variables quantitatives.	126
Équation 25 : Définition du coefficient de concordance entre deux variables quantitatives.	127
Équation 26 : Formule de la distance Euclidienne utilisée dans l'algorithme des plus proches voisins.	128
Équation 27 : Modélisation du modèle de régression logistique.	129

LISTE DES ÉQUATIONS

Équation 28 : Estimation du risque relatif entre deux profils à partir d'une modélisation par modèle de régression logistique.	129
Équation 29 : Définition de la fraction attribuable.	130
Équation 30 : Formulation mathématique de la fraction attribuable proposée par Levin, 1953.	131
Équation 31 : Formulation mathématique de la fraction attribuable proposée par Miettinen, 1974.	131
Équation 32 : Formulation mathématique de la fraction attribuable proposée par Bruzzi, 1985.	131
Équation 33 : Modélisation de la fonction de survie de base pour le score préménopausique.	171
Équation 34 : Modélisation de la fonction de survie de base pour le score postménopausique.	172

LISTE DES ABRÉVIATIONS

ADN	Acide désoxyribonucléique
AIC	Critère d'information d'Akaike (<i>Akaike information criteria</i>)
AUC	Aire sous la courbe ROC (<i>Area under the ROC curve</i>)
AVC	Accident vasculaire cérébral
Bac	Baccalauréat
BMB	Banque de matériel biologique
BRCA	<i>Breast Cancer</i> (= cancer du sein)
CESP	Centre de recherche en épidémiologie et santé des populations
CIM	Classification internationale des maladies
CIM-O	Classification internationale des maladies oncologiques
CIRC	Centre international de recherche sur le cancer
CNIL	Commission nationale de l'informatique et des libertés
CNR	Comité national des registres
CPP	Comité de protection des personnes
DDASS	Direction départementale des affaires sanitaires et sociales
E3N	Étude épidémiologique auprès de femmes de l'Éducation nationale
E4N	Étude épidémiologique des enfants des femmes E3N
EE	Éthinyl-œstradiol
EGF	Facteur de croissance épidermique (<i>Epidermal growth factor</i>)
EPIC	<i>European prospective investigation into cancer and nutrition</i>
FISH	Hybridation <i>in situ</i> en fluorescence (<i>Fluorescent in situ hybridization</i>)
Francim	France cancer incidence et mortalité
HCL	Hospices civils de Lyon
HER2	Récepteur 2 du facteur de croissance épidermique humain (<i>Human epidermal growth factor receptor-2</i>)
HHV-8	Herpès-virus humain 8
HR	<i>Hazard ratio</i>
HTA	Hypertension artérielle
HTLV-1	Virus T-lymphotrope humain de type 1 (<i>Human T-lymphotropic virus type 1</i>)
IDM	Infarctus du myocarde
IHC	Immunohistochimie
INCa	Institut national du cancer

LISTE DES ABRÉVIATIONS

INPES	Institut national de prévention et d'éducation pour la santé
Insee	Institut national de la statistique et des études économiques
InVS	Institut de veille sanitaire
JRC	Centre commun de recherche de la commission européenne (<i>Joint research centre</i>)
LMNH	Lymphome malin non hodgkinien
MET	Équivalent métabolique d'effort physique (<i>Metabolic equivalent task</i>)
MGEN	Mutuelle générale de l'Éducation nationale
MVTE	Maladie veineuse thrombo-embolique
ORL	Oto-rhino-laryngologie
OMS	Organisation mondiale de la santé
PAF	Fractions attribuables (<i>Population attributable fraction</i>)
PTEN	Phosphatase et homologue de la tensine (<i>Phosphatase and tensin homolog</i>)
RNIPP	Répertoire national d'identification des personnes physiques
RO	Récepteurs à l'œstrogène
ROC	Courbe d'efficacité du récepteur (<i>Receiver operating characteristic</i>)
RP	Récepteurs à la progestérone
RR	Risque relatif
SBR	Scarff-Bloom-Richardson
SIDA	Syndrome d'immunodéficience acquise
THM	Traitements hormonaux de la ménopause
TNM	Tumeur, ganglion lymphatique, métastase (<i>Tumour, node, metastasis</i>)
UICC	Union internationale contre le cancer (<i>Union for international cancer control</i>)
UV	Ultraviolet
VIH	Virus de l'immunodéficience humaine

INTRODUCTION GÉNÉRALE

Partie 1 Épidémiologie des cancers

1.1. Processus biologique

1.1.a. *D'où provient le mot « cancer » ?*

Le cancer¹ est une maladie décrite depuis l'Antiquité. Son nom provient des mots « karkinos » et « karkinoma » (crabe et écrevisse en grec ancien), en raison d'une comparaison qu'aurait faite le médecin grec Hippocrate entre les prolongations de certaines tumeurs* et les pattes d'un crabe.

1.1.b. *La cancérogenèse*

Le fonctionnement de l'organisme humain repose sur l'équilibre entre la production de nouvelles cellules (mitoses* lors des cycles cellulaires) et la destruction des cellules vieilles ou abîmées (apoptose*). Le cancer est une maladie qui se développe en plusieurs phases, dont une phase de latence, pouvant se compter en dizaines d'années, entre l'exposition à un cancérogène* et l'apparition de symptômes cliniques.

Le développement d'un cancer résulte de l'accumulation d'altérations génétiques (figure 1) au cours de la vie et est accéléré par la diminution de la réparation génique (par exemple, inactivation de l'apoptose), l'activation d'oncogènes*, la répression de suppresseurs de tumeurs ou la présence d'agents mutagènes. Par doublements cellulaires successifs, la masse de tissu tumoral ne cesse d'augmenter (à partir d'une seule cellule maligne et après 20 doublements cellulaires, un million de cellules tumorales sont formées, soit environ un milligramme de tissu malin). Lorsque sa taille atteint une dimension suffisante, la masse tumorale se vascularise en sécrétant des molécules qui provoquent la formation de nouveaux vaisseaux sanguins (angiogenèse*). Ces vaisseaux fournissent aux cellules cancéreuses l'irrigation et l'alimentation en oxygène et en nutriments nécessaires à leur croissance et à leur propagation vers les tissus sains voisins (invasion) ou lointains (métastases*).

¹ Un astérisque symbolise un terme présent dans le glossaire page 198. Seule la première apparition du mot est signalée par un astérisque. Dans la version électronique, un hyperlien renvoie vers la définition correspondante dans le glossaire. Depuis le glossaire, un hyperlien renvoie vers le mot correspondant étoilé (*i.e.* la première apparition du mot dans le texte).

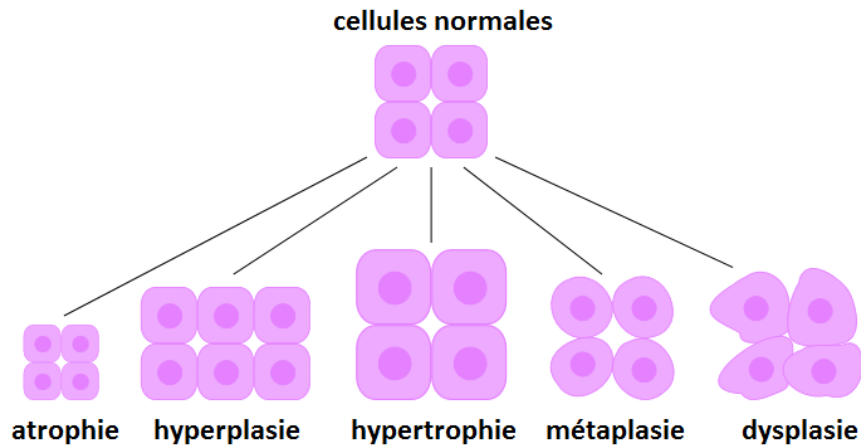


Figure 1 : Représentation des différentes mutations génétiques les plus fréquemment rencontrées.

Note : Les définitions des différentes mutations (atrophie, hyperplasie, hypertrophie, métaplasie et dysplasie) sont présentes dans le glossaire.

Source : The biology of cancer, LaMorte WW., Boston University School of Public Health

La cancérogenèse* correspond à une succession d'étapes qui peuvent se chevaucher dans le temps pour aboutir à l'apparition clinique de la maladie, due à l'accumulation de plusieurs mutations* de l'ADN (acide désoxyribonucléique), en nombre et natures différents selon les types de cancers. Ces mutations entraînent l'apparition de cellules qui se multiplient de façon anarchique, car elles échappent aux processus normaux de régulation de la croissance tissulaire, dont l'apoptose. Au cours de la première étape, dite phase d'initiation, une mutation majeure de l'ADN se transmet de façon irréversible de cellules en cellules lors de la division cellulaire et peut persister dans l'organisme pendant de nombreuses années à l'état latent. Lors de la deuxième phase, dite de promotion, d'autres altérations ont lieu au niveau des cellules de la lignée de la cellule initiée et aboutissent à la prolifération excessive de cellules transformées et différenciées*, c'est-à-dire ayant conservé les caractéristiques du tissu d'origine. Si cette prolifération contient des cellules moins différenciées, susceptibles de se transformer en cellules cancéreuses, on parlera de lésion précancéreuse*. Lors de la troisième phase, dite de progression, les cellules précancéreuses acquièrent les caractéristiques d'une cellule cancéreuse en se multipliant de façon anarchique et en perdant, en partie, leur caractère différencié, formant alors une tumeur maligne *in situ**. L'évolution et la perte des processus normaux de la régulation de la croissance cellulaire peuvent entraîner la contamination des tissus voisins - on parle alors de tumeur maligne* invasive* – puis des tissus plus lointains *via* le passage des cellules cancéreuses dans le sang ou la lymphe* – on parle alors de métastases.

Dans certains tissus épithéliaux*, comme le sein ou le côlon, ces étapes vont correspondre à des phases successives d'hyperplasie* (initiation secondaire à la première mutation), de dysplasie*

(promotion) et de progression (évolution vers un cancer *in situ* puis invasif) (figure 2). Depuis de nombreuses années, il est reconnu que le développement d'un carcinome*, forme la plus répandue des cancers, est déterminé non seulement par une succession de mutations survenant dans les cellules épithéliales, mais aussi par leurs interactions avec l'environnement extracellulaire (microenvironnement* ou stroma*). Les éléments du stroma, tissu conjonctif qui entoure les cellules épithéliales, participent aux étapes de la progression tumorale vers des carcinomes invasifs et métastatiques, en facilitant l'invasion et la dissémination des cellules cancéreuses (1).

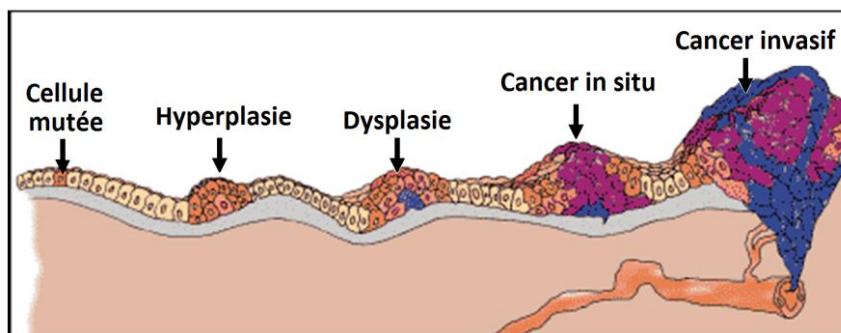


Figure 2 : Représentation de l'évolution d'une tumeur : de la mutation jusqu'au cancer invasif.

Source : The biology of cancer, LaMorte WW., Boston University School of Public Health

1.2. Épidémiologie de l'incidence de cancer en France et dans le monde

Dans cette partie, les statistiques majoritairement décrites sont les taux d'incidence* de cancer, afin de permettre la comparabilité des statistiques sur plusieurs années. En effet, le nombre de cas diagnostiqués dépend fortement de la croissance démographique, qui évolue au cours du temps.

1.2.a. L'incidence du cancer en 2012

Le nombre de cancers, toutes localisations confondues, diagnostiqués chez la femme en 2012, a été estimé à 6,66 millions à travers le monde. Avec une estimation de 1,68 millions de cas diagnostiqués dans le monde, le cancer du sein est le cancer le plus fréquent chez la femme (devant les cancers du côlon-rectum, de l'utérus et du poumon) (2).

La France ne dispose pas de registre national pour mesurer l'incidence du cancer sur l'ensemble du territoire. L'observation et la surveillance des cancers sont réalisées grâce à un travail partenarial entre le réseau des registres des cancers Francim (France cancer incidence et mortalité), le service de biostatistiques des Hospices civils de Lyon (HCL), l'Institut de veille sanitaire (InVS) et l'Institut national du cancer (INCa). Le réseau des registres Francim a été créé en 1997 et regroupe l'ensemble des registres des cancers français (généraux ou spécialisés)

qualifiés par le Comité national des registres (CNR), qui assurent l'enregistrement exhaustif de nouveaux cas de cancers dans une zone géographique délimitée. La base de données commune aux 25 registres français (15 registres généraux et 10 registres spécialisés) regroupe des informations décrivant l'ensemble des données d'incidence et de survie collectées par les registres depuis leur création. Elle permet de décrire plus d'un million de tumeurs, couvrant 20 à 25 % de la population française selon les types de cancers (3). Avec un taux d'incidence de près de 90 nouveaux cas en France en 2012 pour 100 000 femmes, le cancer du sein représente le premier cancer chez la femme (3), soit 31 % des cas de cancer diagnostiqués en France en 2012 (figure 3).

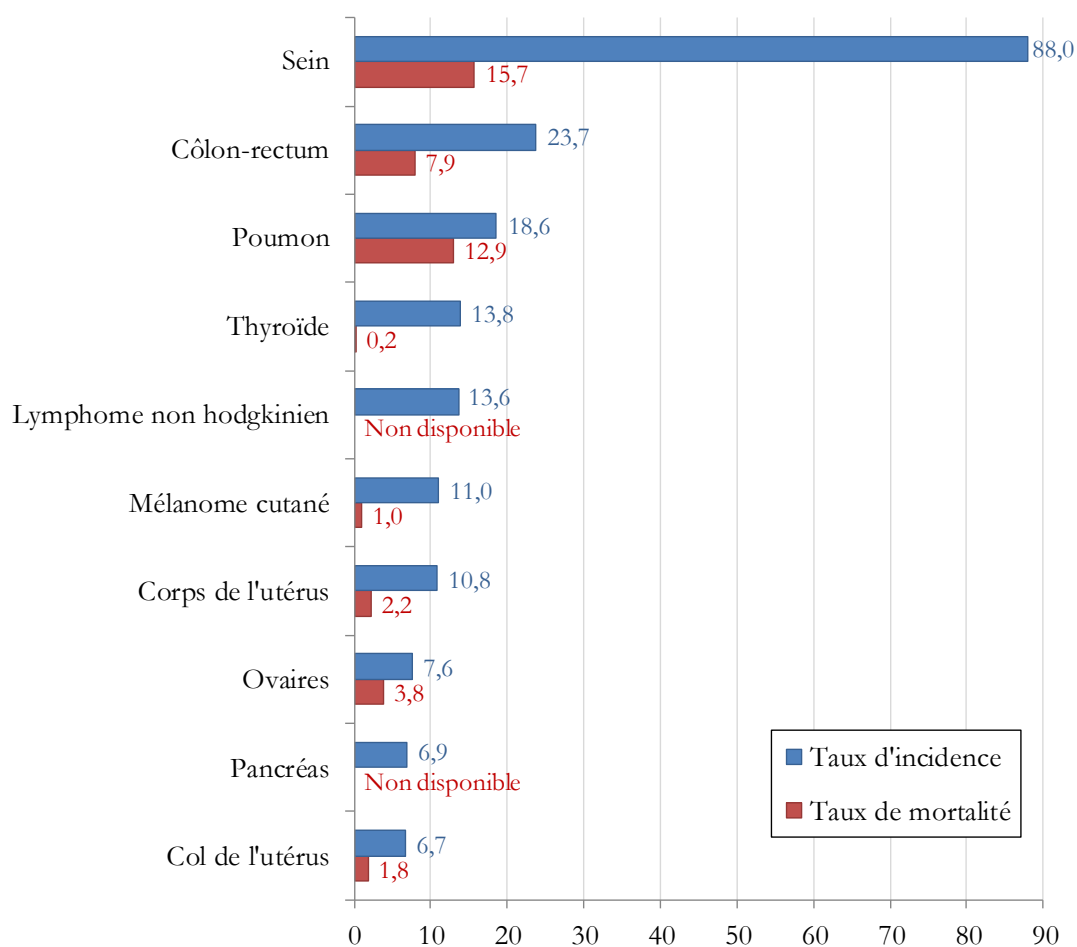


Figure 3 : Taux d'incidence et de mortalité par cancer, standardisés sur la structure d'âge de la population mondiale, pour 100 000 femmes en 2012 en France métropolitaine.

Note : Les localisations sont présentées par ordre décroissant d'incidence. Les localisations de cancer dont le taux d'incidence est inférieur à 5 pour 100 000 femmes ont été omises (rein, lèvre et cavité buccale, pharynx, système nerveux central, lymphome de Hodgkin, estomac, vessie, leucémie aiguë myéloïde, foie, syndrome myéloprolifératif chronique, syndrome myélodysplasique, œsophage et larynx). Les données de mortalité des cancers du pancréas ne sont pas présentées du fait d'une moindre fiabilité de ces données par rapport aux autres localisations, de même que les données de mortalité des lymphomes non hodgkiniens, car le codage des données issues des certificats de décès ne permet pas de classer les décès selon la classification adoptée par Monnereau *et al.* 2013 (4) pour les hémopathies malignes.

Source : Binder-Foucard *et al.* 2013 (3), Monnereau *et al.* 2013 (4)

1.2.b. Évolution de l'incidence du cancer entre 1980 et 2012

Le taux d'incidence du cancer en France, standardisé sur la structure d'âge de la population mondiale est l'un des plus élevés d'Europe et ne cesse d'augmenter depuis les années 1990, avec une tendance à se stabiliser depuis quelques années (figure 4).

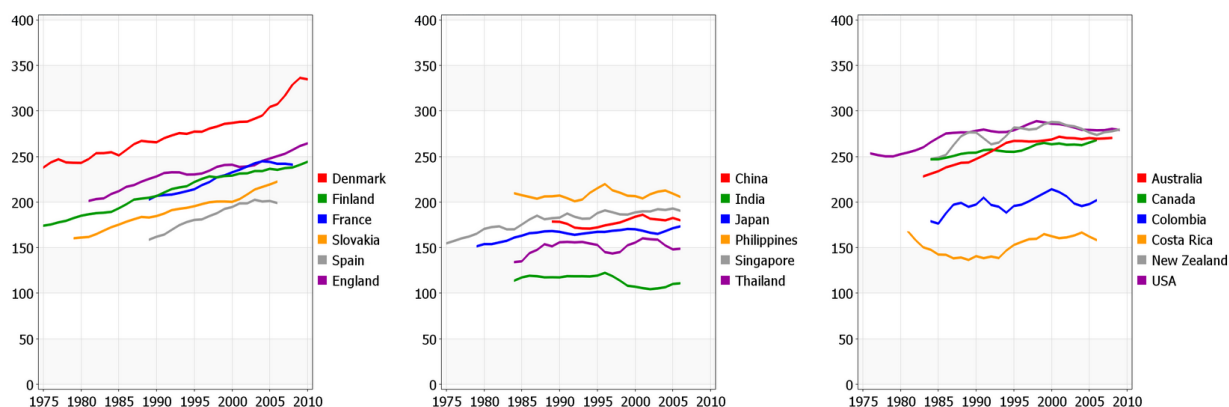


Figure 4 : Évolution du taux d'incidence de cancer pour 100 000 femmes, standardisé sur la structure d'âge de la population mondiale, pour certains pays d'Europe (à gauche), d'Asie (au milieu) et d'Amérique et d'Océanie (à droite) de 1975 à 2010.

Source : Globocan 2012, Section of Cancer Information (site consulté le 11/06/2014) (2)

L'évolution de l'incidence du cancer diffère selon la localisation de la tumeur considérée. En effet, si les taux d'incidence des cancers de la thyroïde, du poumon, du pancréas et des mélanomes cutanés ne cessent d'augmenter depuis les années 80, les taux d'incidence des cancers du côlon-rectum et du corps de l'utérus semblent stables et ceux des cancers du col de l'utérus et des ovaires sont en constante diminution (figure 5). En ce qui concerne le taux d'incidence du cancer du sein, il a presque doublé entre 1980 (56,3 cas pour 100 000 femmes) et 2005 (97,8 cas pour 100 000 femmes), en raison, notamment, d'un diagnostic plus précoce lié au développement du dépistage en France (5). Celui-ci diminue depuis 2005, avec une estimation du taux d'incidence s'élevant à 88,0 cas pour 100 000 femmes en 2012 (3).

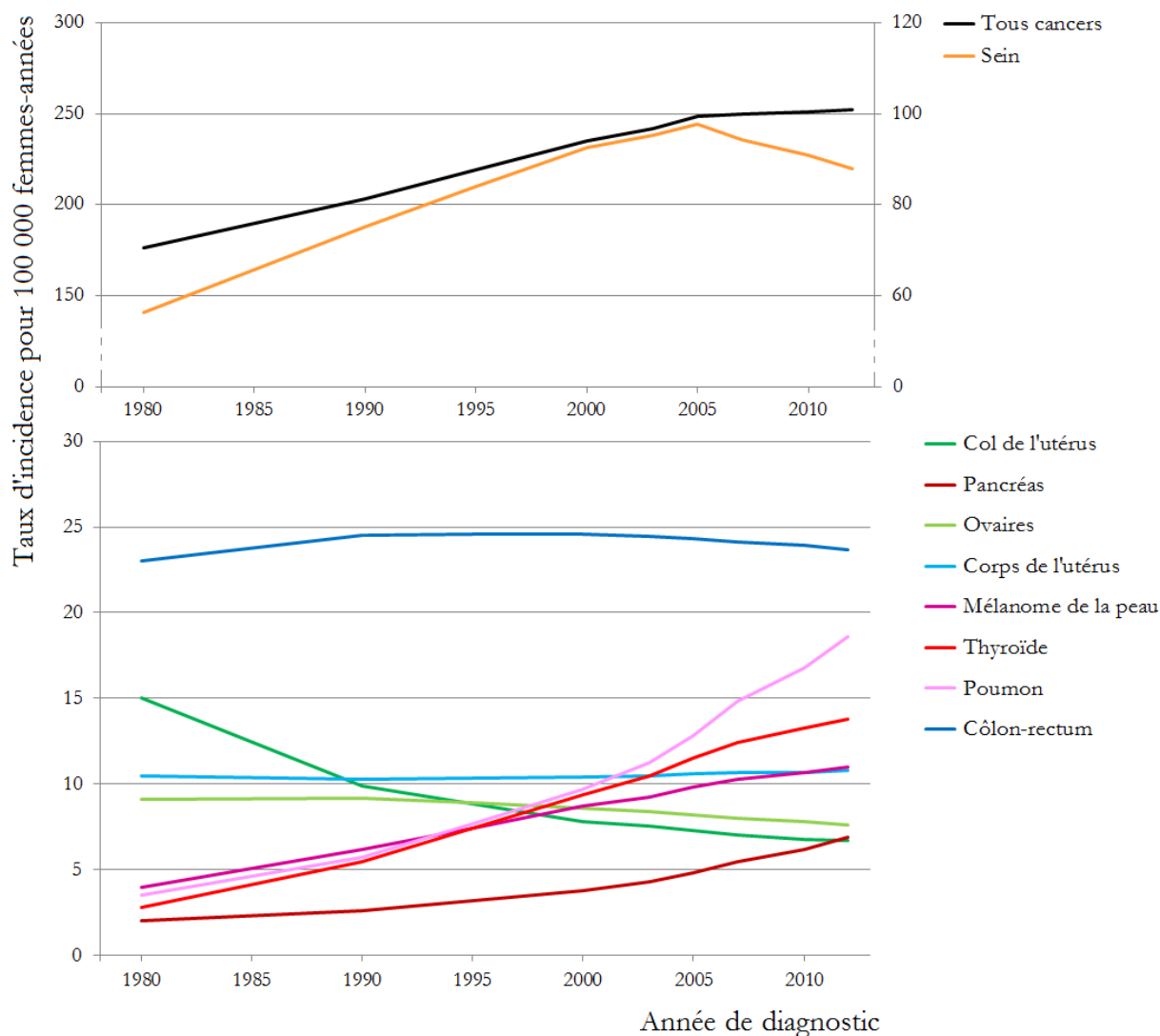


Figure 5 : Évolution du taux d'incidence pour 100 000 femmes en France métropolitaine, standardisé sur la structure d'âge de la population mondiale, pour le cancer, toutes localisations confondues (en haut), du sein (en haut) et d'autres localisations (en bas) entre 1980 et 2012.

Note : Sur le graphique du haut, l'évolution du taux d'incidence de l'ensemble des localisations de cancer est décrite par l'axe des ordonnées de gauche, celle du cancer du sein par l'axe des ordonnées de droite. L'évolution du taux d'incidence pour les lymphomes non hodgkiniens n'est pas présentée en raison d'une hétérogénéité des données disponibles par années recueillies dans Monnereau *et al.* 2013 (4).

Source : Binder-Foucard *et al.* 2013 (3)

1.2.c. Les différents sous-types de cancer du sein

Environ 80 % des cancers du sein sont composés de cellules cancéreuses présentant des récepteurs hormonaux à leur surface, on parle alors de cancers hormono-dépendants. Il existe deux types de récepteurs : les récepteurs à l'œstrogène* (RO) et les récepteurs à la progestérone* (RP), qui déterminent l'hormonosensibilité de la tumeur. La liaison entre les hormones et leurs récepteurs sur les cellules déclenche la stimulation de la croissance des cellules cancéreuses. La présence ou l'absence de ces récepteurs influence le traitement. Les cancers hormono-dépendants ont de meilleurs pronostics à cinq ans, mais de moins bons pronostics une fois ce délai de cinq ans après le diagnostic du cancer écoulé (6).

Plus qu'un facteur pronostic, la présence de récepteurs hormonaux est considérée comme un facteur prédictif de la sensibilité du cancer aux traitements anti-œstrogènes (hormonothérapie) : le tamoxifène* ou les inhibiteurs de l'aromatase*, dont les actions biologiques sont illustrées dans la figure 6.

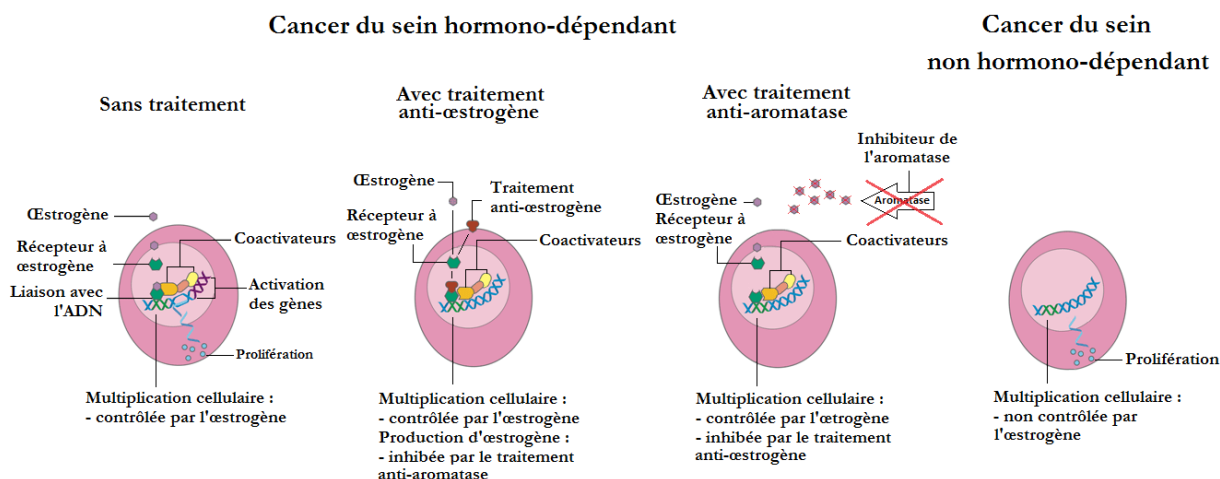


Figure 6 : Action biologique de l'hormonothérapie sur les cellules mammaires cancéreuses.

Source : National Cancer Institute, artwork by Jeanne Kelly (2010)

Différents traitements, autres que l'hormonothérapie, peuvent être utilisés : la chirurgie, la radiothérapie, la chimiothérapie et les thérapies ciblées. Ces traitements sont les seuls envisageables en présence d'une tumeur non-hormonosensible.

Le gène ErbB2, plus connu sous le nom de gène HER2 (récepteur 2 du facteur de croissance épidermique humain, ou en anglais *human epidermal growth factor receptor-2*), est un gène apparenté à celui du récepteur de l'EGF (*epidermal growth factor*), un facteur de croissance, et peut donc favoriser la croissance d'une tumeur. Le gène HER2 encode un récepteur situé à la surface des cellules mammaires et chaque cellule saine en contient deux copies. Dans environ 12 à 20 % des

cancers, un nombre trop élevé de copies est présent dans les cellules (amplification), entraînant une surproduction (ou surexpression) de la protéine HER2. Les tumeurs présentant une surexpression de cette protéine sont plus agressives et tendent à se développer et à se propager plus rapidement que les tumeurs qui produisent la protéine HER2 en quantité normale (7). La surexpression de cette protéine est un facteur péjoratif, néanmoins une thérapie ciblée par traitement au trastuzumab (Herceptin®) peut alors être proposée. Ce médicament est spécialement conçu pour se fixer aux protéines HER2 en excès et faire cesser la croissance du cancer, permettant l'amélioration du pronostic.

La classification moléculaire des cancers du sein, en constante évolution, apporte une valeur pronostique, ainsi qu'une aide à la décision thérapeutique. À partir de critères moléculaires, plusieurs catégories ont été initialement proposées : les cancers luminaux (A et B), HER2 enrichi et *basal-like* (souvent assimilés aux triples négatifs) (8;9). Cette description a été par la suite enrichie par l'identification des sous-types *molecular apocrine*, *claudin-low* et *interferon-rich* (10;11).

1.2.d. Les facteurs de risque de cancer du sein

Le risque de développer un cancer du sein étant très inégal à travers le monde, nous pouvons faire l'hypothèse que les facteurs de risque varient considérablement d'une population à une autre. De nombreux facteurs de risque ont déjà été identifiés, même si l'étiologie du cancer du sein demeure incomplètement connue. On distingue quatre grands types de facteurs de risque : héréditaires, hormonaux, environnementaux et les autres marqueurs de santé.

- **Les facteurs héréditaires**

Les antécédents familiaux constituent un facteur de risque majeur de cancer du sein (12). Le risque de développer un cancer du sein augmente avec le nombre d'antécédents dans la famille au premier degré (mère, sœurs, filles), avec une multiplication du risque allant de 1,80 à 3,90 pour les femmes ayant un à trois antécédents familiaux de cancer du sein (12).

Pourtant, seulement 5 à 10 % des cancers du sein sont directement attribuables à une cause héréditaire, c'est-à-dire résultant de la mutation génétique d'un gène très fortement associé au risque de cancer du sein (13). La présence de cette mutation augmente de 6 à 8 fois le risque de cancer du sein et le risque de cancer précoce (avant 40 ans) chez l'individu porteur (14). Les principales mutations associées au risque de cancer du sein sont celles impliquant les gènes BRCA1* sur le chromosome 17 et BRCA2* sur le chromosome 13 (BRCA pour *breast cancer*). Entre une femme sur 500 et une femme sur 300 serait porteuse d'une altération d'un des deux

gènes (15). Leur risque de développer un cancer du sein avant l'âge de 70 ans est d'environ 60 % pour le premier gène et 55 % pour le second (16). Plusieurs autres gènes de prédisposition génétique au cancer du sein ont été identifiés, comme le gène P53 sur le chromosome 17 qui, associé au syndrome de Li-Fraumeni*, serait responsable de 7 % des cas de cancer du sein avant 40 ans et le gène PTEN (phosphatase et homologue de la tensine) sur le chromosome 10, qui a une faible prévalence, mais induirait un risque de cancer du sein de l'ordre de 85 % chez les femmes porteuses (14). D'autres mutations ont également été impliquées (ATM, CHECK2) et sont plus faiblement associées au risque de cancer. Environ un quart des cancers du sein présentant des antécédents dans la famille au premier degré est expliqué par ces mutations du fait de leur faible fréquence (17;18).

Les facteurs environnementaux partagés dans une famille (mode de vie identique) ne peuvent expliquer les 75 % restants des cas de cancer du sein présentant des antécédents familiaux dans la famille au premier degré. Le modèle retenu est celui de l'association de multiples variants* de gènes, chacun faiblement associé au risque de cancer du sein (augmentant le risque en moyenne de 10 à 20 %), et pouvant interagir avec des facteurs environnementaux (partagés par les membres de la famille) en amplifiant leur impact sur le risque de cancer du sein (17;19).

- **Les facteurs hormonaux et reproductifs**

Le cancer du sein est un cancer hormono-dépendant (20), exceptionnel chez l'homme, et n'apparaissant qu'après la puberté chez la femme. La durée d'exposition aux hormones ovariennes a une influence majeure sur le risque de survenue d'un cancer du sein (21;22).

L'âge aux premières règles (ménarche*), induites par l'apparition de la production d'hormones stéroïdiennes* (œstrogène et progestérone) par les ovaires, est associé au risque de survenue de cancer du sein : le risque est d'autant plus élevé que les premières règles surviennent précocement (23;24). Des cycles menstruels courts, devenant rapidement réguliers, pourraient augmenter le risque de cancer du sein (25;26). En ce qui concerne la ménopause*, une femme ménopausée, dont les ovaires ne produisent plus ni d'œstrogènes ni de progestérone, est à moindre risque qu'une femme du même âge qui n'a pas atteint la ménopause (24). De plus, une survenue tardive de la ménopause est associée à une augmentation du risque de cancer du sein (24).

L'histoire reproductive des femmes influe sur le risque de cancer du sein à court et à long terme. En effet, comparées aux femmes n'ayant jamais donné naissance à un enfant, les femmes ayant mené au moins une grossesse à terme sont à risque augmenté de cancer du sein dans les 15 premières années suivant l'accouchement, quel que soit l'âge à l'accouchement (27). *A contrario*, le

fait d'avoir mené une grossesse à terme est associé à une diminution de risque de cancer du sein à long terme (27), chaque grossesse à terme supplémentaire diminuant le risque d'environ 8 % (23;28). De plus, le risque de cancer du sein est d'autant plus grand que l'âge à la première grossesse à terme est tardif (14;29). Enfin, une diminution du risque de cancer du sein d'environ 4 % par année d'allaitement cumulée supplémentaire a été observée (28). Les grossesses interrompues avant terme, qu'il s'agisse d'un avortement spontané ou provoqué, ne sont pas associées au risque de cancer du sein (30).

Des taux élevés d'hormones circulantes, notamment d'œstradiol*, d'œstrone*, d'androstènedione*, de sulfate de déhydroépiandrostérone* et de testostérone*, ont été associés à un risque augmenté de cancer du sein, avec une augmentation plus marquée en postménopause* (31;32). En revanche, les études menées sur la progestérone n'ont pas permis de mettre en évidence une association entre le niveau endogène* de progestérone et le risque de cancer du sein, que ce soit en préménopause* (33;34) ou en postménopause (34;35). Cependant, ces études sont difficiles à interpréter du fait du caractère fluctuant des hormones stéroïdiennes plasmatiques, particulièrement en préménopause, et du fait que certaines hormones peuvent être métabolisées en d'autres.

Les stéroïdes sexuels exogènes* ont récemment été classés comme cancérigènes par le Centre international de recherche sur le cancer (CIRC) (36). L'utilisation en cours de contraceptifs oraux - consistant généralement en une association d'œstrogènes et de progestatifs* - serait associée à une légère augmentation du risque de cancer du sein, mais les résultats de la littérature divergent à ce sujet (37-40). Le traitement hormonal de la ménopause (THM) consiste généralement en l'administration d'œstrogènes (pour pallier les troubles climatériques* de la ménopause) associés à un progestatif en l'absence d'hystérectomie*, en raison de l'augmentation du risque de cancer de l'endomètre* avec l'utilisation d'œstrogènes seuls (41). Un bon nombre d'études, observationnelles et randomisées, s'accordent sur le fait que l'utilisation de THM est associée à un risque accru de cancer du sein (42). Elles mettent en évidence une augmentation du risque de cancer du sein avec l'utilisation d'un traitement combinant un œstrogène et un progestatif de synthèse (43;44). Cependant, l'augmentation du risque serait moindre pour un traitement associant un œstrogène avec de la progestérone micronisée* ou de la dydrogestérone* que pour un traitement combinant un œstrogène à un autre type de progestatif de synthèse (43;44). Cette augmentation de risque semble d'autant plus importante que le traitement est débuté peu de temps après la ménopause et que la durée d'utilisation est longue (43;45). Un sur-risque persisterait 5 à 10 ans après l'arrêt du traitement (46).

- *Les facteurs environnementaux*

L'impact de l'alimentation sur le risque de cancer du sein n'est pas clairement élucidé. Des hypothèses ont été suggérées quant aux effets de l'apport énergétique, de vitamine D, de graisses alimentaires et de fibres*, mais toutes ces hypothèses ont été largement remises en cause par les grandes études prospectives internationales (47). Le seul facteur de risque alimentaire du cancer du sein clairement identifié est la consommation d'alcool : chaque verre d'alcool consommé par jour serait associé à une augmentation de l'ordre de 7 % du risque de cancer du sein (48;49). Néanmoins, ces observations peuvent refléter l'association du risque de cancer du sein avec un comportement alimentaire particulier, plus qu'avec des aliments ou nutriments isolés. En effet, une alimentation de type occidental (composée principalement de viande, apéritifs, riz, pâtes, pommes de terre, pizzas, tartes, œufs, boissons alcoolisées, gâteaux, mayonnaise, beurre et crème) serait à risque augmenté de cancer du sein, comparée à une alimentation de type méditerranéen (composée principalement de légumes, fruits, poissons, fruits de mer, huile d'olive et huile de tournesol) (50).

La pratique d'une activité physique régulière a été classée comme facteur associé à une réduction du risque de cancer, de façon probable en postménopause et possible en préménopause (51). La réduction de risque associée à la pratique d'une activité physique serait dépendante de son intensité : une activité physique vigoureuse étant plus profitable qu'une activité de faible intensité (52-54).

Des études ont été menées pour identifier l'influence du tabagisme, actif ou passif, sur le risque de cancer du sein et les résultats issus de ces travaux divergent. Le tabagisme actif ne semble pas associé au risque de cancer du sein, sauf en cas d'initiation avant la première grossesse (55;56). Toutefois, il a également été montré une augmentation du risque de cancer avec le tabagisme actif à l'âge adulte en prenant en référence les femmes n'ayant jamais fumé et n'ayant jamais été exposée à du tabagisme passif (57). Enfin, le tabagisme passif, subi par les femmes n'ayant jamais fumé, ne semble pas associé au risque de cancer du sein (58).

La corpulence tout au long de la vie a une influence sur le développement d'un cancer du sein à l'âge adulte. Un gros poids à la naissance a été associé à une augmentation du risque de cancer du sein à l'âge adulte, avec une influence plus marquée pour un diagnostic précoce (59). Il a été également observé qu'une silhouette ronde au moment de la puberté augmente le risque de cancer du sein, en particulier pour un diagnostic tardif (60). Le surpoids et l'obésité à l'âge adulte, définis par l'Organisation mondiale de la santé (OMS) à partir de la valeur de l'indice de masse

corporelle (IMC) (tableau 1), ont été associés au risque de cancer du sein de façon inverse en fonction du statut ménopausique : le surpoids diminue le risque de cancer du sein avant la ménopause, sauf chez les Asiatiques (61), mais l'augmente après la ménopause (la taille et l'adiposité abdominale sont alors des facteurs de risque de cancer du sein) (51;62). En postménopause, les tissus adipeux sont, en effet, l'unique source de production d'œstrogènes. En revanche, l'obésité est inversement corrélée avec les taux plasmatiques d'œstradiol en préménopause, car elle induit des troubles du fonctionnement ovarien. Par ailleurs, l'obésité abdominale est associée à une augmentation de la synthèse de l'insuline, autre facteur hormonal associé au risque de cancer du sein. Enfin, il a été observé une association entre une grande taille à l'âge adulte et le risque de développer un cancer du sein (63).

Tableau 1 : Interprétation de l'IMC par l'OMS.

IMC (kg.m⁻²) = $\frac{\text{masse}}{\text{taille}^2}$	Interprétation
< 16,5	Dénutrition, famine
16,5 - 18,5	Maigreur
18,5 - 25	Corpulence normale
25 - 30	Surpoids
30 - 35	Obésité modérée
35 - 40	Obésité sévère
> 40	Obésité morbide, massive

Source : OMS (64), <http://www.who.int/fr/>

• **Autres marqueurs de santé**

La densité mammaire en mammographie est significativement associée au risque de cancer du sein. Sur une image du sein à la mammographie, les rayons X passent facilement à travers le tissu adipeux, qui apparaît en foncé ; par contre, les tissus épithéliaux et conjonctifs sont plus denses et apparaissent en blanc sur l'image aux rayons X (figure 7). Des études ont montré que les femmes avec une densité mammaire élevée ont 3 à 4 fois plus de risque de cancer du sein que les femmes dont les seins sont presque entièrement constitués de tissu adipeux (65;66). La densité mammaire reflète le nombre de cellules mammaires épithéliales à risque de transformation cancéreuse et leur degré de prolifération, qui augmente les chances de mutations de l'ADN menant à la transformation cancéreuse. Elle reflète également le nombre et le degré de prolifération des cellules stromales entourant les cellules épithéliales, qui jouent un rôle majeur dans l'évolution du cancer en permettant l'invasion et la dissémination des cellules cancéreuses (1). La densité mammaire diminue avec l'âge et serait plus élevée chez les femmes n'ayant pas eu d'enfants, ayant fait une première grossesse tardive et ayant une consommation d'alcool ou un IMC élevés en

postménopause. Il semble cependant que les variations inter-individuelles de la densité mammaire s'expliquent en grande partie par des facteurs génétiques (67).

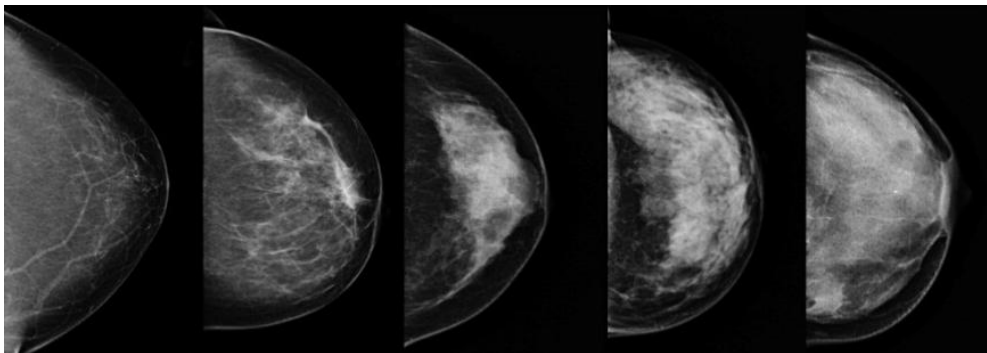


Figure 7 : Représentation de mammographies pour des seins de densité faible (gauche) à élevée (droite).

Source : <http://www.cbcf.org> (site consulté le 23/07/2014)

Un antécédent de maladie bénigne du sein (appelée également mastopathie bénigne) augmente le risque de cancer du sein. Cependant, celui-ci est modulé selon le type de maladie bénigne, le caractère prolifératif de la maladie et l'existence d'antécédents familiaux (68;69). Les maladies bénignes du sein sont usuellement classées en trois catégories : les lésions non prolifératives, les lésions prolifératives sans atypie* et les lésions prolifératives avec atypie. Les lésions non prolifératives et prolifératives sans atypie sont associées à un faible sur-risque de cancer du sein, tandis que les lésions prolifératives avec atypie multiplient le risque environ par 4 (70;71).

Enfin, il y aurait une association entre le diabète* et le risque de cancer du sein. Les femmes diabétiques auraient un risque de 16 à 20 % plus élevé que les femmes non-diabétiques. L'association semblerait plus claire en ce qui concerne le diabète de type II* que de type I* (72;73) et plus marquée après la ménopause (74;75).

1.2.e. Les facteurs de risque des autres localisations de cancer

- ***Le cancer du côlon-rectum***

Le cancer du côlon-rectum est fortement associé au mode de vie. En effet, l'alimentation joue un rôle important dans l'étiologie de ce cancer : une alimentation riche en fibres, céréales complètes et calcium diminuerait le risque de cancer du côlon-rectum, alors qu'une consommation élevée de viandes rouges ou transformées et riches en graisse serait associée à une augmentation du risque de ce cancer. D'autres facteurs de mode de vie influent sur le développement de ce type de cancer, comme le tabagisme, la consommation d'alcool, la sédentarité, le surpoids, l'obésité et l'âge. Le cancer du côlon-rectum est également plus fréquent

chez les hommes et les individus diabétiques, souffrant d'une maladie inflammatoire chronique de l'intestin (telle que la maladie de Crohn* ou la rectocolite hémorragique*) ou infectés par certaines bactéries (telle que l'*Helicobacter Pylori**). Enfin, les facteurs génétiques pourraient influencer la survenue de cancer du côlon-rectum : un sur-risque a été observé chez les individus présentant des antécédents familiaux de cancer du côlon-rectum (47;76-78).

- ***Le cancer du poumon***

Depuis les dernières décennies, le cancer du poumon est devenu un véritable enjeu de santé publique, principalement à cause de la hausse de la consommation de tabac chez les femmes. L'association existante entre l'incidence du cancer du poumon et le tabac est suspectée depuis les années 1920 et confirmée depuis les années 1960. De nos jours, il est estimé que 80 à 85 % des décès par cancer du poumon sont attribuables au tabagisme. Par ailleurs, l'exposition au radon* a été identifiée comme la seconde cause de cancer du poumon après le tabac. Les autres facteurs de risque de cancer du poumon identifiés à ce jour comprennent le tabagisme passif et certaines expositions professionnelles, telles que les expositions à l'amiante*, à la silice*, au chrome*, au cadmium*, au nickel*, à l'arsenic* ou au béryllium*. Le risque du cancer du poumon est plus élevé chez les individus infectés par le virus de l'immunodéficience humaine (VIH) et les femmes survivantes d'un cancer du sein traitées par radiothérapie et fumeuses, à cause de l'effet multiplicateur de l'exposition conjointe des radiations et du tabac (79).

- ***Le cancer de la thyroïde***

Le facteur de risque de cancer de la thyroïde le mieux établi par la littérature est l'exposition aux radiations ionisantes, particulièrement lorsque celles-ci ont lieu dans l'enfance. Les prédispositions familiales semblent tripler le risque de cancer non-médullaire* de la thyroïde (80). La présence de maladie bénigne de la thyroïde, comme les nodules et adénomes thyroïdiens ou le goitre* qui augmentent respectivement de 30 et de 6 fois le risque de cancer. Le cancer de la thyroïde touche trois fois plus de femmes que d'hommes, probablement à cause de l'association existant entre les hormones endogènes et le cancer de la thyroïde. En effet, l'âge au premier enfant et une ménopause artificielle seraient associés à une augmentation modeste du risque de cancer. Les quelques études qui ont analysé le lien entre le cancer de la thyroïde et l'alimentation ont montré que la carence en iode, mais aussi des taux élevés d'iode, sont associés à un risque accru de cancer de la thyroïde, de même qu'une forte consommation de poisson et une faible consommation de légumes. Enfin, le surpoids et l'obésité, ainsi que la sédentarité, sont des facteurs de risque probables de ce type de cancer (81).

- ***Le lymphome malin non hodgkinien (LMNH)***

La présence d'antécédents familiaux de lymphome* augmente le risque d'en développer un soi-même. Le principal autre facteur de risque de LMNH est un système immunitaire affaibli, qu'il soit endommagé à la naissance (congénital) ou plus tard dans la vie (acquis). Les principales raisons d'un système immunitaire affaibli sont la prise d'immunosuppresseur (utilisé principalement à la suite d'une greffe), une infection au VIH et la présence de troubles d'immunodéficience ou auto-immuns. Certaines infections peuvent accroître le risque de LMNH, telles que le virus d'Epstein-Barr (qui provoque la mononucléose infectieuse), le virus T-lymphotrope humain de type 1 (HTLV-1), la bactérie *Helicobacter Pylori*, l'herpès-virus humain 8 (HHV-8) ou le virus de l'hépatite C. D'autres facteurs ont été classés comme facteurs de risque probables de lymphome malin non hodgkinien, tels que l'exposition aux pesticides, au trichloréthylène ou aux colorants capillaires, l'obésité ou une alimentation riche en viande, en produits laitiers et en acides gras saturés, ou faible en légumes (82).

- ***Le mélanome cutané***

Il existe trois grands types de cancer de la peau : le mélanome (touchant les mélanocytes de l'épiderme), le carcinome basocellulaire* (touchant les cellules basales de l'épiderme) et le carcinome épithélial ou spinocellulaire (touchant les cellules de la couche épineuse de l'épiderme) (figure 8). Le carcinome basocellulaire est le cancer cutané le plus fréquent (environ 75 % des cancers), mais également le moins agressif. Contrairement au carcinome basocellulaire qui se développe uniquement localement, les deux autres types de cancer de la peau peuvent franchir la membrane basale et envahir d'autres parties du corps. Le facteur de risque le plus important des cancers de la peau est l'exposition aux rayonnements ultraviolets (UV), qu'ils soient naturels (soleil) ou artificiels (cabines de bronzage) (83). Le risque de cancer de la peau augmente avec la durée d'exposition et est différent selon l'heure d'exposition, avec un risque maximum pour une exposition au soleil entre 10 h et 14 h, heure solaire (en France métropolitaine, en été, cette plage horaire correspond à 12 h - 16 h), au moment où le rayonnement solaire est le plus intense (84). Certaines caractéristiques phénotypiques individuelles prédisposent au cancer de la peau, comme le fait d'avoir une peau claire, des taches de rousseur, des yeux ou des cheveux clairs et une peau incapable de bronzer. Le risque de mélanome est également plus important chez les individus présentant des antécédents familiaux de cancer de la peau, ayant un grand nombre de naevi (appelés couramment grains de beauté) atypiques et avec une exposition prolongée aux hormones ovariennes (83;85).

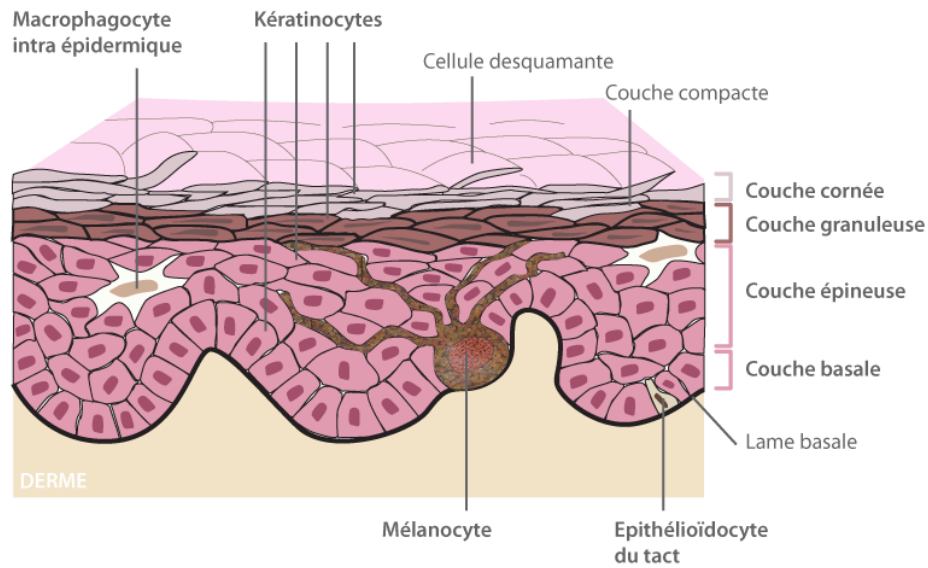


Figure 8 : Représentation d'une coupe transversale de l'épiderme.

Source : <http://www.cosmeticofficine.com/la-peau-et-ses-differentes-couches-tissulaires/lepiderme/> (consulté le 23/07/2014)

• *Les cancers de l'endomètre et des ovaires*

Les cancers de l'endomètre et des ovaires présentent des facteurs de risque reproductifs très similaires. Les facteurs réduisant le risque de ces deux cancers seraient un âge tardif à la ménarche, un âge précoce au premier accouchement, un âge tardif au dernier accouchement, un nombre élevé d'accouchements, une longue période d'allaitement, des cycles courts ou irréguliers et un âge tardif à la ménopause. L'utilisation prolongée de contraceptifs oraux combinés diminue également le risque de ces deux cancers jusqu'à 20 ans après leur arrêt. Les contraceptifs oraux séquentiels à climat œstrogénique dominant, l'utilisation de THM composé d'œstrogène seul et le surpoids seraient associés à une augmentation du risque du cancer de l'endomètre. Le tabagisme serait associé à un risque diminué de cancer de l'endomètre, mais l'association n'est pas clairement définie pour le cancer des ovaires. Les femmes atteintes du syndrome des ovaires polykystiques* présenteraient une augmentation du risque de cancer de l'endomètre et, seulement chez les femmes jeunes, de cancer des ovaires. La présence d'antécédents de fractures et un niveau élevé d'activité physique récréationnelle* seraient associés à un risque diminué de ces deux cancers, avec seulement des preuves limitées pour le risque de cancer des ovaires (86-88).

• *Les cancers du col de l'utérus*

Aujourd'hui, il est estimé que la quasi-totalité des cancers du col utérin est secondaire à une infection par un papillomavirus humain oncogène, essentiellement le 16 et le 18 (HPV), condition nécessaire mais non suffisante d'apparition du cancer du col utérin (89). Ainsi, un dépistage

régulier par frottis tous les trois ans diminue la mortalité par cancer du col de plus de 90 %, le cancer évoluant très lentement (3). La vaccination est recommandée dans de nombreux pays chez les jeunes filles avant le début de leur vie sexuelle (90).

- ***Le cancer du pancréas***

Les principaux facteurs de risque du cancer du pancréas sont un âge avancé, la présence d'antécédents familiaux ou de certaines maladies génétiques, le tabagisme (actif ou passif), l'obésité et certaines expositions professionnelles et industrielles, telles que l'exposition au radon, aux composés et solvants de pétrole, à certains pesticides ou colorants et aux hydrocarbures. Le risque de cancer du pancréas serait également plus élevé chez les hommes, les personnes afro-américaines, diabétiques, ou atteintes de cirrhose alcoolique, du VIH, de l'hépatite B ou de pancréatite chronique. La présence de la bactérie *Helicobacter Pylori* a également été associée à un risque accru de cancer du pancréas, tout comme la présence de certaines mutations génétiques (BRCA2, P16, P53). Enfin, les personnes n'appartenant pas au groupe sanguin O auraient un risque plus élevé de cancer du pancréas (91;92).

Partie 2 Épidémiologie de la mortalité en France

Dans cette partie, nous nous intéresserons essentiellement aux taux de mortalité afin de permettre la comparabilité des statistiques sur plusieurs années. En effet, le nombre de décès dépend fortement de la croissance démographique, qui évolue au cours du temps.

En France, les causes de décès sont recueillies par le Centre d'épidémiologie sur les causes médicales de décès (CépiDc, Inserm) dont les missions essentielles sont le codage des causes médicales de décès, la production annuelle de la statistique des causes médicales de décès en France, la diffusion des données et les études et recherches sur les causes médicales de décès. Leurs données sont issues de deux documents : le certificat médical de décès (rempli par le médecin qui constate le décès) et le bulletin d'État civil de décès (établi par l'officier d'état civil de la mairie après la mort d'une personne) (section Matériel, partie 3 page 86).

2.1. Les principales causes de mortalité chez la femme

2.1.a. *Les principales causes de mortalité en 2011*

En 2011, dernière année pour laquelle les données de mortalité sont disponibles², les principales causes de mortalité des femmes françaises (tous âges confondus) étaient les maladies de l'appareil circulatoire (cardiopathies non-ischémiques, cardiopathies ischémiques et maladies cérébro-vasculaires principalement) et les tumeurs (cancers du sein, du poumon ou du côlon-rectum principalement) (figure 9).

² Le CépiDc met à disposition publiquement les principales statistiques sur les causes de mortalité des années 1979 à 2011, incluses. Les statistiques de l'année 2012 n'étaient pas encore disponibles au 1^{er} septembre 2014 (93).

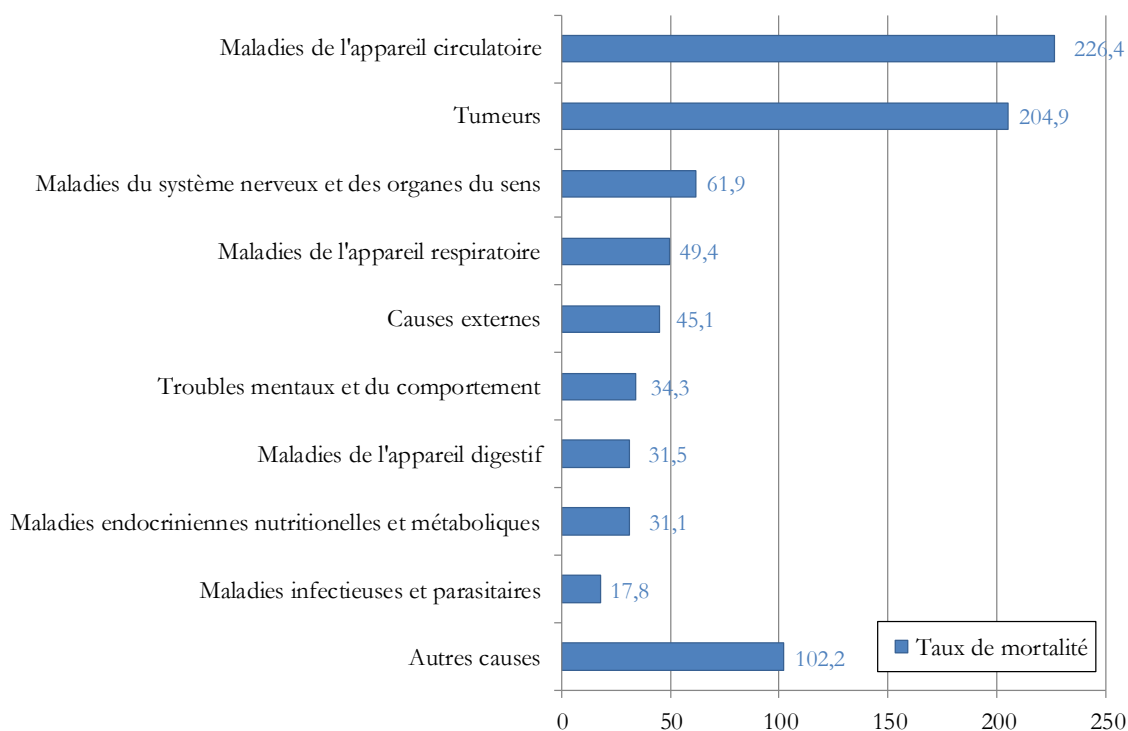


Figure 9 : Taux de mortalité brut pour 100 000 femmes en 2011 en France pour les principales causes de mortalité.

Note : Les maladies infectieuses et parasitaires comprennent le Sida (syndrome d'immunodéficience acquise).

Source : Inserm-CepiDc (Centre d'épidémiologie sur les causes médicales de décès) (93)

Les maladies de l'appareil circulatoire sont la première cause de mortalité chez les femmes (tous âges confondus) avec 226,4 décès pour 100 000 femmes-années en 2011, ce qui représente 73 842 décès, suivies des tumeurs (204,9 décès pour 100 000 femmes-années en 2011, soit un total de 66 829 décès).

Cependant, la distribution des principales causes de mortalité dépend fortement de la tranche d'âge considérée. En effet, avant 75 ans, les tumeurs représentent la principale cause de mortalité (48,5 % des décès), devant les maladies de l'appareil circulatoire (12,7 % des décès), alors qu'au-delà de 84 ans, l'inverse est observé : les maladies de l'appareil circulatoire sont la première cause de mortalité (35,1 % des décès) devant les tumeurs (13,4 %). Les causes de mortalité pour les décès qui surviennent entre 75 et 84 ans sont réparties de façon équitable entre les tumeurs (29,6 % des décès) et les maladies de l'appareil circulatoire (27,5 % des décès) (figure 10).

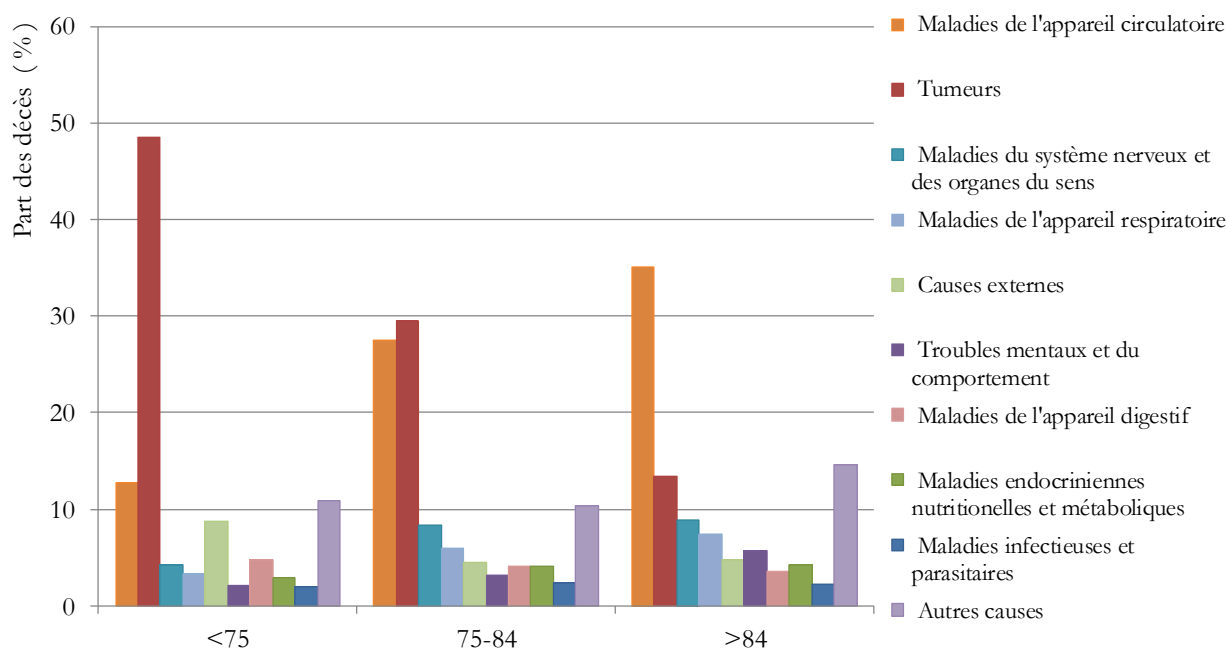


Figure 10 : Part des principales causes de mortalité des femmes, en France, en 2011 parmi l'ensemble des décès en fonction de l'âge au décès (<75, [75 ; 84], >84).

Source : Inserm-CepiDc (Centre d'épidémiologie sur les causes médicales de décès) (93)

2.1.b. Évolution des principales causes de mortalité de 1968 à 2011

La proportion de décès dus aux tumeurs est en constante augmentation chez la femme (figure 11), quelle que soit la tranche d'âge considérée (figure 12), au détriment des décès par maladie de l'appareil circulatoire, qui est en constante diminution (figure 11), quelle que soit la tranche d'âge considérée (figure 12).

On peut également constater que la part de décès dus aux maladies de l'appareil circulatoire chez les femmes de tous âges se rapproche de plus en plus de la part de décès dus aux tumeurs (41,6 % et 20,1 %, respectivement en 1979 contre 28,1 % et 25,5 %, respectivement en 2011). Si la tendance persiste, les tumeurs seront bientôt la première cause de mortalité chez la femme en France.

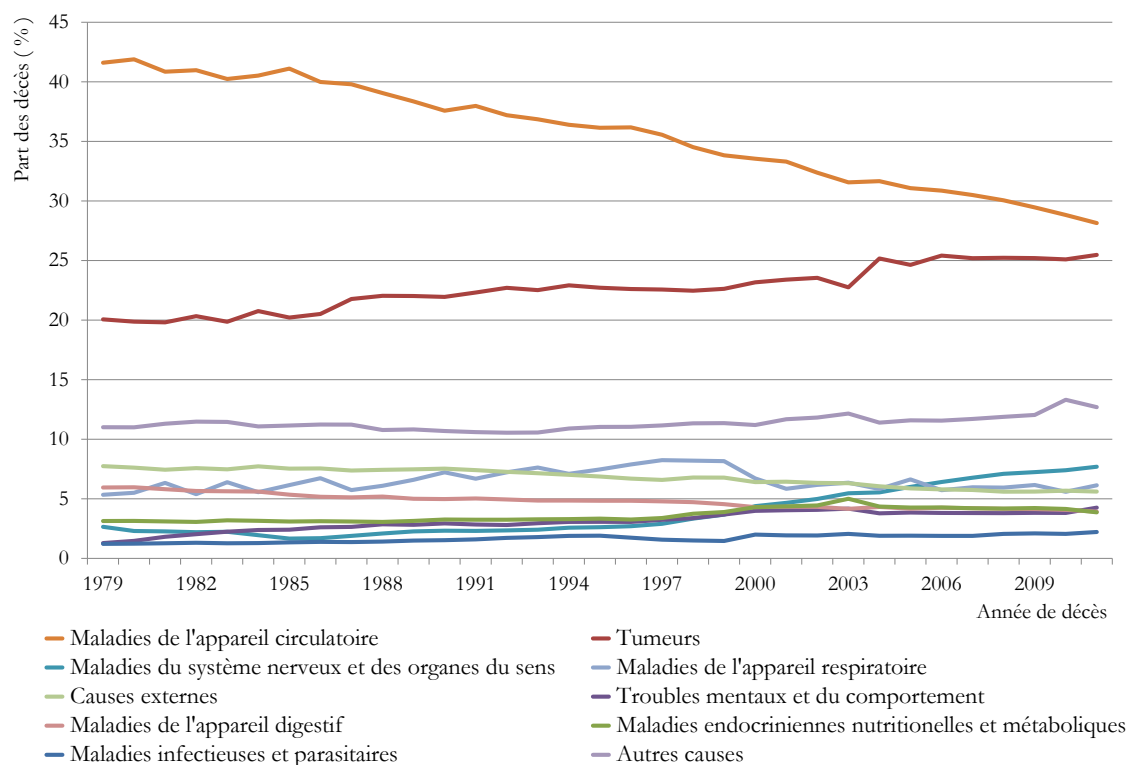


Figure 11 : Évolution de la part des principales causes de décès chez les femmes en France de 1979 à 2011.

Source : Inserm-CepiDc (Centre d'épidémiologie sur les causes médicales de décès) (93)

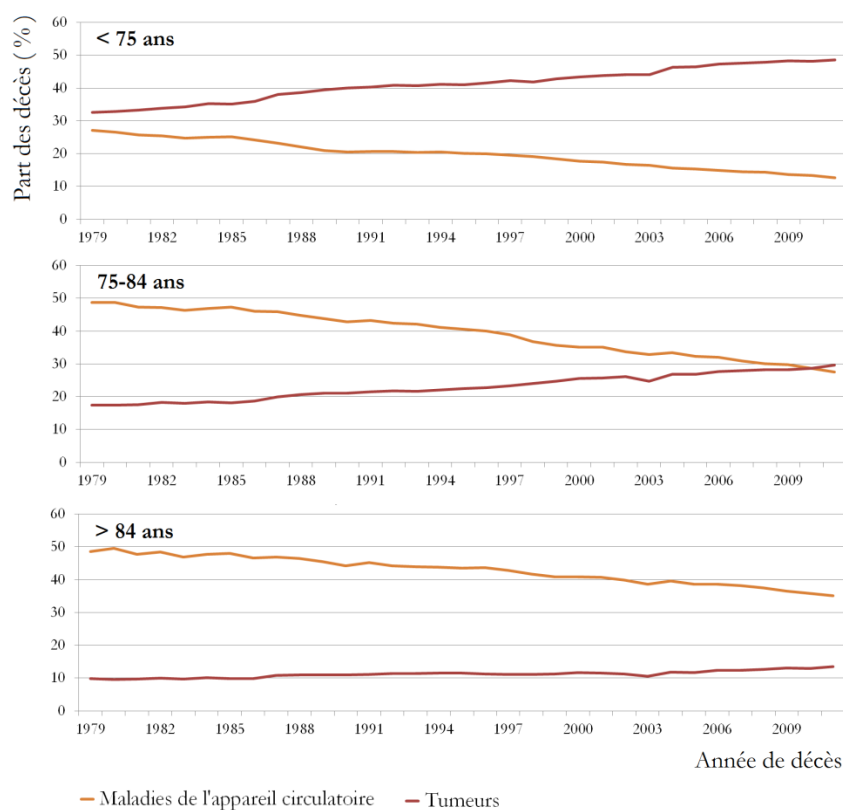


Figure 12 : Évolution de la part des décès par maladie de l'appareil circulatoire et par cancer chez les femmes en France de 1979 à 2011 selon la classe d'âge.

Source : Inserm-CepiDc (Centre d'épidémiologie sur les causes médicales de décès) (93)

2.2. La mortalité par maladie de l'appareil circulatoire chez la femme

Les maladies de l'appareil circulatoire sont des maladies ou des lésions touchant le cœur et les vaisseaux sanguins (artères et veines) transportant le sang du poumon aux organes périphériques, ou inversement. Plusieurs maladies peuvent être identifiées et sont présentées ci-dessous.

La maladie cérébro-vasculaire correspond à une atteinte vasculaire cérébrale, veineuse ou artérielle, avec ou sans interruption de la circulation sanguine. La forme la plus courante de maladie cérébro-vasculaire est l'accident vasculaire cérébral (AVC, appelé communément attaque cérébrale) qui correspond à un arrêt brutal de la circulation sanguine au niveau du cerveau, due, soit à un caillot bouchant une artère à destination du cerveau (AVC ischémique), soit à une hémorragie intracérébrale secondaire à une rupture d'une artère du cerveau (AVC hémorragique).

Les cardiopathies regroupent les maladies du cœur, quelle qu'en soit l'origine. Lorsqu'elles ne sont pas congénitales, les cardiopathies peuvent présenter plusieurs formes. Les cardiopathies coronariennes (ou ischémiques) correspondent à une insuffisance d'oxygénation du muscle cardiaque (myocarde), par rétrécissement ou obstruction complète des artères coronaires irriguant le cœur. La forme la plus courante de cardiopathie ischémique est l'infarctus du myocarde (IDM, appelé communément crise cardiaque), qui est déclenché par l'obstruction d'une artère qui alimente le cœur en sang et donc en oxygène, générant une destruction du tissu non irrigué et donc des problèmes de contraction du myocarde. Les cardiopathies non-ischémiques correspondent aux maladies des valves cardiaques, dues au rétrécissement ou à l'insuffisance d'une des valves cardiaques, à des troubles cardiaques consécutifs à une affection pulmonaire chronique, à des atteintes inflammatoires du myocarde (cardiomyopathie) ou aux maladies du péricarde (membrane fibreuse entourant le cœur).

Les autres maladies de l'appareil circulatoire regroupent les affections cardiopulmonaires et les maladies rhumatismales, ou hypertensives. La thrombose veineuse (appelée également phlébite) désigne la formation d'un caillot dans une veine, qui peut se détacher et provoquer une embolie pulmonaire par l'obstruction de l'artère pulmonaire.

2.2.a. Le taux de mortalité par maladie de l'appareil circulatoire en 2011

En 2011, la mortalité par maladie de l'appareil circulatoire en France métropolitaine chez les femmes s'élevait à plus de 70 000 décès (93), faisant des maladies de l'appareil circulatoire la première cause de décès chez les femmes (figure 9). Avec un taux de mortalité de 75,4 décès pour

100 000 femmes en 2011, les cardiopathies non-ischémiques sont les formes de maladie de l'appareil circulatoire ayant la mortalité la plus élevée (24 602 décès), suivies des maladies cérébro-vasculaires (56,2 décès pour 100 000 femmes et 18 327 décès) et des cardiopathies ischémiques (44,4 décès pour 100 000 femmes et 14 477 décès).

2.2.b. Évolution du taux de mortalité par maladie de l'appareil circulatoire entre 1979 et 2011

Depuis 1979, la part des décès par maladie de l'appareil circulatoire est en constante diminution chez les femmes en France (figure 13) et ce, dans chaque catégorie d'âge. L'évolution de la part relative des principales formes de maladie de l'appareil circulatoire, à savoir les maladies cérébro-vasculaires, les cardiopathies (ischémiques et non-ischémiques) et les autres maladies de l'appareil circulatoire, chez les femmes de tous âges, est présentée dans la figure 13. La part de décès par maladie cérébro-vasculaire parmi les maladies de l'appareil circulatoire est en constante diminution depuis 1979, laissant la place de première cause de mortalité par maladie de l'appareil circulatoire aux cardiopathies non-ischémiques à partir des années 1990.

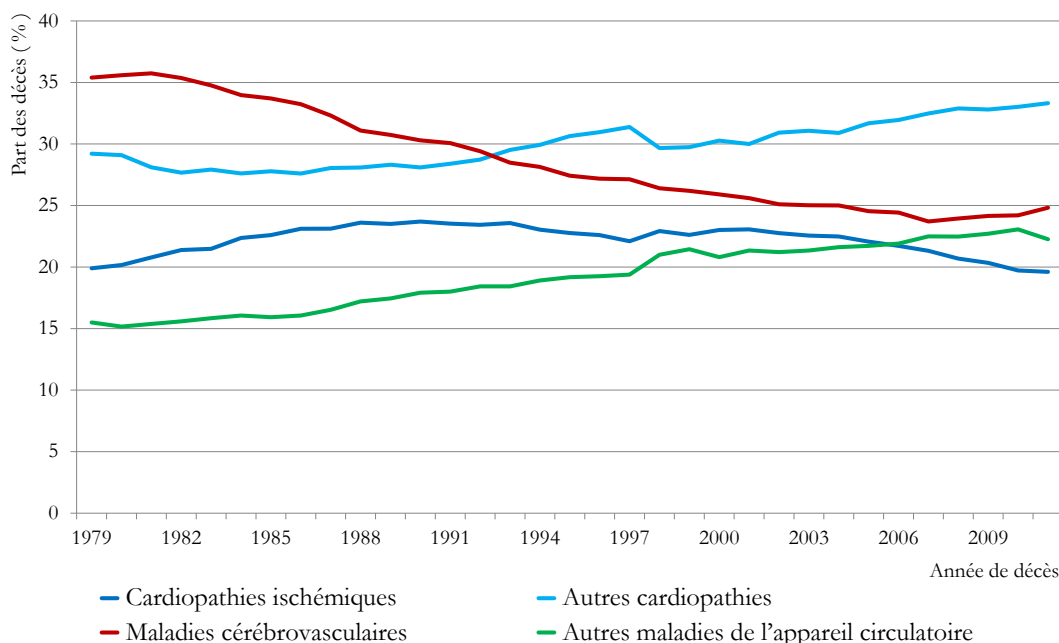


Figure 13 : Évolution de la part des décès des différentes formes de maladie de l'appareil circulatoire parmi l'ensemble des décès par maladie de l'appareil circulatoire, chez les femmes en France, de 1979 à 2011.

Source : Inserm-CepiDc (Centre d'épidémiologie sur les causes médicales de décès) (93)

2.3. La mortalité par cancer chez la femme

Le cancer peut tuer de diverses façons. Une tumeur est un amas de tissu de croissance non contrôlée, fortement demandeur d'un point de vue métabolique, ce qui entraîne de la faiblesse

énergétique dans les formes terminales. Les cellules perdent leur fonction d'origine, ce qui peut bouleverser le fonctionnement normal du corps. Par exemple, les patients atteints de leucémie* sont prédisposés aux infections puisque les globules blancs ne fonctionnent plus correctement. Enfin, les cellules tumorales produisent des médiateurs de l'inflammation (déclenchant l'inflammation) qui causent anorexie, douleur et dysfonctionnement hormonal (94).

2.3.a. Le taux de mortalité par cancer en 2012

En 2012, la mortalité par cancer en France métropolitaine chez les femmes s'élève à plus de 60 000 décès (3), faisant du cancer la deuxième cause de décès après les maladies de l'appareil circulatoire (figure 9), mais la première chez les femmes de moins de 84 ans (figure 10). Avec un taux de mortalité de 15,7 décès pour 100 000 femmes, le cancer du sein est le cancer ayant la mortalité la plus élevée (11 886 décès), suivi des cancers du poumon (12,9 décès pour 100 000 femmes et 8 623 décès) et du côlon-rectum (7,9 décès pour 100 000 femmes et 8 447 décès) (3) (figure 3).

2.3.b. Évolution du taux de mortalité par cancer entre 1980 et 2012

À l'inverse de l'incidence, la mortalité par cancer en France est en constante diminution depuis 1975 (figure 14). Le taux de mortalité en France, standardisé sur la structure d'âge de la population mondiale, est passé de 100,4 en 1980 à 73,2 pour 100 000 femmes en 2012 (3).

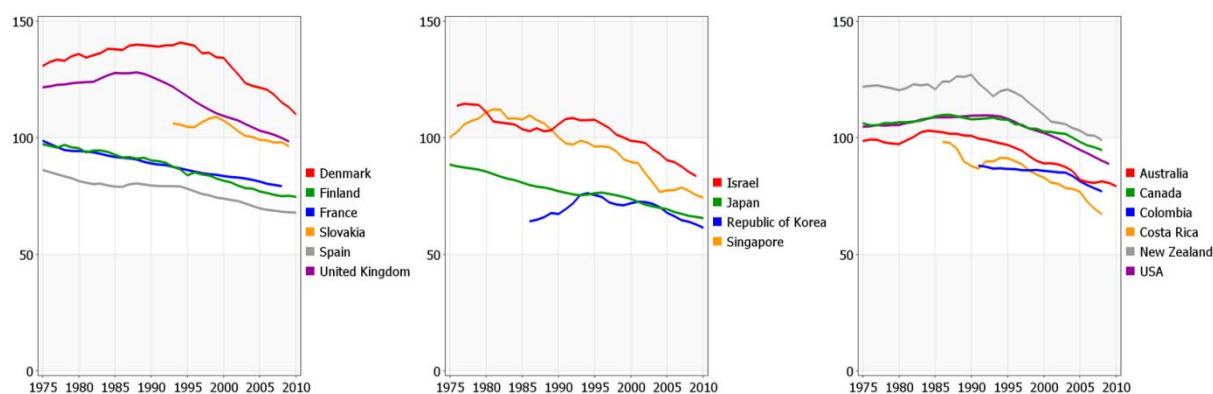


Figure 14 : Évolution du taux de mortalité par cancer pour 100 000 femmes, standardisé sur la structure d'âge de la population mondiale, pour certains pays d'Europe (à gauche), d'Asie (au milieu) et d'Amérique et d'Océanie (à droite) de 1975 à 2010.

Source : Globocan 2012, Section of Cancer Information (11/6/2014) (2)

À l'instar de la constante diminution du taux de mortalité entre 1980 et 2012 pour l'ensemble des localisations de cancer (figure 15), la mortalité pour certaines autres localisations de cancer a également tendance à diminuer (sein, côlon-rectum, ovaires, col de l'utérus). Cependant, certains types de cancer évoluent de manière différente. En effet, les taux de mortalité des localisations de

cancer telles que corps de l'utérus, thyroïde et mélanomes de la peau ont tendance à se stabiliser, alors que la mortalité par cancer du poumon est, au contraire, en nette progression depuis 1980, devenant le deuxième cancer ayant la mortalité la plus élevée, derrière le cancer du sein et devant le cancer du côlon-rectum (figure 15).

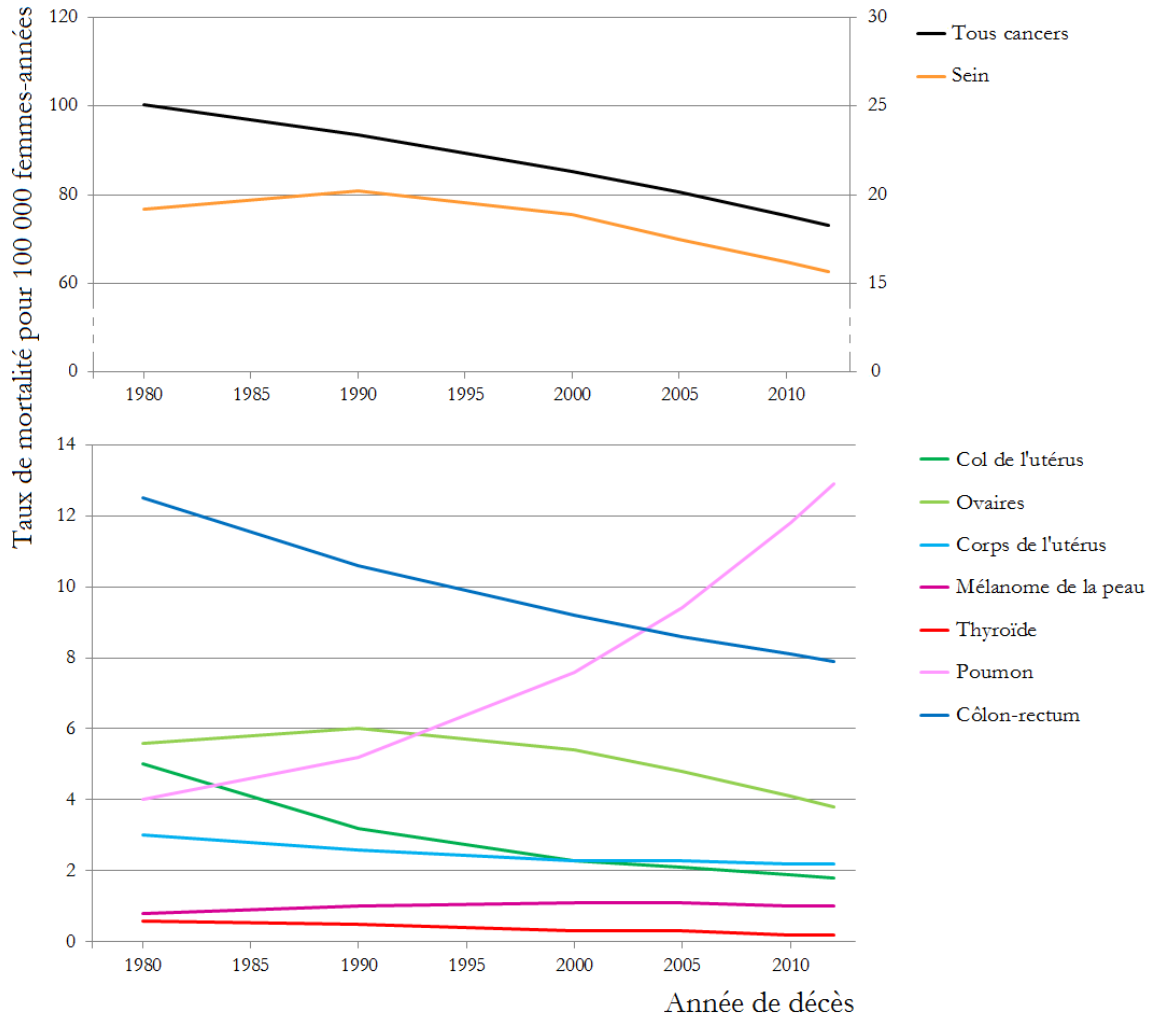


Figure 15 : Évolution du taux de mortalité pour 100 000 femmes en France métropolitaine, standardisé sur la structure d'âge de la population mondiale, pour le cancer, toutes localisations confondues (en haut), du sein (en haut) et d'autres localisations de cancer (en bas) entre 1980 et 2012.

Note : Sur le graphique du haut, l'évolution du taux d'incidence de l'ensemble des localisations de cancers est décrite par l'axe des ordonnées de gauche, celle du cancer du sein par l'axe des ordonnées de droite. Les données de mortalité des cancers du pancréas ne sont pas présentées du fait d'une moindre fiabilité de ces données par rapport aux autres localisations, de même que les données de mortalité des lymphomes non hodgkiniens car le codage des données issues des certificats de décès ne permet pas de classer les décès selon la classification adoptée par Monnereau *et al.* 2014 (4) pour les hémopathies malignes.

Source : Binder-Foucard *et al.* 2013 (3)

2.3.c. Cas particuliers des cancers du sein et du poumon

Le contraste entre l'augmentation de l'incidence du cancer du sein jusqu'en 2005 et la diminution de sa mortalité depuis 1990 peut s'expliquer, en partie par les changements de mode

de vie et l'amélioration des traitements, mais également par le diagnostic plus précoce grâce, notamment, au programme national de dépistage. Si le développement de ce programme a pu faire croître transitoirement le taux d'incidence de cancer du sein, on peut imaginer également que la stabilisation du taux, voire sa diminution, entre 2005 et 2012 pourrait être liée à la baisse de 62 % de l'utilisation de THM (facteur de risque du cancer du sein) entre 2002 et 2006. En effet, la diminution massive de l'utilisation de ces traitements semble la seule modification environnementale majeure pouvant expliquer cette évolution (5;95).

En ce qui concerne l'augmentation de la mortalité par cancer du poumon chez la femme, alors que l'on observe une diminution chez l'homme, cette évolution reflète, pour une grande part, l'entrée plus récente des femmes dans le tabagisme (à la fin des années 1960) (96).

Partie 3 État de l'art

Le sujet de recherche de cette thèse porte sur les liens entre la combinaison de facteurs comportementaux et non-comportementaux et, d'une part, le risque de cancer, en particulier de cancer du sein, et, d'autre part, le risque de mortalité toutes causes. Dans un premier temps, les associations entre le tabagisme, l'indice de masse corporelle, la consommation d'alcool, la consommation de fruits et légumes, le niveau d'activité physique, ainsi que la combinaison de ces cinq facteurs ont été analysées et leur influence sur le risque de cancer et de mortalité globale a été investiguée. Dans un deuxième temps, les relations entre les facteurs de risque de cancer du sein établis dans la littérature et le risque de cancer du sein ont été étudiées afin de rechercher, parmi ces facteurs, ceux ayant le plus d'influence sur ce risque, qu'ils soient comportementaux (les cinq facteurs cités précédemment, le nombre d'enfants, l'âge à la première grossesse, la silhouette à la puberté, la durée d'allaitement cumulée, l'alimentation, l'apport en vitamine D et la prise de traitements hormonaux), ou non-comportementaux (la présence d'antécédents familiaux de cancer du sein, le niveau d'éducation, la taille à l'âge adulte, les antécédents personnels de maladies bénignes du sein, l'âge aux premières règles, le poids de naissance et l'âge à la ménopause). Enfin, dans un troisième temps, nous avons tenté de développer un score de prédiction du risque de cancer du sein. Nous allons passer tout d'abord en revue l'état des connaissances sur ces sujets d'étude.

3.1. L'association entre le mode de vie et le risque global de cancer

3.1.a. *Tabagisme actif et risque de cancer*

La suspicion d'un lien entre le tabagisme et le risque de cancer a commencé à émerger dans les années 1920, mais n'a été confirmée qu'en 1954 à la suite d'une étude portant sur 40 000 médecins anglais (97). Dans une méta-analyse* de 2008, le tabagisme actif était associé à une augmentation du risque de plusieurs localisations de cancer : poumon, larynx, pharynx, cavité buccale, vessie, œsophage, sinus, col utérin, pancréas, estomac, foie et rein. Comparé au fait de n'avoir jamais fumé, le tabagisme actif en cours était associé à une augmentation estimée du risque comprise entre 1,52 (cancer du rein) et 8,96 (cancer du poumon). Cette augmentation semblait persister après l'arrêt du tabac : la multiplication du risque chez les anciens fumeurs a été estimée entre 1,25 (cancer du rein) et 4,65 (cancer du larynx), comparés aux personnes n'ayant jamais fumé (98).

Cette augmentation du risque est due à la présence d'agents chimiques carcinogènes dans la fumée de cigarette, tels que l'arsenic, le benzène*, le benzo[a]pyrène*, le cadmium, le chrome, le formaldéhyde*, le nickel et le N-Nitrosornicotine*. Ainsi, un individu fumeur s'expose à l'inhalation d'environ 1,4 à 2,2 mg d'agents carcinogènes pour chaque cigarette fumée (en comparaison, la nicotine contenue dans chaque cigarette est de l'ordre de 0,8 mg) (99).

3.1.b. Anthropométrie et risque de cancer

L'IMC est la variable la plus fréquemment utilisée dans la littérature pour analyser le lien entre l'anthropométrie et le risque de cancer (tableau 1). Une récente méta-analyse de 2008 a évalué l'influence de l'augmentation de 5 points d'IMC sur le risque de plusieurs localisations de cancer, et en a déduit que cette augmentation serait associée à un risque accru de cancer de l'endomètre, de la vésicule biliaire, de l'œsophage, du rein, de la thyroïde, du sein en postménopause, du pancréas, du côlon et de certaines hémopathies malignes (leucémie, myélome multiple et lymphome malin non hodgkinien). Les risques relatifs associés ont été estimés entre 1,07 (lymphome malin non hodgkinien) et 1,59 (endomètre). Une augmentation de l'IMC serait associée, au contraire, à une diminution du risque de cancer du sein en préménopause (sauf chez les femmes asiatiques) et du poumon, avec des risques relatifs de 0,92 et 0,80, respectivement (61).

L'association inverse observée entre le surpoids et le risque de cancer du sein selon le statut ménopausique serait due aux mécanismes hormonaux impliqués. Durant la ménopause, la production d'hormones ovariennes cesse et les tissus adipeux deviennent la source principale d'œstrogènes *via* une réaction enzymatique* médiée par l'aromatase (permettant de transformer en œstrogènes les androgènes* sécrétés par la surrénale et, de façon résiduelle, par les ovaires). Ainsi, les femmes ménopausées en surpoids ou obèses présentent des niveaux d'œstrogènes plus élevés que les femmes minces. En revanche, en préménopause, l'obésité induit des troubles de l'ovulation induisant une diminution de synthèse de l'œstradiol et de la progestérone ovariens (100). Enfin, selon une hypothèse ancienne, le tissu adipeux séquestrerait l'œstradiol. Ces mécanismes expliqueraient l'association inverse trouvée entre le surpoids et le risque de cancer du sein avant la ménopause (101;102).

3.1.c. Consommation d'alcool et risque de cancer

La consommation d'alcool est associée de façon convaincante au risque de cancer de la cavité buccale, de l'œsophage, du sein et du côlon-rectum (chez les hommes), et un facteur de risque probable de cancer du foie et du côlon-rectum (chez les femmes). Cette augmentation de risque

semble être la même quel que soit le type de boissons alcoolisées consommées (vin, bière, liqueur) et retrouvée dès le premier verre d'alcool consommé, sans seuil sécuritaire (47).

Pour expliquer ces associations, plusieurs mécanismes ont été suggérés. Une des hypothèses avancées serait que la métabolisation* de l'alcool produirait des éléments carcinogènes, comme l'acétaldéhyde* et le diacétyle*. L'alcool potentialise la synthèse d'œstradiol et agirait également comme un solvant, augmentant la perméabilité des membranes cellulaires aux carcinogènes. Enfin, les gros buveurs d'alcool auraient des régimes alimentaires déficients en certains nutriments, rendant le tissu plus réceptif à la carcinogenèse (47).

3.1.d. Consommation de fruits et légumes et risque de cancer

Les fruits et légumes contiennent des vitamines, des minéraux, des fibres alimentaires et d'autres composés bioactifs, tels que les composés phytochimiques*, ce qui peut avoir une influence sur la santé. Une alimentation pauvre en fruits et légumes a été considérée comme un facteur de risque probable pour les cancers de la cavité buccale, de l'œsophage, du poumon et de l'estomac. De plus, la consommation de légumes crus semble diminuer plus fortement le risque de cancer de l'œsophage et de l'estomac que la consommation de légumes cuits (47).

Ces associations pourraient s'expliquer par l'apport par les fruits et légumes de nutriments permettant de prévenir certaines carences. Les substances susceptibles d'être à l'origine d'une réduction du risque de cancer pourraient être le bêta-carotène*, les antioxydants*, les fibres alimentaires ou les folates*, présents en proportion importante dans les fruits et légumes (103).

3.1.e. Niveau d'activité physique et risque de cancer

L'activité physique correspond à tout mouvement faisant appel aux muscles. Il existe plusieurs types d'activités physiques : professionnelle (effectuée pendant l'activité professionnelle), domestique (effectuée chez soi, comme les tâches ménagères), de transport (pour se déplacer) ou récréationnelle (de loisir). Une activité physique peut être classée selon son intensité, qui peut être quantifiée en comparant la dépense énergétique engagée par l'activité physique avec le niveau de dépense énergétique au repos, correspondant à 1 MET-h (*Metabolic equivalent task* en une heure). Ainsi, les activités physiques vigoureuses (course, marche rapide, vélo à vive allure, aérobic, football, volleyball, basketball, tennis en simple, travaux de force), modérées (marche, vélo, ménage, aspirateur, badminton en loisir, tennis en double), légères (marche lente, pêche, cuisine, vaisselle) et sédentaires (activité assise telle que regarder la télé, lire, écrire, travailler à un bureau,

taper sur un clavier, dormir) correspondent respectivement à une dépense énergétique de plus de 6 METs-h, entre 3 et 6 METs-h, entre 1 et 3 METs-h et entre 1 et 1,5 METs-h.

Un faible niveau d'activité physique total a été classé comme facteur de risque convaincant de cancer du côlon-rectum et probable de cancer du sein en postménopause et de l'endomètre (47). Plusieurs explications ont été avancées pour essayer de comprendre les mécanismes sous-jacents à l'association inverse entre l'activité physique et le risque de cancer. Les mécanismes possibles incluent la réduction de la résistance à l'insuline, du fait de la diminution de la masse grasse (en particulier abdominale), la prévention de la prise de poids, la diminution de la synthèse par aromatisation des œstrogènes par réduction du tissu adipeux, la réduction de la durée du transit intestinal ou le renforcement du système immunitaire et des défenses contre le stress oxydatif*.

3.1.f. Combinaison de ces cinq facteurs de mode de vie et risque de cancer

La corrélation* existante entre ces cinq facteurs de mode de vie (104;105), à savoir le tabagisme actif, l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique, incite à leur analyse conjointe, en complément de leur analyse individuelle.

En effet, les adhésions aux recommandations de santé pour les cinq facteurs de mode de vie considérés sont fortement corrélées. Il existe ainsi une corrélation positive entre les deux comportements addictifs de consommation d'alcool et de consommation de tabac, les gros buveurs ayant davantage tendance à fumer que les petits buveurs (106). De façon surprenante, il a été observé une corrélation positive entre la consommation d'alcool et le niveau d'activité physique (107), ce qui pourrait s'expliquer par la recherche d'un style de vie fort en sensations, la réunion après match autour d'un verre ou la volonté de dépenser les calories, apportées par les boissons consommées, à travers l'activité physique. Le tabac diminue les capacités fonctionnelles respiratoires et le niveau de performance du sportif, qu'il soit professionnel ou non. La capacité à pratiquer une activité physique peut être réduite chez les personnes qui fument, en raison d'une réduction de la fonction pulmonaire (108). À l'inverse, la pratique d'une activité physique régulière freine le désir de fumer (109). L'activité physique permet le maintien d'un poids stable (110). Les fumeurs ont un IMC plus faible que les non-fumeurs : les effets de la cigarette sur le poids sont principalement dus à la nicotine qui diminue la prise alimentaire et augmente les besoins énergétiques minimums par modification des neurotransmetteurs du système nerveux central (111). L'arrêt du tabac provoque la reprise de ce poids « manquant », parfois en excès. Les

fumeurs mangent généralement moins de fruits et légumes que les non-fumeurs (112;113), de plus, la consommation de fruits et légumes apporte une sensation de satiété qui diminue l'envie de fumer (114).

L'effet conjoint de la non-adhésion aux recommandations de santé concernant ces facteurs de mode de vie entraîne une multiplication du risque de cancer. En effet, les quelques études analysant l'adhésion simultanée à cinq (115-117) ou quatre (118-128) de ces recommandations ont montré une réduction du risque de cancer plus marquée qu'avec l'adhésion à une seule de ces recommandations. Ces études ont montré une réduction du risque associé à l'adhésion aux recommandations de santé publique concernant l'ensemble des localisations de cancer (117-119;121;122) ou certaines localisations spécifiques de cancer, telles que le cancer du côlon-rectum (120;124), du pancréas (116) et du sein en postménopause (125). L'adhésion à plusieurs recommandations sur ces facteurs de mode de vie serait également associée à une réduction du risque de mortalité par cancer (115;126-128).

L'influence, sur une maladie, d'un facteur de risque ou de la combinaison de plusieurs facteurs de risque dans une population peut être quantifiée par l'estimation de la proportion de cas de cette maladie qui pourraient être attribuables à l'exposition à ce(s) facteur(s) de risque et qui pourraient être théoriquement évités en absence de cette exposition. Les PAFs (pour *population attributable fraction*) sont évaluées à la fois à partir de la force de l'association entre l'exposition et la maladie considérée et de la prévalence de l'exposition au sein de la population d'intérêt (section Méthodes statistiques, partie 3 page 130). Cinq études ont évalué la proportion de cas de cancer attribuables à la non-adhésion à plusieurs recommandations de santé publique (tabagisme, consommation d'alcool, activité physique, tour de poitrine, alimentation, consommation de fruits et légumes, consommation de viande, consommation de sel, IMC) (116-120). Dans ces études, les PAFs combinées ont été estimées entre 11 % et 31 % pour l'ensemble des localisations de cancer (117-119), 6% pour le cancer du côlon-rectum (120) et 27% pour le cancer du pancréas (116).

3.2. Le poids relatif des différents facteurs de risque de cancer du sein

Les facteurs de risque de cancer du sein (largement décrits dans la section Introduction, partie 1.2.d page 38) peuvent être classés en deux catégories : les facteurs directement modifiables (par exemple, la consommation d'alcool ou de fruits et légumes), que nous appellerons facteurs de risque comportementaux, et les facteurs de risque non-comportementaux (par exemple, l'âge aux premières règles ou l'âge à la ménopause).

3.2.a. Les facteurs comportementaux

Les facteurs comportementaux incluent le tabagisme actif, le niveau d'activité physique, le nombre d'enfants, l'âge à la première grossesse à terme, la durée d'allaitement cumulée, la silhouette au moment de la puberté, le type d'alimentation (que nous appellerons profil alimentaire), la consommation d'alcool, l'apport en vitamine D par l'alimentation, l'exposition aux rayons ultraviolets du soleil, l'IMC et l'utilisation de traitements hormonaux (contraceptif oral, progestatif ou THM).

3.2.b. Les facteurs non-comportementaux

Les facteurs de risque non-comportementaux incluent : la présence d'antécédents familiaux de cancer du sein chez les parents au premier degré (mère, sœurs, filles), le niveau d'éducation, la taille à l'âge adulte, les antécédents personnels de maladie bénigne du sein, l'âge aux premières règles, le poids à la naissance et l'âge à la ménopause.

3.2.c. Prise en considération conjointe des facteurs comportementaux et non-comportementaux

L'estimation de la proportion de cas de cancer du sein attribuables à chaque facteur de risque comportemental et non-comportemental permet de quantifier l'impact relatif de ces différents facteurs de risque dans une population donnée. L'évaluation des fractions attribuables à certaines combinaisons de facteurs de risque permet d'estimer l'influence relative et conjointe des facteurs comportementaux et non-comportementaux au sein d'une population.

Plusieurs études ont estimé les proportions de cas de cancer du sein attribuables à la combinaison de plusieurs facteurs de risque dans une population de femmes aux États-Unis (129-132), en Allemagne (133), en Iran (134), en Italie (135;136), en Suède (137) et en Grande-Bretagne (138). Seulement six (131-136) de ces études ont pris en compte le statut ménopausique des femmes dans leurs estimations, malgré les associations différentes entre le cancer du sein en préménopause et le cancer du sein en postménopause pour certains facteurs de risque, dont l'IMC. De grandes disparités dans les estimations peuvent être observées entre les études et entre les populations de femmes en préménopause et en postménopause. Ces disparités peuvent s'expliquer par des prévalences d'expositions différentes selon les pays étudiés, les populations n'ayant pas les mêmes habitudes de vie et caractéristiques physiologiques, et en raison de forces d'association différentes selon le moment du diagnostic du cancer du sein (en fonction de l'âge

ou du statut ménopausique). On peut également remarquer que les différentes études n'analysent pas les mêmes facteurs de risque et ne prennent en compte qu'un nombre réduit de variables.

3.3. Les scores prédictifs de cancer du sein

3.3.a. *Origine des scores prédictifs en santé*

Un modèle de prédiction de risque est un outil statistique permettant d'estimer la probabilité de survenue d'une maladie dans une période de temps définie pour un individu sain avec des caractéristiques individuelles particulières. Ils permettent, pour les professionnels de santé, d'adapter les interventions de prévention et, pour les patients, de se situer par rapport au risque moyen de la population générale.

La cohorte* de Framingham, l'une des plus anciennes cohortes, a permis l'émergence du premier score de prédiction utilisé à l'échelle mondiale. Cette cohorte a été initiée en 1948 dans la ville de Framingham dans le Massachusetts aux États-Unis, choisie pour sa représentativité de la population américaine. L'étude initiale a inclus 5 209 personnes, hommes et femmes, correspondant à 2/3 de la population adulte de la ville, âgées de 30 à 62 ans à l'inclusion. L'objectif du score de Framingham était de prédire la probabilité de survenue d'une maladie coronarienne dans une période de 10 ans. Par la suite, d'autres études d'observation ont permis l'établissement de nombreux scores de prédiction d'autres maladies : diabète en 2003 (139), ostéoporose en 2004 (140) puis en 2010 (141), mélanome en 2005 (142), démence en 2006 (143), cancer du côlon-rectum en 2007 (144) ou cancer du poumon en 2007 (145). La description des scores prédictifs du cancer du sein est développée dans le paragraphe suivant.

3.3.b. *Les scores de prédiction du risque de cancer du sein*

Le tout premier modèle de prédiction du risque de cancer du sein publié a été développé par Gail *et al.* en 1989 à partir d'une étude cas-témoins* (146) et avait pour but de prédire le risque à cinq ans de survenue d'un cancer du sein *in situ* ou invasif. Les facteurs prédictifs pris en compte dans ce modèle sont l'âge, l'âge à la première grossesse, le nombre de biopsies mammaires pratiquées, l'âge à la ménarche et le nombre d'antécédents familiaux de cancer du sein chez les parents au premier degré. Par la suite, d'autres modèles de prédiction ont été construits, soit en établissant un modèle original (147-156), soit en adaptant le modèle de Gail à d'autres populations (157-161) ou en prenant en compte des facteurs de risque complémentaires comportementaux (consommation d'alcool, utilisation de THM, activité physique), non-comportementaux (âge à la ménopause, nombre d'enfants, ethnicité, taille) ou des biomarqueurs

(densité mammaire, facteurs génétiques) (162-166). Tous les modèles de prédiction sont établis à partir d'un modèle statistique (régression logistique, de Poisson ou de Cox). Les variables prises en considération dans la construction du modèle prédictif varient d'un modèle à l'autre, mais on peut observer que tous les modèles prennent en compte l'âge des individus et que la majorité d'entre eux prennent en compte l'âge à la première grossesse, l'âge aux premières règles, le nombre d'enfants, la présence d'antécédents de cancer du sein chez les parents au premier degré et le nombre de biopsies mammaires ou d'antécédents personnels de maladie bénigne du sein. De plus, plusieurs modèles prennent également en compte l'IMC, la taille à l'âge adulte, la consommation d'alcool, l'âge à la ménopause et l'utilisation de THM. Cependant, seulement trois modèles (149;153;156) prennent en considération le statut ménopausique dans la construction du score de risque, dont deux (149;153) construisent deux scores distincts pour la prédiction du risque de cancer du sein chez les femmes en préménopause et en postménopause.

3.4. L'association entre le mode de vie et la mortalité

3.4.a. *Tabagisme et risque de mortalité*

Le tabac demeure la première cause de mortalité évitable dans le monde, en étant responsable de plus de 6 millions de morts par an, un chiffre en constante augmentation : si la tendance actuelle se poursuit, le tabac causera plus de 8 millions de décès en 2030 (167). Il a été estimé que le taux de mortalité des fumeurs était de 1,5 à 3 fois supérieur à celui des non-fumeurs (168) et que s'arrêter de fumer entre 35 et 49 ans procure un gain moyen de 10 années d'espérance de vie (167). Par ailleurs, le tabagisme passif est une exposition courante dans de nombreux pays. En 2004, dans le monde, il a été estimé qu'environ 33 % d'hommes et 35 % de femmes étaient régulièrement exposés au tabagisme passif, qui serait responsable d'environ 600 000 morts en 2004 (169).

Pour combattre cette épidémie, de nombreuses politiques anti-tabac ont été mises en place dans le monde, comme l'installation d'espaces non-fumeurs, l'instauration de programmes de sevrage tabagique, l'interdiction de la publicité, l'augmentation des taxes et la mise en place de mises en garde sanitaires (comme les étiquettes apposées sur l'emballage des produits du tabac : « Fumer tue », « Fumer provoque le cancer du poumon » ou les campagnes anti-tabac dans les médias « Quand on est libre, pourquoi choisir d'être dépendant ? ») (167).

3.4.b. Anthropométrie et risque de mortalité

Le surpoids et l'obésité représentent l'une des principales causes de mortalité dans le monde causant 3,4 millions de morts chaque année (170). Dans une récente méta-analyse collaborative de 57 études prospectives, il a été estimé qu'à partir de 25 kg.m⁻², chaque augmentation de 5 points d'IMC était associée à une augmentation de la mortalité de l'ordre de 30 % (171). À l'échelle mondiale, le surpoids (IMC égal ou supérieur à 25 kg.m⁻², tableau 1 page 42) concerne 34 % d'hommes et 35 % de femmes de plus de 20 ans et l'obésité (IMC égal ou supérieur à 30 kg.m⁻², tableau 1 page 42) respectivement 10 % et 14 % (172). En France, en 2012, il a été estimé que le surpoids et l'obésité touchent près de 21 millions de Français de plus de 18 ans, ce qui représente 14,3 % des hommes et 15,7 % des femmes (173).

Au fil des années, la lutte contre le surpoids et l'obésité est devenue un enjeu de santé publique. Afin de réduire la charge du surpoids et de l'obésité, qui est en constante augmentation, plusieurs actions sont menées, comme promouvoir une alimentation saine (limiter l'apport énergétique provenant de la consommation de graisses totales et de sucres, consommer davantage de fruits et légumes, de légumineuses, de céréales complètes et de noix) et pratiquer une activité physique régulière (150 minutes par semaine pour un adulte). Depuis 2001, de nombreuses mesures sont également entrées en vigueur, comme l'interdiction des distributeurs automatiques de boissons sucrées et confiseries dans les établissements scolaires, la réglementation des publicités sur les produits sucrés et la mise en place du Plan National Nutrition Santé (PNNS) (174), avec, notamment, ses messages sanitaires introduits dans les publicités et autres outils de communication : « Pour votre santé, évitez de manger trop gras, trop sucré, trop salé » ou encore « Pour votre santé, évitez de grignoter entre les repas ».

3.4.c. Consommation d'alcool et risque de mortalité

Dans le monde, la consommation d'alcool serait à l'origine de 3,3 millions de morts en 2012, ce qui correspond à 5,9 % de l'ensemble des décès. L'association entre la consommation d'alcool et le risque de mortalité serait en forme de « J », avec une légère diminution du risque de mortalité pour une faible consommation d'alcool, et une augmentation du risque pour une forte consommation (175). Cependant, cette association ne serait pas confirmée chez les seniors (176). En 2010, il a été estimé qu'environ 6 litres d'alcool sont consommés en moyenne par personne par an dans le monde, quantité environ deux fois plus élevée en Europe (10,9 litres) (177). En France, la consommation moyenne d'alcool par habitant et par an a été estimée à 12,2 litres en

2010 (17,8 litres pour les hommes et 7,1 litres pour les femmes). Les boissons alcoolisées les plus consommées sont le vin (56 %), les spiritueux (23 %) et la bière (19 %) (177).

Afin de lutter contre ce problème de santé publique, plusieurs interventions ont été mises en place à travers le monde. En France, la vente des boissons alcoolisées fait l'objet d'une réglementation spécifique et des campagnes de sensibilisation sont régulièrement diffusées, comme la sensibilisation de l'alcool au volant (« Alcool au volant : quand on tient à quelqu'un, on le retient », « Alcool ou conduire, il faut choisir »), de la consommation d'alcool pendant la grossesse (« Zéro alcool pendant la grossesse, le faire savoir, un devoir », « Quand maman boit, bébé trinque ») ou de la consommation d'alcool excessive chez les jeunes et les moins jeunes (« L'abus d'alcool est dangereux pour la santé, à consommer avec modération »).

3.4.d. Consommation de fruits et légumes et risque de mortalité

Jusqu'à 1,7 millions de vies dans le monde pourraient être épargnées chaque année moyennant une consommation suffisante de fruits et légumes, ce qui place la faible consommation de fruits et légumes parmi les dix principaux facteurs de risque de mortalité dans le monde (178).

En Europe, en 2004, la consommation de fruits et légumes a été estimée à environ 386 g/jour, ce qui correspond à une consommation de 4,8 portions de fruits et légumes par jour, proche des recommandations de santé publique qui préconisent de manger au moins cinq portions de fruits et légumes par jour (400 g/jour). Cependant, peu de pays respectent les recommandations incitant à consommer au moins 400 g de fruits et légumes par jour (Pologne, Allemagne, Italie et Autriche) (179). Les Français sont parmi les meilleurs consommateurs de fruits et légumes en Europe (179), sans pour autant adhérer aux recommandations car il a été estimé que seulement 42,8 % des adultes de plus de 18 ans en ont consommé au moins cinq portions par jour en 2006, les hommes et les jeunes ayant tendance à moins en consommer (180). En France, un message bien connu de l'Institut national de prévention de l'éducation pour la santé (INPES), qui accompagne les publicités pour produits alimentaires, informe : « Pour votre santé, mangez au moins cinq fruits et légumes par jour ».

3.4.e. Niveau d'activité physique et risque de mortalité

Le manque d'activité physique serait responsable d'un décès sur dix dans le monde, un des facteurs de risque de mortalité les plus répandus et les plus persistants. La prévalence d'une activité physique inadéquate (moins de 20 minutes d'activité physique vigoureuse³ 3 j/semaine,

³ cf. section Introduction, partie 3.1.e page 31 pour les définitions des activités physiques vigoureuse et modérée.

ou moins de 30 minutes d'activité physique modérée¹ ou de marche à pied 5 j/semaine) toucherait un adulte sur trois dans le monde (181). La pratique d'une activité physique régulière (au moins une demi-heure par jour) serait associée à une réduction du risque de mortalité de 19 %, comparé à l'absence d'activité physique (182). Cette diminution de risque serait plus marquée chez les femmes et pour des activités physiques intenses (182;183). Cependant, la comparaison des résultats sur le sujet est difficile en raison de divergences dans la définition de l'activité physique entre les études.

En France, peu de mesures ont été mises en place pour promouvoir l'activité physique à tout âge. Seul le PNNS (174), à travers son programme « manger, bouger », son slogan « Pour votre santé, pratiquez une activité physique régulière » et son guide « La santé vient en bougeant », incite les jeunes et les moins jeunes à pratiquer une activité physique régulière.

3.4.f. Combinaison de ces cinq facteurs de mode de vie et risque de mortalité

Tout comme le risque de cancer, on peut observer que le risque de mortalité peut être minimisé par l'adoption conjointe de plusieurs recommandations de santé publique. En effet, une méta-analyse, fondée sur 15 études et datant de 2012, a étudié la réduction du risque de mortalité toutes causes associée à l'adhésion à 1, 2, 3 ou au moins 4 comportements sains pour la santé, estimant la réduction de risque à respectivement 26 %, 42 %, 46 % et 68 % (184). La réduction de risque semble légèrement plus prononcée chez les femmes que chez les hommes (121;127;185-189). Par contre, les résultats sont divergents pour évaluer si la combinaison de plusieurs facteurs de mode de vie réduit plus fortement la mortalité par cancer (117;128;190-192) ou par maladie de l'appareil circulatoire (115;126;127;193-195).

Onze études ont évalué la proportion de décès attribuables à la non-adhésion à plusieurs recommandations de santé publique conjointement (115;117;128;188-190;195-199). Ces fractions attribuables combinées ont été estimées entre 29 % et 60 % pour la mortalité toutes causes (115;117;128;188-190;195-198), entre 22 % et 75 % pour la mortalité par maladie cardiovasculaire (115;117-119;128;190;195;199) et entre 19 % et 60 % pour la mortalité par cancer (115;128;190;195).

Partie 4 Justification du travail doctoral

Les organismes de santé publique publient régulièrement des recommandations afin de mettre en exergue les comportements à risque de cancer ou de mortalité. Ainsi, ces organismes espèrent faire évoluer les mentalités et la manière de vivre afin que l'incidence des cancers diminue et que l'espérance de vie augmente. La littérature, cependant, ne fournit pas d'informations claires sur le poids relatif des facteurs comportementaux (que l'on peut modifier, au moins à un moment de sa vie) et des facteurs non-comportementaux (que l'on ne peut pas directement modifier), ni sur leur importance relative.

L'objectif principal de cette thèse est de préciser le lien existant entre la combinaison de plusieurs facteurs (comportementaux et non-comportementaux) et le risque de cancer (en particulier du sein) et de mortalité. Les données riches et variées de la cohorte E3N permettent d'apporter des résultats supplémentaires à une littérature parfois insuffisante sur l'évaluation de la combinaison de plusieurs facteurs.

La première partie portera sur l'analyse de l'influence du mode de vie, évalué à l'aide de cinq facteurs comportementaux simples, à savoir : le tabagisme, la corpulence *via* l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique, sur le risque de cancer, toutes localisations confondues et par type de cancer. Ce travail a également pour but d'estimer le pourcentage de cas de cancer qui auraient pu potentiellement être évités si toutes les femmes avaient suivi les recommandations de santé publique par rapport à ces cinq comportements.

La deuxième partie ciblera plus particulièrement le cancer du sein, cancer le plus fréquemment diagnostiqué chez les femmes. Le poids des différents facteurs de risque établis de la littérature sera évalué, en les différenciant suivant leur caractère comportemental ou non-comportemental. Étant donné que les facteurs de risque diffèrent suivant que le cancer du sein s'est développé en pré- ou en postménopause, les analyses seront différenciées suivant le statut ménopausique des femmes au moment du diagnostic. L'hétérogénéité entre les différents types de cancer du sein sera également appréhendée en effectuant des analyses séparées selon le statut des récepteurs hormonaux à l'œstrogène et la surexpression de la protéine HER2 de l'ensemble des cas de cancer du sein.

La troisième partie aura pour but la construction d'un score de prédiction de cancer du sein, à partir des facteurs de risque les plus influents, ciblés dans la deuxième partie. Cette partie permettra de déterminer les facteurs de risque qui possèdent les plus forts pouvoirs prédictifs

dans l'estimation du risque individuel de survenue de cancer du sein. Le score établi à partir d'un modèle de régression de Cox sera comparé à deux autres scores prédictifs construits à partir de deux méthodes différentes : l'algorithme des plus proches voisins et le modèle de régression logistique de Gail. Les performances des scores prédictifs seront évaluées en séparant les femmes selon leur statut ménopausique afin de prendre en compte la différence d'étiologie observée entre les cancers du sein avant et après la ménopause.

La quatrième partie portera sur l'analyse de l'association entre les cinq facteurs comportementaux pris en compte dans la première partie et le risque de mortalité, toutes causes confondues et par cause spécifique de mortalité. Les pourcentages de décès attribuables aux recommandations individuelles de santé publique et à leur combinaison seront estimés.

Ce travail doctoral permettra de mieux comprendre l'association entre les facteurs comportementaux et non-comportementaux et le risque de cancer et de mortalité chez la femme, avant et après la ménopause. Il permettra également de compléter la littérature sur la construction d'un score prédictif de cancer du sein, en mettant en avant les facteurs de risque à forte valeur prédictive.

MATÉRIEL

Partie 1 La cohorte E3N-EPIC

1.1. Présentation de la cohorte E3N et objectifs de l'étude

L'étude française E3N (200;201), Étude Épidémiologique auprès de femmes de la Mutuelle générale de l'Éducation nationale (MGEN), est une enquête de cohorte prospective menée auprès de 98 995 femmes volontaires françaises, âgées de 40 à 65 ans (nées entre 1925 et 1950) et résidant en France métropolitaine au moment de l'inclusion en 1989-1990. L'objectif initial d'E3N était l'identification et l'analyse du rôle de certains facteurs de risque hormonaux, reproductifs, alimentaires ou liés au mode de vie, dans la survenue des cancers chez les femmes. Depuis, l'étude s'est élargie à d'autres maladies chroniques et à la mortalité.

Le choix de la collaboration avec la MGEN a été motivé par le besoin d'optimiser le suivi à long terme des participantes. En effet, cette collaboration permet un suivi régulier des femmes, répondantes et non-répondantes, assurant un taux de perdues de vue faible (de l'ordre de 2,4 % en 2012, c'est-à-dire après plus de 20 ans de suivi).

L'étude est dirigée par Françoise Clavel-Chapelon, directrice de recherche à l'Inserm. La logistique de cette étude, de l'envoi des questionnaires au recueil de l'information, ainsi que son analyse, s'effectue au sein de l'équipe 9 « Nutrition, hormones et santé des femmes » du Centre de recherche en épidémiologie et santé des populations (CESP, Inserm UMR 1018), implantée à l'institut de cancérologie Gustave Roussy basé à Villejuif.

1.2. Le recueil des données

1.2.a. *Le commencement de l'étude*

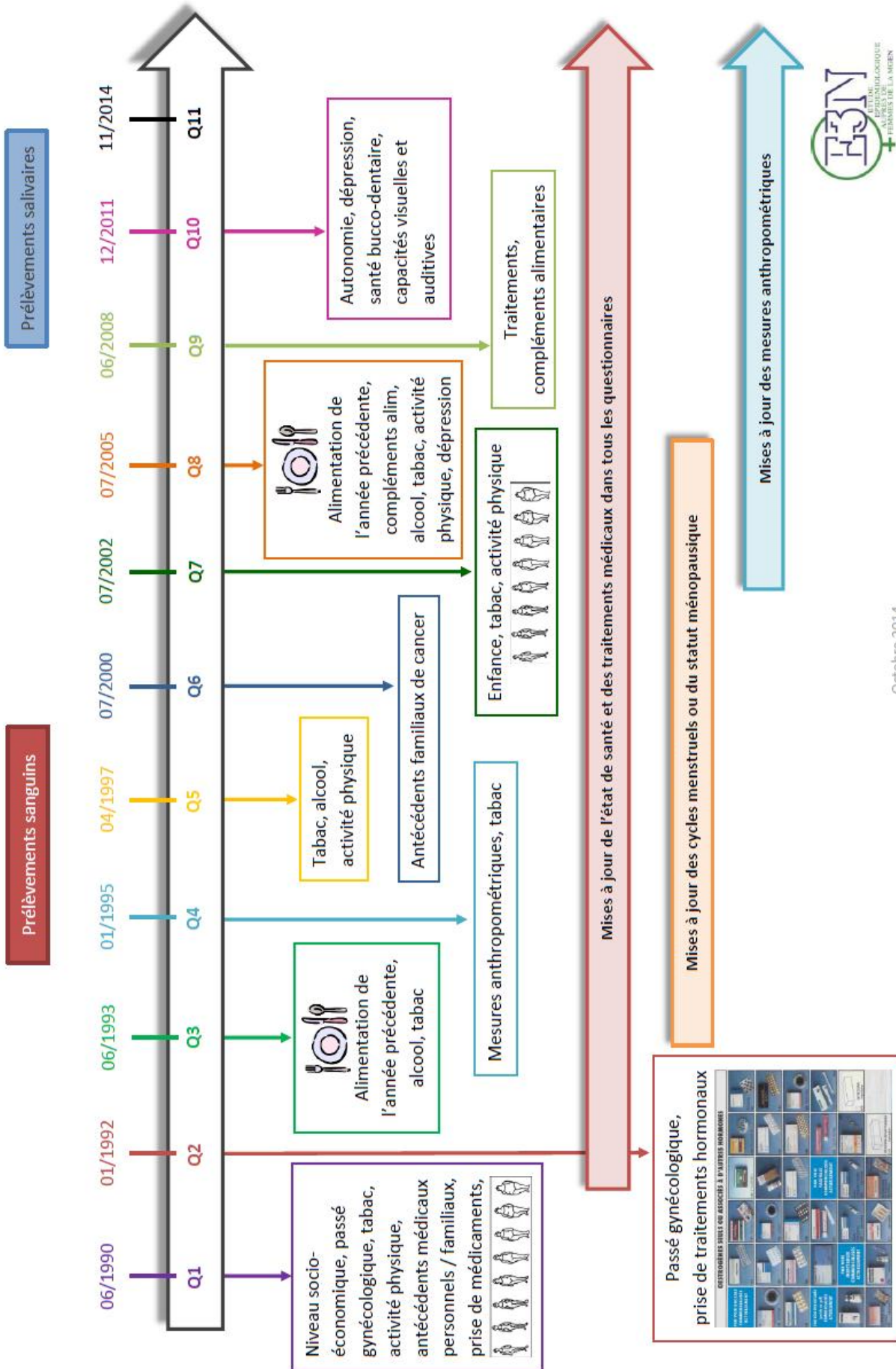
Une phase pilote, initiée en janvier 1989, a été réalisée auprès de l'ensemble des femmes résidant dans trois départements spécifiques (le Nord, le Pas-de-Calais et le Tarn-et-Garonne), affiliées à la MGEN et âgées de 40 à 65 ans. Après une seconde relance en juin 1989, sur les 28 584 femmes invitées, 2 720 ont accepté de participer en retournant le questionnaire pilote, qui avait pour but d'estimer le taux de réponse (9,5 %) et de tester la compréhension et l'acceptabilité des questions posées. L'étude nationale a été lancée en juin 1990 auprès de 499 668 femmes, initialement adhérentes à la MGEN et nées entre 1925 et 1950, dont près de 100 000 ont accepté de participer (soit un taux de réponse de 20 %, meilleur que celui de l'étude pilote). E3N a été

approuvée par la Commission nationale de l'informatique et des libertés (CNIL) et chaque femme incluse dans l'étude a signé un accord de participation au moment de son recrutement.

1.2.b. Le suivi des participantes

Depuis 1990, les femmes sont régulièrement suivies par auto-questionnaires, envoyés par courrier sous format papier, tous les 2 à 3 ans environ (figure 16). Suite à un questionnaire initial au moment du recrutement (Q1), neuf questionnaires de suivis (Q2 à Q10) ont été envoyés aux participantes de l'étude (figure 17). Le onzième questionnaire (Q11) a été envoyé en octobre 2014. Alors que les deuxième (Q2), troisième (Q3) et quatrième (Q4) questionnaires ont été adressés aux femmes ayant répondu au questionnaire précédent, les autres questionnaires (Q5 à Q11) ont été envoyés à l'ensemble des participantes. En cas de non-retour d'un questionnaire de suivi, au moins deux relances postales ont été effectuées.

Le questionnaire initial (Q1) permettait de renseigner des facteurs socio-économiques (le niveau d'études et la profession), certains paramètres anthropométriques (taille, poids et silhouettes à différents âges), des facteurs reproductifs et hormonaux (dont l'âge à la ménopause chez les femmes concernées), l'utilisation de méthodes contraceptives, l'état de santé, les éventuels antécédents personnels ou familiaux de maladies, le niveau d'activité physique et la consommation de tabac. Les questionnaires suivants (Q2 à Q10) comportaient des questions permettant de mettre à jour l'état de santé des participantes et les informations relatives au poids, à la consommation de tabac, au statut ménopausique ou à la prise de traitements hormonaux. Le deuxième questionnaire (Q2) détaillait plus précisément l'histoire reproductive et la prise de traitements hormonaux (contraceptifs oraux et traitement de la ménopause). Le quatrième questionnaire (Q4) précisait des mesures anthropométriques supplémentaires (tour de poitrine, tour de taille, tour de hanches et taille assise) ; ces différentes mesures étaient ensuite mises à jour systématiquement à partir du septième questionnaire (Q7). Les troisième et huitième questionnaires (Q3 et Q8) comportaient un questionnaire spécifique d'histoire alimentaire. Les taux de réponses à chaque questionnaire sont présentés dans le tableau 2.



Octobre 2014

Figure 16 : Chronologie des questionnaires E3N et recueil des différentes données.

MATÉRIEL - PARTIE 1

Tableau 2 : Taux de réponse et de participation aux questionnaires E3N, cohorte E3N (N = 98 995).

Questionnaire	Date d'envoi	Nombre d'envois	Nombre de retours de questionnaires exploitables ^a	Taux de réponse
Q2	Janvier 1992	98 995	86 164	87,0 %
Q3 santé	Juin 1993	93 067	78 087	83,9 %
Q3 alimentaire ^b	Juin 1993	93 067	74 524	80,1 %
Q4	Décembre 1994	75 347	69 148	91,8 %
Q5	Avril 1997	97 860	86 825	88,7 %
Q6	Juin 2000	96 605	83 196	86,1 %
Q7	Juillet 2002	96 277	81 453	84,6 %
Q8 ^b	Juillet 2005	94 613	71 411	75,5 %
Q9	Juillet 2008	92 889	73 703	79,3 %
Q10	Décembre 2011	87 751	70 519	80,4 %

a : Un questionnaire est exploitable lorsque celui-ci est renvoyé complet (aucune page manquante), dans les temps et est lisible.

b : Le questionnaire alimentaire envoyé en 1993 a fait l'objet d'un questionnaire séparé du questionnaire santé envoyé en même temps, alors qu'en 2005, lors de l'envoi du questionnaire Q8, le questionnaire alimentaire faisait suite au questionnaire santé.

Figure 17 shows the first pages of questionnaires Q1 (left) and Q6 (right). Q1 is a paper form with a header from 'MUTUELLE' and 'INSERM'. It includes a 'PERSONNELLE' identification section with fields for name, sex, and address. The main body contains various health and lifestyle questions, such as 'Quelle est la couleur qui se rapproche le plus de la couleur naturelle de vos cheveux?' and 'Avez-vous des tâches de rousseur?'. Q6 is a blue-themed paper form with a header 'E3N Questionnaire n°6' and a barcode. It features a 'TRÈS IMPORTANT' section with instructions: 'Ecrire avec un stylo noir', 'Ne pas déborder des cases', and 'Ne pas rayer si vous n'avez rien à remplir'. The form includes sections for 'Date à laquelle vous répondez à ce questionnaire', 'Renseignements généraux' (birth date, profession, weight, height), 'Transit intestinal', 'Tabac', and 'Examens pratiqués'.

Figure 17 : Première page des questionnaires Q1 (à gauche) et Q6 (à droite).

1.2.c. Les données alimentaires

L'information sur l'alimentation des participantes a été recueillie à deux reprises, à Q3 (1993) et à Q8 (2005), à l'aide d'un questionnaire d'histoire alimentaire. Ce questionnaire a été validé par des rappels de 24 heures, réalisés par deux diététiciennes expérimentées, durant lesquels les participantes ont été questionnées sur leurs consommations alimentaires de la veille (202). Ces

questionnaires ont été accompagnés d'un livret de photographies, également validé (203), facilitant les estimations des portions d'aliments ou de certaines boissons.

Le questionnaire alimentaire Q3 a été envoyé à 93 067 participantes et le taux de réponse a été de 80 %, après deux relances postales. Parmi les 78 278 questionnaires retournés, 74 524 ont pu être analysés. L'objectif était de décrire les habitudes alimentaires de l'année précédente, en tenant compte des variations saisonnières et journalières. Le questionnaire comprenait deux parties, l'une quantitative, l'autre qualitative (figure 18).

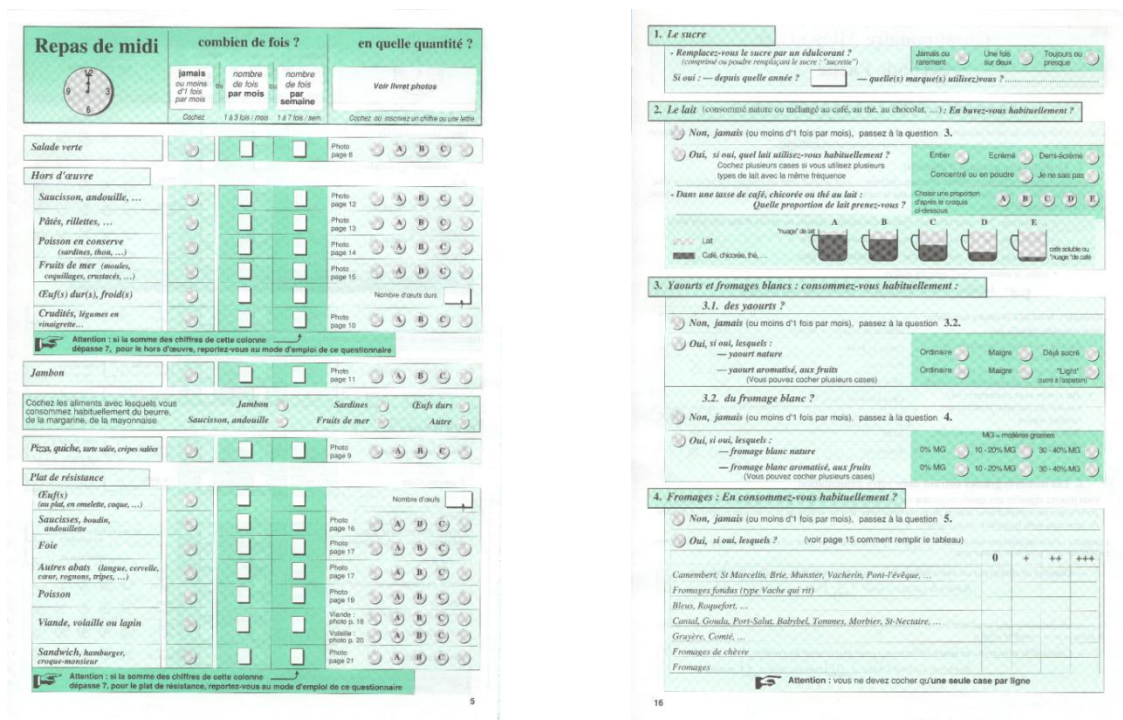


Figure 18 : Exemple de page des parties quantitative (à gauche) et qualitative (à droite) du questionnaire alimentaire Q3 (1993).

La première partie comprenait un questionnaire de fréquence quantitatif, recueillant les fréquences de consommation et les tailles de portions habituellement consommées d'une liste de 66 aliments. Le questionnaire est structuré autour de huit repas journaliers : petit-déjeuner, collation de la matinée, apéritif précédant le déjeuner, repas de midi, collation de l'après-midi, apéritif précédant le dîner, repas du soir, collation suivant le dîner. Pour chaque type d'aliment, la participante devait indiquer la fréquence de consommation parmi onze modalités* proposées (jamais ou moins d'une fois par mois, 1 à 3 fois par mois, 1 à 7 fois par semaine) et la quantité consommée, estimée soit en unité standard (nombre d'œufs, nombre de cuillères à café), soit en portions avec l'aide du livret de photographies en couleurs.

La seconde partie du questionnaire revêtait un aspect plus qualitatif. Elle visait à préciser l'alimentation du sujet en détaillant, dans chaque groupe alimentaire (tel que fruit, fromage, crudité, yaourt, fromage blanc), les types d'aliments simples que les femmes avaient l'habitude de consommer, leurs fréquences relatives de consommation (jamais, 1 à 3 fois par mois ou 1 à 7 fois par semaine), ainsi que des questions détaillées sur les matières grasses, le sucre et les modes de cuisson utilisés.

Après pondération de la première partie par la seconde partie, le questionnaire a permis d'évaluer la consommation quotidienne moyenne d'un total de 208 types d'aliments et de boissons.

1.2.d. Les données biologiques

Afin de compléter les données issues des auto-questionnaires, des données biologiques ont été collectées auprès de 25 000 volontaires, entre novembre 1994 et février 1999 (figure 16), et stockées à des fins d'analyses biologiques ultérieures. La création de cette banque de matériel biologique (BMB) a été approuvée par le Comité de protection des personnes (CPP). Pour des raisons matérielles, les femmes invitées à donner leur sang ont été sélectionnées dans des départements métropolitains, choisis selon l'effectif initial de participantes E3N (plus de 1 000 participantes) et de facilité d'organisation. Le taux de participation parmi les femmes invitées était de 37 %, représentant 24 505 prélèvements. Ces prélèvements ont ensuite été aliquotés* et séparés en 28 paillettes de 500 µL par sujet (sérum*, plasma*, couche leucoplaquettaire*, érythrocytes*) stockées dans de l'azote liquide (à -196°C) au Centre International de Recherche sur le Cancer (CIRC) à Lyon et à l'Établissement Français du Sang à Annemasse.

La constitution d'une biothèque d'échantillons de salive a été réalisée entre novembre 2009 et mai 2011, auprès des femmes de la cohorte n'ayant pas fourni de prélèvement sanguin. Cette collection, qui permettra de récupérer l'ADN des femmes participantes, compte 44 775 prélèvements salivaires. Les échantillons de salives sont stockés dans une chambre froide à une température de 4°C dans les locaux de Gustave Roussy à Villejuif.

L'ADN extrait des échantillons de salive et de sang est stocké dans un congélateur à une température de -20°C dans les locaux de Gustave Roussy à Villejuif.

1.2.e. Les données de la MGEN

Tous les trois mois, la MGEN met à disposition de l'équipe E3N des données sur le statut vital des participantes et la mise à jour de leur adresse postale. En complément, depuis le 1^{er}

janvier 2004, elle fournit également l'ensemble des fichiers de remboursement des médicaments, permettant d'avoir une indication sur l'achat par les participantes de médicaments remboursés, sans garantie, cependant, qu'elles les aient effectivement consommés.

1.3. EPIC : une dimension européenne

L'ensemble des participantes ayant renvoyé le premier questionnaire alimentaire (Q3) constitue la composante française d'une vaste étude prospective européenne : EPIC (*European prospective investigation into cancer and nutrition*). Soutenue par le programme « Europe contre le Cancer » et coordonnée par le CIRC, EPIC est aujourd'hui menée dans dix pays européens, à savoir l'Allemagne, le Danemark, l'Espagne, la France, la Grèce, l'Italie, la Norvège, les Pays-Bas, le Royaume-Uni et la Suède (figure 19), rassemblant ainsi près de 520 000 hommes et femmes (204-207). L'objectif principal d'EPIC est l'analyse des relations entre l'alimentation, le mode de vie, les facteurs environnementaux et la survenue de cancers ou d'autres pathologies chroniques. La richesse d'EPIC est de regrouper un important panel de la population européenne, permettant de comparer des habitudes alimentaires très différentes.

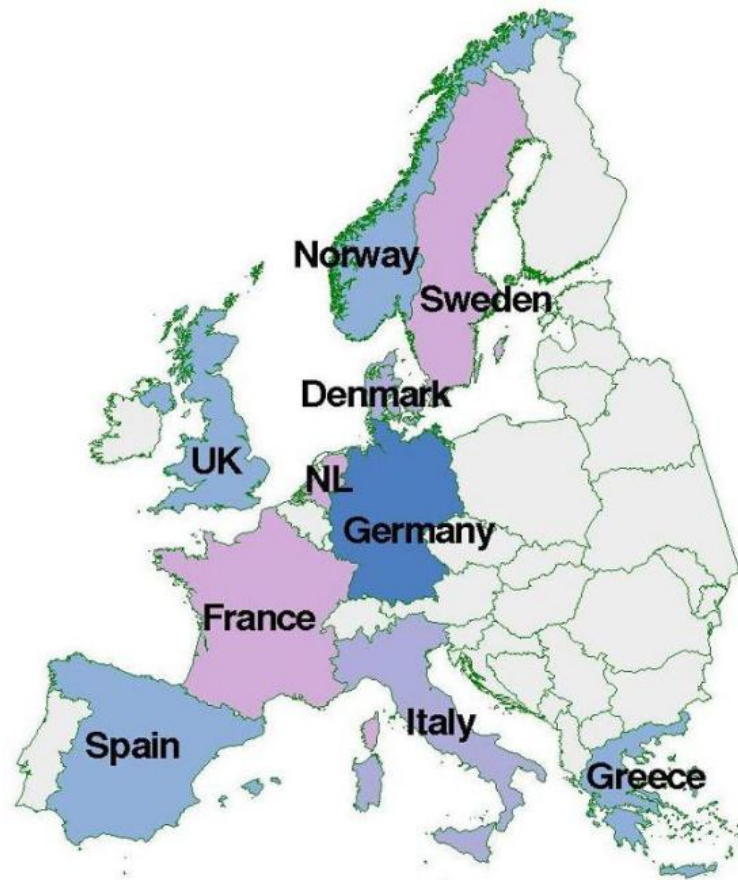


Figure 19 : Pays participants à l'étude EPIC (*European Prospective Investigation into Cancer and Nutrition*).

1.4. Perspectives : d'E3N à E4N

L'évolution majeure de la cohorte E3N consiste en la mise en place d'une nouvelle cohorte de jeunes adultes (appelée E4N pour étude épidémiologique auprès des enfants des femmes E3N), fils et filles des femmes de la cohorte E3N. Avec un taux de participation attendu de 30 %, une cohorte de 50 000 descendants de première génération des femmes E3N devrait pouvoir être constituée (sur les 170 000 enfants E3N). Par ailleurs, afin d'obtenir l'ensemble des informations sur l'environnement familial et génétique, l'inclusion de l'ensemble de leur famille (père des enfants des femmes E3N, petits-enfants et parents des petits-enfants des femmes E3N) est prévue (figure 20) et un recueil de salive sera proposé.

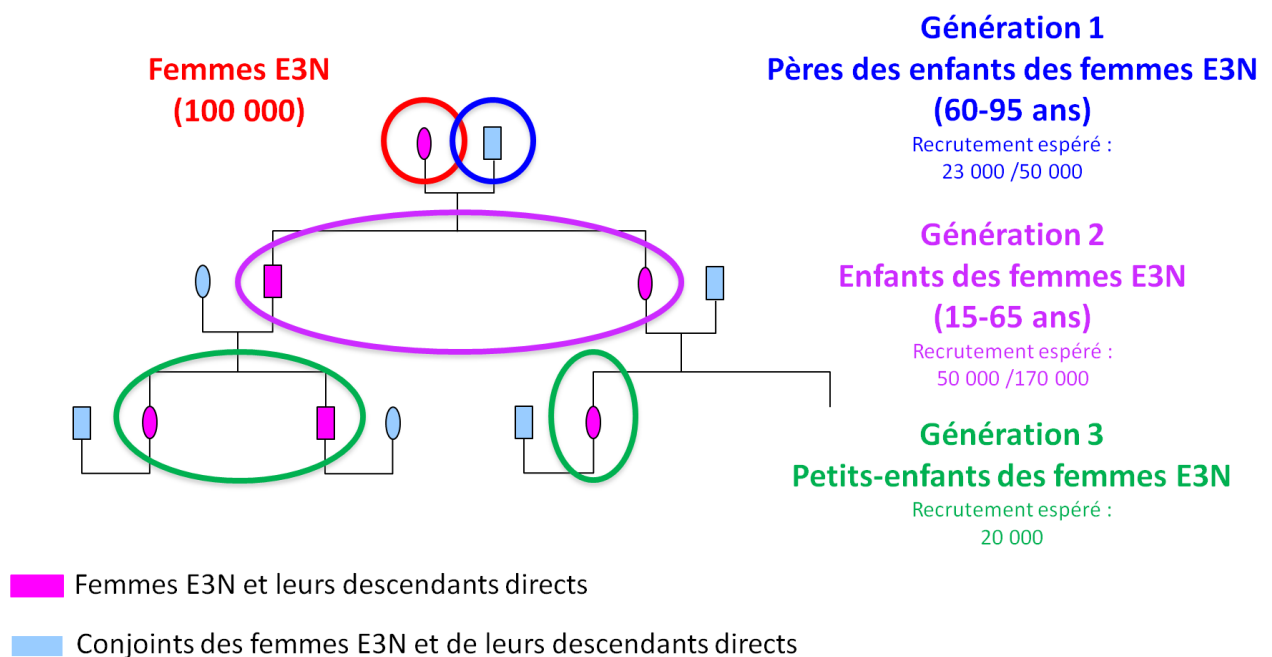


Figure 20 : E4N, étude des enfants d'E3N, une cohorte familiale.

Partie 2 Recueil de l'information sur le cancer

2.1. Les cas de cancer

Chaque questionnaire comporte une section consacrée au suivi de l'état de santé de la femme. Celle-ci y déclare, en particulier, le ou les cancers éventuels qu'elle a eus et la date de diagnostic correspondante. Chaque participante est également invitée à communiquer les coordonnées de son ou ses médecins traitants et les comptes rendus histologiques en sa possession. Le cas échéant, un courrier est adressé aux médecins traitants en vue de la confirmation des cancers déclarés et l'obtention du compte rendu histologique correspondant. Ces comptes rendus nous fournissent des informations précieuses, telles que la date de diagnostic, la localisation de la tumeur, le grade ou le degré d'envahissement. En complément, les décès par cancer sont également pris en considération pour identifier les cas de cancer non rapportés dans les questionnaires de suivis (section Matériel, partie 3 page 86).

À ce jour, les comptes rendus histologiques ont été obtenus pour 87,4 % des cas incidents de cancer identifiés entre 1990 et 2008, ce pourcentage allant de 71,3 % pour les cancers de la peau à 100 % pour les cancers de l'endomètre. Un total de 92,4 % des cas incidents de cancer du sein ont été validés. Le recueil des comptes rendus histologiques pour les cancers déclarés lors du dixième questionnaire (Q10, envoyé en décembre 2011) est en cours d'achèvement.

2.2. Classification des cancers

Le compte rendu histologique des cancers renseigne sur la localisation et la morphologie de la tumeur, codées à l'aide de trois tables de classification internationale établies par l'OMS (208;209) et l'UICC (*Union for international cancer control*) (210).

2.2.a. La CIM (*Classification internationale des maladies*)

La CIM est la norme internationale permettant d'organiser et coder les données de mortalité et de morbidité*. Elle organise et code les informations sanitaires qui sont utilisées pour les statistiques et l'épidémiologie, la prise en charge des soins, l'allocation des ressources, le suivi et l'évaluation, la recherche, les soins primaires, la prévention et le traitement. Tous les États membres de l'OMS utilisent la CIM, traduite en 43 langues, ce qui permet de fournir un langage commun pour notifier et suivre les maladies. Cela permet au monde entier de comparer et d'échanger des données de manière cohérente et normalisée, et de faciliter la collecte et le stockage des données. Plusieurs révisions de la CIM ont vu le jour, dont les trois plus récentes

sont la CIM 8 (8^{ème} révision), utilisée entre 1968 et 1975, la CIM 9 (9^{ème} révision), utilisée entre 1975 et 2000 et la CIM 10 (10^{ème} révision) utilisée depuis 2000. La onzième révision est en cours d'élaboration et la version définitive est prévue pour 2015. Les révisions successives de la CIM sont nécessaires afin de refléter au mieux les progrès accomplis dans les sciences et la pratique médicales. Par exemple, le tableau 3 présente la codification des tumeurs malignes du sein à l'aide des classifications CIM 8 à CIM 10.

Tableau 3 : Évolution de la codification des tumeurs malignes du sein de la CIM 8 à la CIM 10.

Localisation	CIM 8	CIM 9	CIM 10
Tumeur maligne du sein	174	174 (chez la femme)	C50
Mamelon et auréole	-	174.0	C50.0
Partie centrale du sein	-	174.1	C50.1
Quadrant supéro-interne	-	174.2	C50.2
Quadrant inféro-interne	-	174.3	C50.3
Quadrant supéro-externe	-	174.4	C50.4
Quadrant inféro-externe	-	174.5	C50.5
Prolongement axillaire	-	174.6	C50.6
Autres	-	174.8	-
Lésions à localisations contiguës du sein	-	-	C50.8
Sein, sans précision	-	174.9	C50.9

2.2.b. La CIM-O (Classification internationale des maladies oncologiques)

Depuis près de 25 ans, la CIM-O constitue l'ouvrage de référence pour coder les diagnostics de néoplasmes* dans les registres des tumeurs et du cancer et les laboratoires d'anatomopathologie. La CIM-O, qui en est maintenant à sa troisième révision (CIM-O-3, utilisée depuis 2000), est une double classification avec des systèmes de codage topographique et morphologique. Le code topographique décrit le point d'origine du néoplasme et utilise les mêmes catégories à trois ou quatre caractères que la section sur les tumeurs figurant dans le chapitre II de la codification CIM-10. Le code morphologique décrit les caractéristiques de la tumeur, notamment son type cellulaire et son activité biologique.

Les codes morphologiques comportent le préfixe « M », suivi de cinq chiffres, et sont compris entre M-8000/0 et M-9989/3. Les quatre premiers chiffres représentent le terme histologique précis (type de tumeur ou de cellule) et le cinquième correspond au code de comportement, indiquant si la tumeur est maligne, bénigne, *in situ* ou de caractère malin ou bénin non assuré. Un caractère numérique supplémentaire peut être ajouté au code morphologique et indique le grade histopathologique ou le degré de différenciation (degré de ressemblance entre la tumeur et le tissu

dans lequel elle a pris naissance). Les différents groupes histopathologiques de tumeurs malignes sont présentés dans le tableau 4, les différents types de comportements évolutifs d'une tumeur dans le tableau 5, et les différents grades histopathologiques dans le tableau 6.

Tableau 4 : Groupes histopathologiques des tumeurs selon la CIM-O.

Groupes de tumeurs	Code
Carcinomes épidermoïdes	M-805 à M-808, M-812, M-813
Carcinomes basocellulaires	M-809 à M-811
Adénocarcinomes	M-814, M-816, M-819 à M-822, M-826 à M-833, M-835 à M-855, M-857, M-894
Autres carcinomes particuliers	M-803 à M-871, M-815, M-817, M-818, M-823 à M-825, M-834, M-856, M-858 à M-867
Carcinomes non spécifiés	M-801, M-802
Sarcomes* et tumeurs des tissus mous	M-868 à M-871, M-880 à M-892, M-899, M-904, M-912, M-913, M-915 à M-925, M-973, M-954 à M-958
Lymphomes	M-959 à M-972
Leucémies	M-980 à M-994, M-995, M-996, M-998
Sarcome de Kaposi*	M-914
Mésothéliome*	M-905
Autres types de cancer, spécifiés	M-872 à M-979, M-893, M-895 à M-898, M-900 à M-903, M-906 à M-911, M-926 à M-936, M-938 à M-953, M-973 à M-975, M-976
Types de cancer non spécifiés	M-800, M-997

Tableau 5 : Code de comportement évolutif selon la CIM-O.

Comportement évolutif	Code
Bénin	/0
Indéterminé si bénin ou malin, à la limite de la malignité, à faible potentiel de malignité ou à potentiel de malignité non assuré	/1
Carcinome <i>in situ</i> , intra-épithélial, non-infiltrant et non-invasif	/2
Malin, siège primitif	/3
Malin, siège métastatique ou secondaire	/4
Malin, incertain si primitif ou métastatique	/9

Tableau 6 : Code pour le grade histopathologique et la différenciation selon la CIM-O.

Différenciation	Code	Grade
Bien différencié, différencié, ou sans précision	1	Grade I
Moyennement différencié, moyennement bien différencié, ou différenciation intermédiaire	2	Grade II
Peu différencié	3	Grade III
Indifférencié ou anaplasique	4	Grade IV
Cellule T	5 ^a	-
Cellule B	6 ^a	-
Cellule nulle	7 ^a	-
Cellule NK	8 ^a	-
Grade ou différenciation non déterminé(e), non cité(e) ou inapplicable	9	-

a : Les codes 5 à 8 ne sont utilisés que pour désigner la lignée cellulaire des leucémies et des lymphomes.

2.2.c. La classification TNM (Tumor, Node, Metastasis) des tumeurs malignes

Le système de classification TNM permet de se rendre compte du stade d'un cancer. La lettre T (comme tumeur) renseigne sur la taille de la tumeur, la lettre N (comme *node*, ou en français ganglion) indique si les ganglions lymphatiques ont été, ou non, envahis et la lettre M (comme métastase) signale la présence ou l'absence de métastases. Les catégories TNM sont spécifiques à la localisation de tumeur. Les classifications présentées dans les tableaux 7 et 8 ont pour but d'être les plus génériques possible, les classifications sont ensuite adaptées selon la localisation de la tumeur.

Tableau 7 : Classification anatomique d'une tumeur selon la TNM.

Classification anatomique		Signification
T : Tumeur primitive	T0	Pas d'évidence de tumeur primitive
	Tis	Tumeur <i>in situ</i>
	T1	Tumeur limitée à l'organe d'origine
	T2	Tumeur intéressant un ou plus d'un organe ou tissu adjacent à la tumeur d'origine, ou intéressant plusieurs sites du même organe
	Tx	Les données sur la tumeur primitive ne sont pas disponibles
N : Ganglions à proximité de la tumeur	N0	Absence de ganglions à proximité de la tumeur
	N1	Existence de ganglions à proximité de la tumeur
	Nx	Pas de données sur les ganglions à proximité de la tumeur
M : Métastases à distance	M0	Absence de métastases
	M1	Existence de métastases
	Mx	Pas de données sur les métastases

Tableau 8 : Stade d'une tumeur selon la TNM.

Stade	T	N	M
0	Tis	N0	M0
I	T1	N0; Nx	M0
II	T2	N0; Nx	M0
III	T1 ou T2	N1	M0
IV	T1 ou T2	N0 ou N1	M1

2.2.d. Classification des localisations de cancer utilisée dans le présent travail

La classification des différentes localisations de cancer, qui a été utilisée pour coder les cancers déclarés au cours du suivi ou pour repérer les cancers spécifiés dans les causes de décès, est présentée dans le tableau 9.

Tableau 9 : Classification des localisations de cancer utilisée dans le présent travail.

Localisation	CIM 9	CIM 10
Sein	174	C50
Col de l'utérus	180	C53
Endomètre	182	C54
Ovaires	1830-1832	C56
Gynécologique ^a , autres que col de l'utérus, endomètre et ovaires	179, 181, 1833-1839, 184-187	C51-C52, C55, C57-C58
Cutané non basocellulaire	172-173 Exclusion des CIM-O-3 M809-M811 ^b	C43-C44
Côlon-rectum	1530-1534, 1536-1538, 1540-1541	C180, C182-C188, C19-C20
Pancréas	157	C25
Digestifs ^c , autres que côlon-rectum ou pancréas	150-152, 1535, 1538, 1539, 1542-1548, 155-156, 158, 1590, 1592-1599	C15-C17, C181, C189, C22-C24, C260, C268-C269
Thyroïde	193	C73
Poumon	162	C33-C34
Hémopathies malignes ^d	200-208, 1591	C261, C81-C96
Cerveau	191-192	C70-C72
ORL ^e (Oto-rhino-laryngologue)	140-149, 160-161	C00-C14, C30-C32
Urologique ^f	188-189	C64-C68
Sarcome	170-171	C40-C41, C45-C49
Autres localisations	163-165, 190, 194-199	C21, C37-C39, C51, C69, C74-C80, C97

a : Les cancers gynécologiques incluent les tumeurs malignes de la vulve, du vagin, de l'utérus sans précision, des autres organes génitaux (sauf trompes de Fallope) et du placenta.

b : Les codes CIM-O-3 M809-811 correspondent aux tumeurs basocellulaires.

c : Les cancers digestifs incluent les tumeurs malignes de l'œsophage, de l'estomac, de l'intestin grêle, de l'appendice, du côlon sans précision, du foie, des voies biliaires, de la vésicule biliaire et du tractus intestinal et les lésions à localisations contiguës de l'appareil digestif.

d : Les hémopathies malignes incluent les tumeurs malignes de la rate et des tissus lymphoïdes, hématopoïétiques et apparentés, la maladie de Hodgkin, les LMNH, les maladies immunoprolifératives malignes, les myélomes multiples et les leucémies.

e : Les cancers ORL incluent les tumeurs malignes de la lèvre, de la cavité buccale, du pharynx, des fosses nasales et de l'oreille moyenne.

f : Les cancers urologiques incluent les tumeurs malignes des voies urinaires (rein, bassinet, uretère, vessie, urètre et glande urétrale) et les lésions à localisations contiguës des organes urinaires.

Dans l'ensemble du présent travail, les cancers cutanés basocellulaires ont été considérés comme non-cas de cancers en raison du fait que ces cancers sont très peu agressifs et ne métastasent jamais.

2.3. Codification des récepteurs hormonaux des cancers du sein

2.3.a. Le grade Scarff-Bloom-Richardson (SBR)

Le grade SBR est un facteur pronostic déterminé à partir du degré de différenciation cellulaire*, du degré d'anisonucléose* et de l'activité mitotique*. Chaque paramètre varie de 1 à 3, permettant d'obtenir un score entre 3 et 9 pour classer la tumeur en trois grades de pronostic (tableau 10).

Tableau 10 : Méthode de calcul du grade SBR.

Côte \ Critère	Degré de différenciation	Degré d'anisonucléose	Activité mitotique en un champ tumoral ^a
1	Cavités glanduliformes	Noyaux réguliers	Au maximum une mitose
2	Association de cavités et de travées	Irrégularités modérées	Au maximum deux mitoses
3	Aucune cavité glanduliforme trouvée	Nombreuses monstruosité nucléaires	Au moins trois mitoses
Score SBR	Grade	Pronostic	
3, 4, ou 5	I	Favorable	
6 ou 7	II	Moyen	
8 ou 9	III	Sombre	

a : Tient compte du champ le plus riche en mitose.

2.3.b. Les récepteurs hormonaux

Les récepteurs hormonaux déterminent l'hormonosensibilité de la tumeur et sa capacité à répondre à un traitement d'hormonothérapie. Il existe deux types de récepteurs : le récepteur à l'œstrogène (RO) et le récepteur à la progestérone (RP) (section Introduction, partie 1.2.c page 37). L'expression de ces récepteurs est déterminée par immunohistochimie* (IHC) et ne constitue pas en soi un facteur pronostique. La présence des récepteurs est plutôt un facteur prédictif de réponse à l'hormonothérapie qui, elle, confère un meilleur pronostic.

2.3.c. La protéine HER2

HER2 est une protéine, encodée par un gène situé sur le chromosome 17q12, dont la surexpression est un facteur de mauvais pronostic, mais permet toutefois un traitement par trastuzumab (Herceptin) (section Introduction, partie 1.2.c page 37). L'IHC permet d'évaluer la surexpression de cette protéine, exprimée sous forme de score de 0 à 3+ (tableau 11). L'examen de FISH (*Fluorescent in situ hybridization*), permettant de détecter une amplification du gène HER2, corrélée à la surexpression de la protéine, est effectué lorsque l'examen IHC s'avère équivoque (score de 2+).

Tableau 11 : Calcul du score HER2.

Score IHC	Coloration	Examen de FISH	Surexpression de HER2
0	Aucune coloration n'est observée ou la coloration de la membrane est observée dans moins de 10% des cellules tumorales.	-	Négative
1+	Une coloration faible ou à peine perceptible de la membrane est observée dans plus de 10% des cellules tumorales. Les cellules ne sont colorées que sur une partie de leur membrane.	-	Négative
2+	Une coloration faible à modérée de toute la membrane est observée dans plus de 10% des cellules tumorales	FISH - FISH +	Négative Positive
3+	Une coloration modérée à forte de toute la membrane est observée dans plus de 10% des cellules tumorales	-	Positive

Partie 3 Recueil de l'information sur la mortalité

3.1. La connaissance des décès

Le décès d'une femme de la cohorte peut être connu à partir de différentes sources d'information : le fichier fourni trimestriellement par la MGEN contenant, notamment, le statut vital de l'ensemble des femmes ayant donné leur accord de participation et toujours adhérentes à la MGEN, la famille (époux, enfants) ou les voisins, que ce soit par téléphone, courrier ou retour des questionnaires, les médecins ou les services postaux par retour de non réception du questionnaire.

3.2. Demande d'informations complémentaires sur les décès

Dès qu'un décès est suspecté, quelle que soit la source d'information, un courrier est envoyé à la mairie de naissance et/ou de domicile. Ce courrier permet d'obtenir des informations complémentaires sur le décès, telles que la date du décès déclarée par les autorités ou la commune de décès.

3.3. Recueil des causes de décès auprès du CépiDc⁴

3.3.a. Présentation du CépiDc

Le CépiDc est un laboratoire de l'Inserm dont les missions essentielles sont la production annuelle de la statistique des causes médicales de décès en France (540 000 décès par an), la diffusion des données et les études et recherches sur les causes médicales de décès.

Depuis 1968, le CépiDc est chargé d'élaborer annuellement la statistique nationale des causes médicales de décès en collaboration avec l'Institut national de la statistique et des études économiques (Insee). Cette statistique est établie à partir de deux documents : le certificat médical de décès et le bulletin d'État civil de décès. Le CépiDc gère une base de données comportant actuellement près de 20 millions d'enregistrements (tous les décès depuis 1968).

3.3.b. Le certificat médical de décès

L'article L. 363-1 de la loi 93-23 du 8 janvier 1993 dispose que « L'autorisation de fermeture du cercueil ne peut être délivrée qu'au vu d'un certificat, établi par un médecin, attestant le décès.

⁴ Source : <http://www.cepidc.inserm.fr/site4/> ; <http://www.certdc.inserm.fr/> (Consultés en juillet 2014)

Ce certificat, rédigé sur un modèle établi par le ministère chargé de la Santé, précise, de manière confidentielle, la ou les causes du décès à l'autorité sanitaire de la santé dans le département. Ces informations ne peuvent être utilisées que par l'État pour la prise de mesures de santé publique ou pour l'établissement de la statistique nationale des causes de décès par l'Institut national de la santé et de la recherche médicale. »

La version du certificat de décès actuellement utilisée existe depuis 1998 et comporte deux parties qui doivent donc être remplies par un médecin (figure 21). La partie supérieure comporte l'identification de la commune de décès et l'identification de la personne décédée. Cette partie nominative permet également au médecin de spécifier la date du décès, la présence ou non d'un obstacle médico-légal et diverses informations telles que la mise en bière ou le don d'organe. La partie médicale (inférieure) permet de spécifier les causes médicales du décès. Cette partie comporte certains renseignements individuels (lieu de décès, sexe, date de naissance et date de décès), les causes du décès et des informations complémentaires sur le décès.

La partie médicale du certificat de décès français est conforme au modèle international préconisé par l'OMS dans la codification CIM. Elle comprend elle-même deux parties. La première partie comporte quatre lignes qui permettent au médecin de décrire l'enchaînement causal des maladies qui ont directement conduit à la mort, de la cause immédiate rapportée sur la première ligne à la cause initiale mentionnée sur la dernière ligne. La seconde partie permet de notifier les autres états morbides qui ont pu contribuer au décès. Selon la définition de l'OMS, la cause initiale est « la maladie ou le traumatisme qui a déclenché l'évolution morbide conduisant directement au décès ou les circonstances de l'accident ou de la violence qui ont entraîné le traumatisme mortel ». La cause initiale est donc la cause sur laquelle il faut agir pour prévenir le décès et celle qui est utilisée dans les analyses épidémiologiques.

3.3.d. Le circuit administratif du certificat de décès

Après le remplissage du certificat de décès par le médecin, celui-ci est transmis à la mairie du lieu de décès, qui rédige alors deux documents : l'avis 7 bis et le bulletin 7. Le circuit administratif du certificat de décès est représenté sur la figure 22. L'avis 7 bis comporte le nom de la personne décédée et les informations d'état civil qui permettront à l'Insee de mettre à jour le répertoire national d'identification des personnes physiques (RNIPP). Le bulletin 7 comprend les mêmes informations individuelles sur la personne décédée, mais anonymisées. La mairie envoie l'avis 7 bis à l'Insee et le bulletin 7, accompagné de la partie inférieure du certificat toujours close, à la DDASS (direction départementale des affaires sanitaires et sociales) du département. Le médecin de la DDASS ouvre le certificat et prend connaissance des causes du décès (dans un contexte de surveillance des causes de décès). Le certificat est ensuite envoyé à l'Inserm, toujours accompagné du bulletin 7. Parallèlement à cette procédure, en cas de mort suspecte, le corps est transféré dans un institut médico-légal qui rédige le certificat médical de décès définitif. Ce circuit a pour objectif de garantir la confidentialité des causes de décès : l'Insee sait qui est décédé mais, ne connaît pas les causes médicales du décès, alors que l'Inserm connaît les causes de décès, mais ne sait pas qui est la personne décédée.

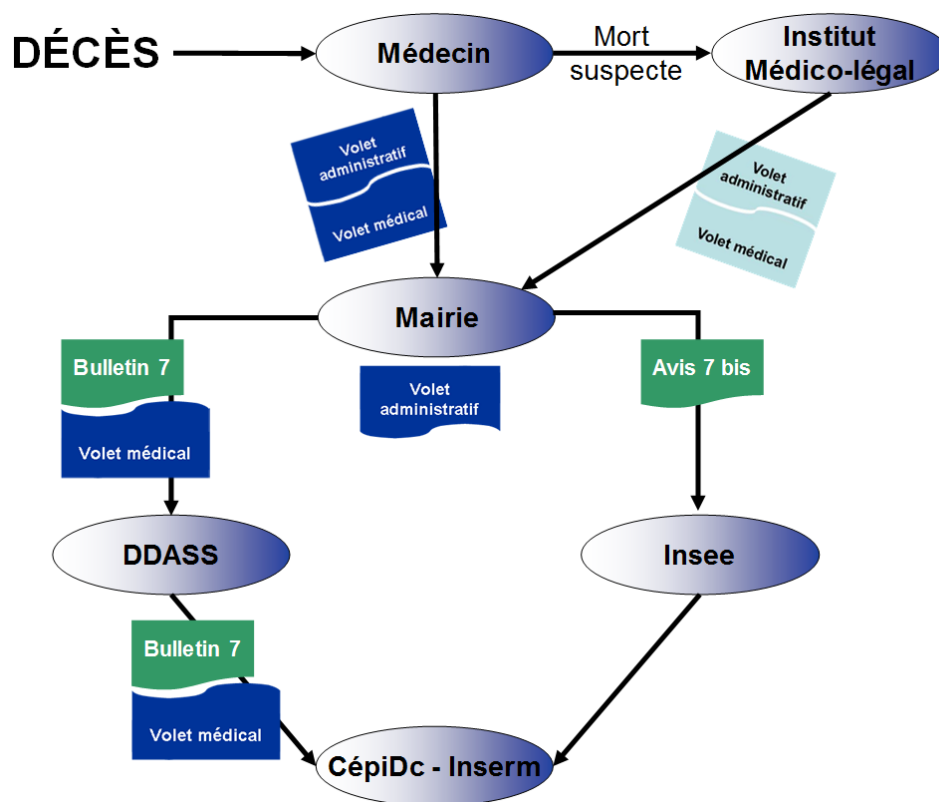


Figure 22 : Circuit d'un certificat de décès traditionnel.

Source : <http://www.cepidc.inserm.fr/site4/index.php?p=codification>

3.3.e. Du certificat papier au certificat électronique

Le délai de transmission du certificat de décès sous forme papier permet difficilement l'utilisation des données sur les causes médicales de décès dans un contexte d'alerte. Compte tenu des évolutions technologiques, le passage à un mode de certification électronique des causes médicales de décès améliorera considérablement le circuit actuel du certificat de décès. Cette orientation est d'ailleurs recommandée par le groupe de travail européen sur les causes médicales de décès piloté par Eurostat (212). Le principe de ce système consiste à donner au médecin la possibilité de certifier directement le décès sur un poste informatique en supprimant le document papier pour la partie médicale du certificat. La certification électronique, disponible en France depuis 2008, mais seulement utilisée pour à peine 5 % des certifications en 2011 (213), permettra la diminution des délais de production des indicateurs d'alerte et des statistiques de mortalité (la transmission des données saisies est presque immédiate), l'amélioration de la qualité et de la fiabilité des données en utilisant des possibilités d'aide en ligne à la certification, et le renforcement de la confidentialité des données par des procédures de chiffrement.

3.3.f. Codification des décès et extraction de la cause initiale de décès

Le travail de codification effectué par le CépiDc comporte deux tâches distinctes et successives : l'attribution d'un code à chaque maladie, traumatisme ou cause externe de décès mentionné sur le certificat de décès et la sélection de la cause initiale de décès.

La codification de l'ensemble des causes de décès spécifiées sur le certificat de décès est effectuée par la complémentarité d'un système de codage automatique (logiciel IRIS) et l'expertise des codeurs du CépiDc pour le traitement des cas complexes. Les causes des décès survenus entre 1968 et 1978 sont codées à l'aide de la huitième révision de la codification CIM, les causes des décès survenus entre 1979 et 1999 à l'aide de la neuvième révision et pour les causes des décès survenus depuis 2000 à l'aide de la 10^{ème} révision.

Pour l'ensemble des certificats, le choix de la cause initiale est régi soit par la règle générale (voir ci-après), soit par les règles 1 ou 2. La règle 3 peut, dans certains cas, s'appliquer en plus de l'une des précédentes⁵ :

Règle générale : Choisir l'affection mentionnée seule à la dernière ligne employée de la partie 1, à moins qu'il ne soit tout à fait improbable que cette affection ait pu donner naissance à toutes les autres affections mentionnées aux lignes précédentes.

⁵ Extrait du site du CépiDc www.cephidc.inserm.fr/site4/ (93)

Règle 1 : Lorsqu'il est indiqué un enchaînement aboutissant à l'affection mentionnée en premier lieu sur le certificat, choisir la cause initiale de cet enchaînement. Si plusieurs enchaînements sont indiqués, choisir la cause initiale du premier enchaînement mentionné en premier lieu.

Règle 2 : Lorsqu'il n'est indiqué aucun enchaînement aboutissant à l'affection mentionnée en premier lieu sur le certificat, choisir cette première affection comme cause initiale.

Règle 3 : Lorsque l'affection choisie par application de la règle générale ou des règles 1 et 2 peut être considérée comme la séquelle directe d'une autre affection mentionnée en I ou en II, choisir cette affection primitive comme cause initiale. Si plusieurs affections primitives sont indiquées, retenir celle qui est mentionnée en premier lieu.

La cause initiale choisie par application des règles ci-dessus n'est pas nécessairement la plus utile ou la plus intéressante aux fins des analyses sur la mortalité. Neuf règles de modifications ont donc été mises en place dans le but d'augmenter l'utilité et la précision des statistiques de mortalité et doivent être appliquées une fois que la cause initiale a été choisie conformément aux règles de sélection.

Exemple 1 : règle 4 - sénilité. Lorsque la cause initiale retenue peut être classée sous la rubrique sénilité et qu'il est signalé une affection pouvant être classée sous des rubriques autres que symptômes, signes et états morbides mal définis, il ne faut pas tenir compte de la sénilité pour choisir la cause initiale, sauf si la sénilité modifie le classement.

Exemple : **I** (a) Sénilité et pneumonie hypostatique
 (b) Arthrite rhumatoïde
 Classer à arthrite rhumatoïde.

Exemple 2 : règle 7 - causes liées. Lorsque la cause initiale retenue est liée à une ou plusieurs autres affections mentionnées sur le certificat par une disposition figurant dans la classification, ou dans les notes destinées au codage des causes initiales de mortalité, c'est d'après l'association d'affections qu'il faut classer.

Exemple : **I** (a) Dilatation cardiaque et néphrosclérose
 (b) Hypertension
 Classer à cardio-néphropathie hypertensive, combinaison des trois affections.

3.3.g. Appariement des données E3N et des données du CépiDc

Les données de la cohorte E3N et du CépiDc étant anonymes, les causes de décès des femmes E3N décédées sont obtenues par appariement sur un ensemble de 9 variables : la commune et le département de décès, le jour, le mois et l'année de décès, le sexe et le jour, le mois et l'année de naissance. Pour chaque femme décédée, les variables dont l'information est disponible dans E3N sont alors fournies au CépiDc afin qu'il effectue une recherche de correspondances. Pour chacune d'entre elles, le CépiDc nous fournit aucune, une seule ou plusieurs correspondances. Pour l'ensemble des correspondances trouvées, le CépiDc spécifie si les variables d'appariement correspondent ou si l'une d'entre elles diverge entre les données E3N et du CépiDc. Au-delà de deux informations divergentes, les correspondances ne sont pas transmises.

Les correspondances pour lesquelles une des informations est différente entre les données E3N et les données du CépiDc sont alors vérifiées au cas par cas, en s'appuyant sur les données fournies par le CépiDc (variables d'appariement, commune et département de naissance, statut marital) et les données E3N recueillies tout au long du suivi de la femme concernée.

L'information dont nous disposons sur la commune de décès et la date de décès peut manquer de précision. En effet, la commune de domicile peut s'être substituée à la commune de décès, et la date de décès fournie par la MGEN correspond à la date à laquelle la MGEN a eu l'information du décès de la femme, laquelle peut différer de quelques jours voire de quelques mois de la date officielle de décès. Pour ces raisons, une seconde étape est nécessaire afin d'optimiser nos recherches de causes de décès : celle de l'investigation des correspondances non trouvées en libérant certaines variables, telles que la commune de décès, ainsi que le jour et le mois de décès successivement.

L'ensemble de ces étapes nous a permis de recueillir les causes de 99,5 % des décès E3N survenus jusqu'en 2009, inclus.

3.4. Classification des causes de décès utilisée dans le présent travail

Dans le présent travail, la classification des différentes causes de décès prise en compte dans les analyses est présentée dans le tableau 12.

Tableau 12 : Classification des causes de décès utilisée dans le présent travail.

Cause de décès	CIM 9	CIM 10
Maladies infectieuses	001-139	A-B
Tumeurs malignes invasives	140-209	C
Tumeurs bénignes*	210-229	D10-D36
Tumeurs malignes <i>in situ</i>	230-234	D00-D09
Tumeurs malignes à évolution imprévisible	235-239	D37-D48
Maladies du sang et des organes hématopoïétiques	280-289	D50-D89
Maladies endocriniennes, nutritionnelles et métaboliques	240-279	E
Troubles mentaux et du comportement	290-319	F
Maladies du système nerveux	320-359	G
Maladies des organes des sens	360-389	H
Maladies de l'appareil circulatoire	390-459	I
Maladies de l'appareil respiratoire	460-459	J
Maladies de l'appareil digestif	520-579	K
Maladies de la peau et du tissu cellulaire sous-cutané	680-709	L
Maladies du système ostéo-articulaire, des muscles, et du tissu conjonctif	710-739	M
Maladie de l'appareil génito-urinaire	580-629	N
Grossesse, accouchement, puerpéralité	630-676	O
Certaines affections dont l'origine se situe dans la période périnatale	760-779	P
Malformations congénitales et anomalies chromosomiques	740-759	Q
Symptômes, signes et états morbides mal définis	780-799	R
Causes externes de mortalité	800-999+E800-E999	S-Y

Partie 4 Recueil des données d'exposition⁶

4.1. Les données alimentaires

Pour des informations générales sur le recueil des données alimentaires dans la cohorte E3N, se reporter à la section Matériel, partie 1.2.c page 74.

4.1.a. La consommation d'alcool

Les femmes de la cohorte ont pu estimer, lors du troisième questionnaire envoyé en 1993, leur consommation de boissons alcoolisées, en détaillant le type d'alcool et la fréquence de consommation aux différents repas proposés (apéritif avant le déjeuner, repas de midi, apéritif avant le dîner, repas du soir) (figure 23), avec l'aide d'un livret photographique (figure 24). À partir d'une table de composition alimentaire, établie à partir des données du Ciqual 1995 (214), la quantité d'alcool pur consommée quotidiennement a pu être estimée. Cette variable s'exprime en grammes d'éthanol (alcool pur) par jour (g/jour).

		combien de fois ?			en quelle quantité ?	
		jamais ou moins d'1 fois par mois	ou nombre de fois par mois	ou nombre de fois par semaine	Voir livret photos	
		Cochez	1 à 3 fois / mois	1 à 7 fois / sem.	Cochez ou inscrivez un chiffre ou une lettre	
Bière	Standard <input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 1	Nombre de verres <input type="text"/>
	"De luxe" <input type="radio"/> Sans alcool <input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 1	Nombre de verres <input type="text"/>
Cidre		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 1	Nombre de verres <input type="text"/>
Vin, champagne		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 1	Nombre de verres <input type="text"/>
Muscat, porto, vermouth, ...		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 33	Nombre de verres (= nombre de fois le volume représenté) <input type="text"/>
Whisky, gin, vodka, ...		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 33	Nombre de verres (= nombre de fois le volume représenté) <input type="text"/>
Alcool anisé (Ricard, pastis, ...)		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 33	Nombre de verres (= nombre de fois le volume représenté) <input type="text"/>
Cocktail, punch, kir, ...		<input type="radio"/>	<input type="checkbox"/>	<input type="checkbox"/>	Type de verre : Photo page 33	Nombre de verres (= nombre de fois le volume représenté) <input type="text"/>

Figure 23 : Questionnaire alimentaire de fréquence et de quantité de boissons alcoolisées consommées.

⁶ Pour les données d'exposition pour lesquelles l'information est recueillie à plusieurs questionnaires, le nettoyage des données inclue des contrôles de cohérence entre les différentes réponses fournies.

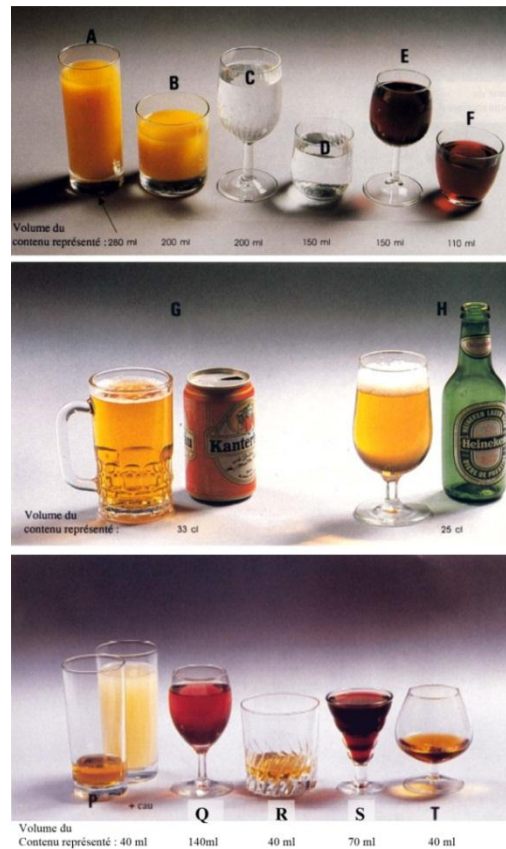


Figure 24 : Photographies d'aide au remplissage du questionnaire alimentaire. Volumes standards d'eau, de vin, de jus de fruits et de sodas (en haut), de bière (au milieu), d'apéritifs et de digestifs (en bas).

4.1.b. La consommation de fruits et légumes

Le questionnaire alimentaire envoyé en 1993 permettait aux femmes de déclarer leur consommation moyenne de fruits et légumes sur l'année précédente (figures 25 et 26), définie comme la somme des consommations en légume cuit, salade verte, crudité et légume en vinaigrette (à l'exception de la consommation de maïs, de lentille et de pomme de terre en salade, considérés comme des féculents) et fruit frais non pressé.

Repas de midi

combien de fois ?

en quelle quantité ?

jamais ou moins d'1 fois par mois ou *nombre de fois par mois* ou *nombre de fois par semaine*

Photo page 8

Photo page 10

Photo page 26

Photo page 28

5. Salade verte, crudités, légumes en salade : En consommez-vous habituellement ?

- 5.1. Non, jamais (ou moins d'1 fois par mois), passez à la question 6.

Oui, si oui, lesquels ? (voir page 15 comment remplir le tableau)

	0	+	++	+++		0	+	++	+++
Salade verte					Céleri-rave				
Endives en salade					Tomates en salade				
Poireaux en salade					Avocat				
Artichauts					Concombre				
Chou-fleur en salade					Maïs en salade				
Chou blanc ou rouge, cru					Lentilles en salade				
Carottes rapées					Pommes de terre en salade				
Radis					Autres (précisez) :				
Betteraves									

Figure 25 : Questionnaires quantitatifs (à gauche) et qualitatifs (à droite) utilisés pour estimer la consommation de fruits et légumes.

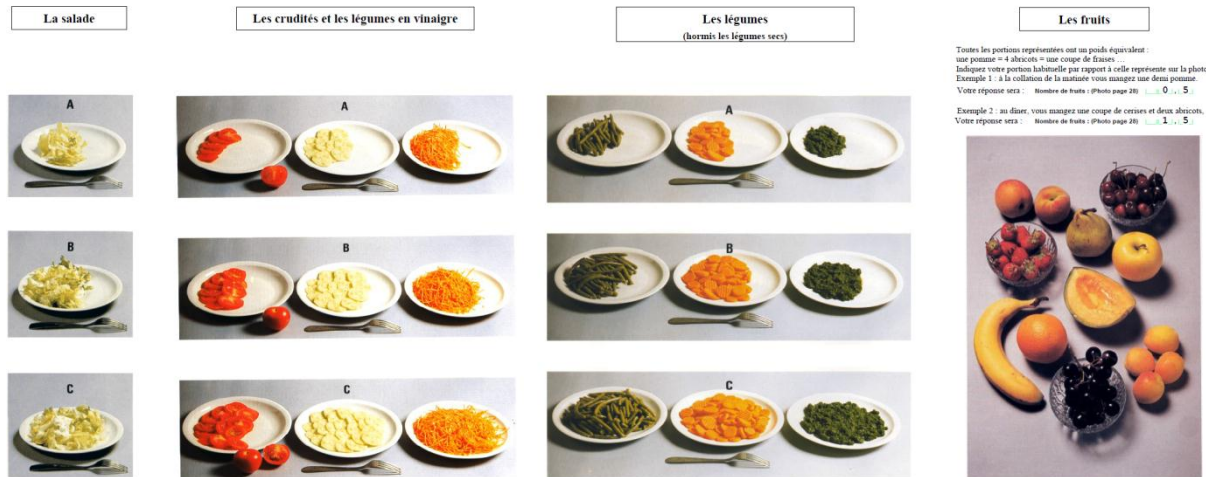


Figure 26 : Photographies d'aide au remplissage du questionnaire alimentaire. De gauche à droite : portions standards de salade verte, de crudités et légumes à la vinaigrette, de légumes cuits (hors légumes secs) et de fruits.

4.1.c. L'apport en vitamine D alimentaire

L'apport en vitamine D issue de l'alimentation a été estimé à l'aide de la table de composition alimentaire (214) et à partir de la déclaration des consommations d'aliments sources de vitamine D à Q3 (1993). Les principaux aliments contributeurs à l'apport en vitamine D sont les poissons, les produits laitiers et les œufs. Dans le cas d'analyses sur les effets des apports alimentaires en vitamine D, les hypothèses de la littérature invitent à évaluer un effet différentiel des associations selon la dose d'exposition solaire perçue par l'individu. À cet effet, nous disposons, grâce à la collaboration avec le Docteur Verdebout du Centre commun de recherche de la commission européenne, ou JRC (*Joint research centre of the european Commission*), de données mensuelles satellitaires évaluant les radiations ultraviolettes (UVA, UVB et UVC) de chaque département métropolitain français entre 1984 et 2003. À partir de celles-ci, des doses journalières moyennées sur l'année pour chaque département français ont été déterminées (figure 27) (215). L'algorithme qui a permis ces mesures tenait compte précisément de la saison et de la nébulosité*.

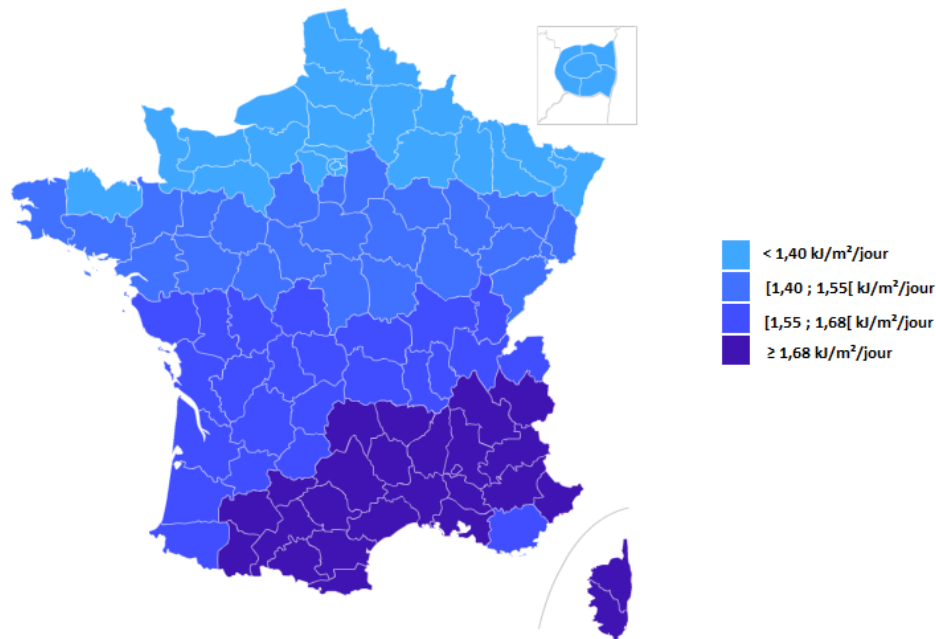


Figure 27 : Dose journalière moyenne de rayonnement solaire perçu par département français.

4.1.d. Les profils alimentaires

À partir des réponses des femmes au questionnaire alimentaire (Q3, 1993), il est possible de leur assigner des profils alimentaires évalués à partir des 57 groupes alimentaires définis dans le tableau 39 (en annexe page 217), en groupant les aliments corrélés entre eux. Une analyse en composantes principales* a été utilisée pour combiner l'information contenue par les réponses à ces 57 groupes alimentaires et extraire deux profils alimentaires, que l'on appellera « Profil alimentaire sain » et « Profil alimentaire occidental » (50). Cette analyse permet de former des combinaisons linéaires de groupes alimentaires dont les coefficients estimés pour les deux profils alimentaires, multipliés par les consommations déclarées pour le groupe alimentaire, définissent un score d'appartenance aux patterns alimentaires. Les informations concernant les deux profils alimentaires peuvent alors être combinées en créant une variable croisée à quatre catégories, en dichotomisant chaque score au niveau de la médiane observée dans la population d'analyse.

4.1.e. L'énergie consommée totale

À partir des informations issues du questionnaire alimentaire Q3 (1993) et de la table de composition alimentaire (214), l'apport énergétique (en kcal) associé à chaque type d'aliment est récupéré et multiplié par la quantité consommée par jour de l'aliment concerné. L'apport énergétique total par jour correspond à la somme de ces quantités et est exprimé en kcal/jour.

4.2. Les données anthropométriques

4.2.a. Anthropométrie à la naissance et à la puberté

Lors du questionnaire Q7 (2002), un ensemble de questions portait sur l'anthropométrie à la naissance. Il était demandé aux femmes de renseigner si elles étaient nées prématurément, c'est-à-dire à moins de 37 semaines entières de gestation. Deux questions permettaient de recueillir le poids de naissance : une question qualitative demandant aux femmes si on leur avait dit qu'elles étaient un bébé de petit, moyen ou gros poids, et une question quantitative leur demandant d'indiquer, si possible, leur poids de naissance en kilogramme. Ces deux questions ont été combinées afin de créer une variable à trois catégories de poids de naissance, donnant priorité aux informations quantitatives : petit poids ou poids inférieur à 2,5 kg, moyen poids ou poids compris entre 2,5 et 4 kg, gros poids ou poids supérieur à 4 kg. De façon similaire, une question qualitative et une question quantitative ont été posées afin de recueillir la taille à la naissance (en cm), catégorisée en trois modalités : petite taille ou taille inférieure à 48 cm, moyenne taille ou taille comprise entre 48 et 51 cm, grande taille ou taille supérieure à ≥ 51 cm.

Une estimation de la corpulence au moment de la puberté a été obtenue à Q1 (1990), grâce à l'échelle des silhouettes de Sørensen (216) (figure 28). Les femmes étaient invitées à choisir, parmi huit silhouettes, celle qui correspondait le mieux à la leur à différents âges, et en particulier au moment de la puberté (*i.e.* au moment de la ménarche). Les catégories les plus élevées (silhouettes 4 à 8) ont été regroupées à cause d'effectifs faibles.

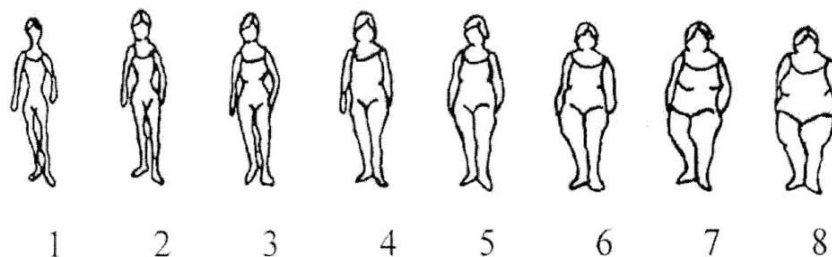


Figure 28 : Silhouettes de Sørensen (féminines).

4.2.b. Anthropométrie au cours du suivi

Les données anthropométriques ont été recueillies dans différents questionnaires (figure 16 page 73). La taille à l'âge adulte a été recueillie lors des questionnaires Q1 (1990) et Q4 (1995), puis à chaque questionnaire de Q6 (2000) à Q11 (2014). Une taille standardisée unique a été calculée pour chaque femme, correspondant à sa taille avant un éventuel tassement dû à l'âge. Le

poids a été renseigné à chacun des questionnaires Q1 à Q11, ce qui a permis le calcul de l'IMC à chaque questionnaire selon la catégorisation du tableau 1 page 42.

Une étude de validation a été effectuée auprès de 152 femmes de la cohorte, résidant en région parisienne et ayant accepté un prélèvement sanguin, afin de valider les mesures auto-déclarées des données anthropométriques le jour du prélèvement (217). La veille du prélèvement sanguin, les femmes ont rempli un auto-questionnaire spécifique dans lequel elles ont rapporté les valeurs des mesures anthropométriques (dont le poids et la taille). Le jour du prélèvement, des techniciens étaient chargés de prendre ces mêmes mesures selon un protocole standardisé. Les coefficients de corrélation de Pearson* entre les mesures auto-déclarées et cliniques étaient de 0,89 pour la taille et 0,94 pour le poids.

4.3. Autres données de mode de vie

4.3.a. Le statut tabagique

À chaque questionnaire, de Q1 (1990) à Q8 (2005) et de Q10 (2011) à Q11 (2014), les femmes ont renseigné leur statut tabagique, à savoir si elles étaient, au remplissage du questionnaire, fumeuses régulières (au moins une cigarette par jour), fumeuses occasionnelles, anciennes fumeuses régulières, anciennes fumeuses occasionnelles ou non fumeuses (n'ayant jamais fumé). À partir de Q3 (1993), l'information sur le statut tabagique est complétée par des données sur la durée depuis l'arrêt du tabac pour les anciennes fumeuses, permettant de les différencier suivant la durée depuis l'arrêt.

4.3.b. Le niveau d'activité physique

Lors du questionnaire Q3 (1993), les femmes ont renseigné leur activité professionnelle en indiquant, au cours d'une semaine type, combien d'heures elles avaient consacrées à la marche (par exemple, pour aller au travail, faire des courses), à la bicyclette (pour aller au travail ou en loisir), au bricolage, au jardinage et au sport (comme la natation, la gymnastique, le tennis), ainsi que le nombre d'étages montés à pied, en différenciant selon la saison de l'année (hiver ou été). À chacune de ces activités, un nombre de MET-h (section Introduction, partie 3.1.e page 59), établi à partir du compendium de l'activité physique d'Ainsworth (218), a été assigné. Les durées pour chaque activité déclarées en hiver et en été ont été moyennées. L'activité physique récréationnelle a été définie comme la somme de l'activité physique provenant de la marche (3 METs-h), de la bicyclette (6 METs-h) et du sport (6 METs-h). L'activité physique totale est la somme de l'activité

physique récréationnelle et de l'activité physique provenant du bricolage (4,5 METs-h), du jardinage (4 METs-h) et du nombre d'étages montés à pied (0,067 METs-h par étage).

4.4. Les données des facteurs reproductifs endogènes

4.4.a. *L'âge à la puberté*

L'âge à la puberté est défini comme l'âge aux premières règles, qui a été recueilli dans les deux premiers questionnaires en 1990 et 1992. Si les informations déclarées dans les deux questionnaires différaient de moins d'un an, la moyenne entre les âges aux premières règles déclarés était prise en compte, si la différence était strictement supérieure à un an, la valeur était considérée comme manquante.

4.4.b. *Le statut ménopausique*

La ménopause est un phénomène progressif et correspond à l'arrêt du fonctionnement ovarien, se traduisant par l'arrêt définitif des règles. Il est généralement convenu que la ménopause est confirmée après douze mois d'aménorrhée* (sauf en cas d'hystérectomie) ou après ovariectomie bilatérale*, qui, contrairement à l'hystérectomie ou à l'ovariectomie unilatérale, provoque la ménopause. À chaque questionnaire, une femme était considérée comme préménopausée si elle se déclarait non ménopausée, était réglée sans utiliser de traitement hormonal ou réglée sous pilule contraceptive ou progestatif seul. Une femme était considérée comme postménopausée si elle se déclarait ménopausée, si elle déclarait une absence de règles depuis plus d'un an sans prise de progestatif seul continu et sans hystérectomie ou si elle avait subi une ovariectomie bilatérale ou une chimiothérapie ou une radiothérapie ayant provoqué l'arrêt définitif des règles.

4.4.c. *L'âge à la ménopause*

L'âge à la ménopause est défini comme l'âge à la dernière menstruation, l'âge au moment de l'ovariectomie bilatérale, de la radiothérapie ou de la chimiothérapie ayant provoqué l'arrêt définitif des règles, l'âge de ménopause déclaré par la femme, l'âge au début d'utilisation de THM, ou l'âge aux premiers effets secondaires de la ménopause (43;44;219).

4.4.d. Le nombre d'enfants, l'âge à la première grossesse à terme et la durée d'allaitement cumulée

Le nombre d'enfants, ainsi que le nombre de grossesses menées à terme et l'âge à la première grossesse à terme ont été demandés aux femmes à Q1 (1990) et Q2 (1992). Le questionnaire Q2 permettait de compléter l'information afin d'obtenir, pour chaque grossesse, la durée d'allaitement (en mois) de l'enfant issu de la grossesse. La durée d'allaitement cumulée est la somme des durées d'allaitement de chaque grossesse. Toutefois, aucune information n'est disponible sur la quantité de lait donnée lors de chaque grossesse.

4.5. Les traitements hormonaux exogènes

Les données concernant les traitements hormonaux ont été recueillies à partir de Q2 (1992). Les femmes devaient y indiquer si elles avaient déjà utilisé des traitements hormonaux, avant ou après la ménopause, comme contraceptif oral, dans le cas de maladies du sein, de l'utérus, des ovaires ou dans tout autre cas. Elles devaient également préciser, pour chaque épisode de traitement, le nom du traitement, l'âge de début de l'utilisation et la durée d'utilisation (en mois). Pour faciliter l'effort de mémoire, ce questionnaire était accompagné d'un dépliant présentant les photographies en couleur des conditionnements des traitements commercialisés en France (figures 29 à 31). Tous les questionnaires postérieurs proposaient une mise à jour de l'utilisation des traitements hormonaux (nom des médicaments, date de début d'utilisation et durée d'utilisation). Ainsi, à partir des réponses déclarées à tous les questionnaires, l'histoire d'utilisation de contraceptifs oraux, de progestatifs seuls et de THM a pu être reconstituée jusqu'en 2008 (43;44;219). Pour chaque épisode de traitement, nous disposons de la date de début de l'épisode, de la date de fin et du nom des traitements utilisés. Nous avons pu déduire du nom des traitements les molécules utilisées, leur voie d'administration (orale, cutanée, vaginale, autre) et éventuellement, pour les THM, leur schéma d'administration (séquentiel, combiné continu, combiné discontinu).

4.5.a. Les contraceptifs oraux œstro-progestatifs⁷

Les contraceptifs oraux œstro-progestatifs peuvent être différenciés selon le type d'œstrogène et de progestatif utilisé, leur dosage et le nombre de pilules différentes utilisées sur un cycle. Les pilules utilisant un même dosage sur l'ensemble du cycle sont dites monophasiques (comme Leelo[®], Minidril[®], Melodia[®], Jasminelle[®], Zoely[®]), celles en utilisant plusieurs sont dites

⁷ Les contraceptifs pris en considération dans ce présent travail ne concernent que les pilules contraceptives et exclut tout autre forme de contraceptif hormonal (comme par exemple le stérilet avec hormones).

biphasiques (telle Adépal[®]) ou triphasiques (comme Triella[®], Triminulet[®], Qlaira[®]). Les œstrogènes utilisés dans ces pilules sont l'éthinyl-estradiol (EE) ou, pour les plus récentes, l'œstradiol (Zoely[®]) ou le valérate d'œstradiol (Qlaira[®]), tandis que de multiples progestatifs sont utilisés. Les différentes pilules peuvent être divisées en quatre générations selon le progestatif utilisé. Ainsi, les pilules de première génération utilisent de la noréthistérone (telle Triella[®]), celles de deuxième génération du lévonorgestrel (comme Leelo[®], Minidril[®] ou du norgestrel (Stediril[®]), celles de troisième génération du gestodène (comme Melodia[®]), du norgestimate (comme Cilest[®]) ou du désogestrel (comme Cycleane[®]) et les pilules de quatrième génération utilisent de la drospirénone (Jasminelle[®]), de l'acétate de cyprotéone (Diane[®]), du nomégestrol (Zoely[®]), de la chlormadinone (Belara[®]) ou du diénogest (Qlaira[®]). En complément des pilules associant un œstrogène et un progestatif, il existe également des dosages composés uniquement d'un progestatif. Celles actuellement commercialisées utilisent le lévonorgestrel (comme Microval[®]), l'acétate de noréthistérone (comme Milligynon[®]) ou le désogestrel (comme Cérazette[®]).

Les photographies des conditionnements des contraceptifs oraux œstro-progestatifs du dépliant accompagnant le questionnaire Q2 sont présentées dans la figure 29.



Figure 29 : Photographies des conditionnements des contraceptifs oraux œstro-progestatifs accompagnants le questionnaire Q2 (1992).

4.5.b. Les progestatifs seuls

Les traitements hormonaux composés de progestatifs seuls et utilisés avant la ménopause sont prescrits avec une autorisation de mise sur le marché comme contraceptif (voir partie précédente)

ou pour prévenir des douleurs mammaires, des pathologies de l'endomètre, des troubles liés à la périménopause, des maladies bénignes du sein ou des syndromes prémenstruels. Les types de progestatifs utilisés dans ces traitements incluent l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, la dydrogesterone, le lynestrénol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de nomégestrol, l'acétate de noréthistérone, la progestérone naturelle ou la promégestone.

Les photographies des conditionnements des progestatifs seuls du dépliant accompagnant le questionnaire Q2 sont présentées dans la figure 30.



Figure 30 : Photographies des conditionnements des progestatifs seuls accompagnants le questionnaire Q2 (1992).

4.5.c. Les traitements hormonaux de la ménopause

Les traitements hormonaux de la ménopause (THM) sont actuellement indiqués dans le traitement des troubles du climatère affectant la qualité de vie. Certains de ces troubles peuvent être transitoires, d'autres plus durables. À court terme, les manifestations les plus spécifiques sont vasomotrices* (bouffées de chaleur, sudation nocturne), articulaires (douleurs), génitales (sécheresse et atrophie* vaginale) et urinaires (dysurie*). Les THM se composent soit d'un œstrogène seul (chez les femmes ayant eu une hystérectomie), soit de l'association d'un œstrogène et d'un progestatif (pour prévenir la survenue d'un cancer de l'endomètre lié aux œstrogènes). Les traitements œstro-progestatifs associent un œstrogène (en France, quasi-exclusivement de l'œstradiol) à de la progestérone micronisée, à de la dydrogesterone (dérivé isomère de la progestérone) ou à un autre type de progestatif (incluant l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, le diénogest, la drospirénone, le

diacétate d'éthinodiol, le gestodène, le lévonorgestrel, le lynestrénol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de nomégestrol, l'acétate de noréthistérone ou la promégestone). La tibolone est un stéroïde d'action sélective possédant un effet œstrogénique et des actions progestative et androgénique, également utilisé comme THM. Nous avons classé les différents types de THM en fonction de la force de leur association avec le risque de cancer du sein observée dans la cohorte E3N (44;45).

Les photographies des conditionnements des THM du dépliant accompagnant le questionnaire Q2 sont présentées dans la figure 31.



Figure 31 : Photographies des conditionnements des THM accompagnants le questionnaire Q2 (1992).

4.6. Autres caractéristiques personnelles ou familiales

4.6.a. L'âge

L'âge au moment de la réponse aux différents questionnaires a été déterminé à partir de la date de naissance des femmes et de la date de réponse déclarée sur les questionnaires, ou de la date de retour si celle-ci n'était pas spécifiée.

4.6.b. Le niveau d'étude

À l'inclusion (Q1), une question visait à recueillir le niveau d'étude des participantes, regroupé en trois catégories : inférieur au Baccalauréat (Bac), Bac à Bac+2 et supérieur à Bac+2.

4.6.c. Les antécédents personnels de maladie bénigne du sein

Les données renseignant le diagnostic de maladies bénignes du sein ont été recueillies à tous les questionnaires. Les maladies bénignes du sein incluent les adénomes, les fibro-adénomes, les maladies fibro-kystiques et les mastoses. Bien qu'un antécédent de mastose ne constitue pas un facteur de risque de cancer du sein, ces maladies ont été incluses dans les maladies bénignes du sein car elles sont souvent confondues avec les maladies fibro-kystiques. Les antécédents personnels de maladie bénigne du sein à un questionnaire correspondent aux maladies déclarées lors du remplissage du questionnaire d'intérêt et des questionnaires précédents.

4.6.d. Les antécédents personnels de maladie chronique (cancer, maladie cardiovasculaire et diabète)

Les antécédents personnels de cancer, diabète et maladie cardiovasculaire ont été recueillis dans l'ensemble des questionnaires de suivi. Pour la validation des cas de cancer, se référer à la section Matériel, partie 2.1 page 79. La validation des cas de diabète a été effectuée auprès des femmes ayant déclaré, au cours du suivi, avoir un diabète, suivre un régime alimentaire antidiabétique, avoir pris des médicaments antidiabétiques, ou avoir été hospitalisée pour cause de diabète, à l'aide des données de remboursement d'antidiabétique par la MGEN à partir du 1^{er} janvier 2004 ou à partir d'un questionnaire de validation précisant les conditions de diagnostic, les traitements et le dernier taux d'hémoglobine glyquée. La validation des cas de maladie cardiovasculaire est actuellement en cours de réalisation et inclut l'infarctus du myocarde (IDM), l'accident vasculaire cérébral (AVC), la maladie veineuse thrombo-embolique (MVTE) et l'hypertension artérielle (HTA).

4.6.e. Les antécédents familiaux de cancer

Les antécédents familiaux de cancer ont été recueillis à Q1 (1990) et Q6 (2000). Dans ces questionnaires, il a été demandé aux participantes de renseigner si leur(s) père/mère, frère(s)/sœur(s), fils/fille(s) avai(en)t été atteint(s) d'un cancer du sein, de l'intestin ou d'une autre localisation, ce qui a permis de récupérer le nombre d'antécédents familiaux de cancer toutes localisations confondues et de cancer du sein en particulier, chez les parents au premier degré.

4.6.f. *Le statut marital*

À plusieurs questionnaires (Q1, Q5, Q7, Q8, Q9), les femmes précisait si elles étaient, au moment du remplissage du questionnaire, célibataires, mariées ou en couple, divorcées ou séparées, ou veuves.

4.6.g. *La zone de résidence*

Afin de déterminer la région de résidence des femmes à un questionnaire, le code postal de l'habitation de la femme lors du remplissage de ce questionnaire a été utilisé. Trois zones résidentielles (Nord, Centre et Sud) ont été définies selon la répartition présentée sur la figure 32.

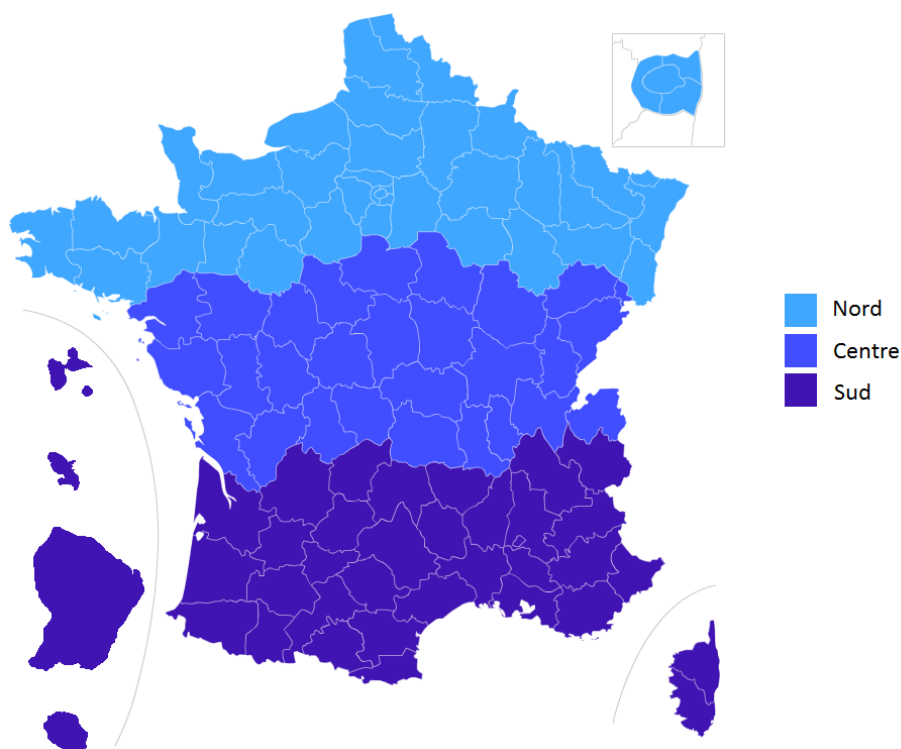


Figure 32 : Répartition en zones de résidence des départements français.

Note : La zone assignée aux territoires non présents sur cette carte est « Nord » pour Saint-Pierre et Miquelon et « Sud » pour Saint-Martin, Saint-Barthélemy, Monaco, la Polynésie française, la Nouvelle Calédonie, les îles de l'océan indien, Wallis et Futuna et l'île de Clipperton. Une valeur manquante a été assignée aux femmes résidant à l'étranger.

4.6.h. *Les antécédents de mammographie*

En 1990 (Q1), les femmes ont été invitées à renseigner si elles avaient déjà effectué une mammographie, information mise à jour dans l'ensemble des questionnaires suivants.

MÉTHODES STATISTIQUES

Toutes les analyses ont été réalisées à l'aide du logiciel SAS® (version 9.3, SAS Institute Inc., Cary, North Carolina, USA). Le seuil choisi pour déterminer la significativité des tests statistiques effectués est de 0,05.

Partie 1 Analyses descriptives

1.1. Les statistiques descriptives utilisées

Pour décrire les différentes populations, les caractéristiques des femmes ont été établies à l'inclusion des différentes analyses, soit selon le statut cas/non-cas, soit selon les modalités de la variable d'exposition principale (la variable analysée).

Dans les deux cas, si on définit un vecteur $X = (x_1, x_2, \dots, x_n)$ contenant les valeurs de n individus, les statistiques usuelles utilisées pour décrire les populations d'analyse sont les suivantes :

Pour les variables quantitatives* :

- La moyenne : $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$. Cette statistique permet de résumer l'information fournie par un ensemble de données de taille n .
- L'écart-type : $\sigma_X = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$. Cette statistique permet de mesurer la dispersion d'un ensemble de valeurs autour de la moyenne. Plus l'écart-type est faible, plus la population est homogène.
- La médiane : m telle que $P(X \leq m) = 0,5$. Cette statistique renseigne la valeur qui partage la distribution d'un ensemble de données en deux parties de même effectif.

Pour les variables qualitatives* :

- La fréquence et le pourcentage des différentes modalités.

1.2. Comparaison des statistiques descriptives entre deux échantillons

1.2.a. Comparaison de deux moyennes

La comparaison des moyennes \bar{x}_1 et \bar{x}_2 observées au sein de deux échantillons indépendants, respectivement de variances σ_1^2 et σ_2^2 et de taille n_1 et n_2 , est réalisée à l'aide du test statistique de l'écart réduit (appelé aussi test de Z), dont la statistique de test est décrite dans l'équation 1.

Équation 1 : Statistique de test associée à la comparaison de la moyenne d'une variable au sein de deux échantillons indépendants.

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

L'hypothèse nulle H_0 sous-jacente au test est l'absence de différence entre les moyennes observées des deux échantillons. Sous l'hypothèse H_0 , la statistique de test Z est approximée, lorsque n_1 et n_2 sont suffisamment grands (supérieurs ou égaux à 30), par une loi normale centrée réduite $N(0,1)$. L'hypothèse nulle est rejetée au seuil de 5 % lorsque Z est supérieure à 1,96, car $P(|Z| > 1,96) = 0,05$.

1.2.b. Comparaison de deux proportions

La comparaison des proportions p_1 et p_2 observées au sein de deux échantillons indépendants, de taille respective n_1 et n_2 , est réalisée à l'aide d'un test du χ^2 , dont la statistique de test est décrite dans l'équation 2.

Équation 2 : Statistique de test associée à la comparaison de deux proportions de deux échantillons indépendants.

$$K = \frac{|p_1 - p_2|}{\sqrt{\frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2} \left(1 - \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

L'hypothèse nulle H_0 sous-jacente au test est l'absence de différence entre les proportions observées des deux échantillons. Sous l'hypothèse H_0 , la statistique de test K est approximée, lorsque n_1 et n_2 sont suffisamment grands (supérieurs ou égaux à 30), par une loi normale centrée réduite $N(0,1)$. L'hypothèse nulle est rejetée au seuil de 5 %, lorsque K est supérieure à 1,96, car $P(|K| > 1,96) = 0,05$.

Partie 2 Analyses statistiques : le modèle de régression de Cox

2.1. Les analyses de survie

L'objectif des analyses de survie est de modéliser un phénomène dépendant du temps, à savoir la durée jusqu'à la survenue d'un événement, appelée survie, et définie par l'équation 3. La survie d'un individu à l'instant t est définie comme la probabilité pour cet individu de ne pas expérimenter l'événement d'intérêt jusqu'à un moment t .

Équation 3 : Définition de la fonction de survie.

$$S(t) = P(T \geq t)$$

Avec T la durée de suivie

À partir de cette définition de la survie d'un individu, on peut exprimer le risque instantané $h(t)$ (équation 4) comme la probabilité d'un individu de subir l'événement d'intérêt à un moment t (modélisé par l'intervalle de temps $[t; dt]$, dt étant, par définition, un intervalle de temps très petit qu'on fait tendre vers 0), sachant qu'il a survécu jusque-là.

Équation 4 : Définition de la fonction de risque instantané.

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt}$$

On peut définir la fonction de risque cumulé $H(t)$ et la relier aux autres fonctions utilisées en analyses de survie, à savoir la fonction de survie $S(t)$, la fonction de risque instantané $h(t)$ et la fonction de risque cumulée $H(t)$ par de simples équations présentées dans l'équation 5.

Équation 5 : Relations directes entre les fonctions de survie, de risque instantané et de risque cumulé.

$$h(t) = \frac{d}{dt} H(t) = -\frac{d}{dt} \ln(S(t))$$

$$\text{ou encore : } S(t) = \exp \left[-\int_{u=0}^t h(u) du \right]$$

2.2. Le modèle de régression semi-paramétrique de Cox

Le modèle de régression semi-paramétrique de Cox constitue la méthode la plus employée en épidémiologie et en biostatistique parmi l'ensemble des analyses de survie (220).

Le modèle de régression de Cox modélise la fonction de risque instantané $h(t)$ et permet l'introduction de variables explicatives (appelées également covariables). La formalisation de la fonction de risque de survenue de l'événement d'intérêt, à l'instant t , pour un individu sain jusqu'à l'instant t , et ayant le profil de covariables $X = (X_1, X_2, \dots, X_p)'$ est présentée dans l'équation 6.

Équation 6 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox.

$$h(t, X) = h_0(t) \cdot \exp(\beta' \cdot X) = h_0(t) \cdot \exp\left(\sum_{i=1}^p \beta_i \cdot X_i\right)$$

L'originalité de cette formulation est qu'elle permet d'exprimer la fonction de risque sous la forme de la multiplication d'un terme non paramétrique $h_0(t)$, appelé risque instantané de base, et d'un terme paramétrique $\exp(\beta' \cdot X)$, ce qui lui donne le nom d'approche semi-paramétrique. La transformée logarithmique du modèle fera l'objet de l'estimation proprement dite. La quantité $\ln(h(t, X))$ se compose de deux termes : un terme fonction du temps et indépendant des covariables $\alpha_0(t) = \ln(h_0(t))$, et un terme s'exprimant comme une fonction linéaire des caractéristiques individuelles des individus $\sum_{i=1}^p \beta_i \cdot X_i$ (équation 7).

Équation 7 : Formalisation de la transformée logarithmique du risque instantané par le modèle de régression de Cox.

$$\ln(h(t, X)) = \alpha_0(t) + \beta' \cdot X = \alpha_0(t) + \sum_{i=1}^p \beta_i \cdot X_i$$

L'estimation des paramètres β s'effectue en maximisant la vraisemblance partielle (équation 8), qui dépend du temps T_i , de la censure* δ_i (égale à 0 si l'individu est censuré, 1 sinon), du vecteur de covariables $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ de chaque individu i , et du vecteur de coefficients de régression théoriques β que l'on cherche à estimer.

Équation 8 : Fonction de vraisemblance partielle d'un modèle de régression de Cox.

$$PL_n[(T_i, \delta_i, X_i)_{1 \leq i \leq n}, \beta] = \prod_{i=1}^n \left[\frac{\exp(\beta' \cdot X_i)}{\sum_{j \in R_{T_i}} \exp(\beta' \cdot X_j)} \right]^{\delta_i}$$

Avec R_{T_i} le nombre de personnes à risque (individus non censurés n'ayant pas encore déclaré l'événement d'intérêt) à T_i . L'estimateur de β , que l'on note communément $\hat{\beta}$, obtenu par la

maximisation de la fonction de vraisemblance partielle converge en probabilité vers β ($\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P(|\hat{\beta}_n - \beta| > \varepsilon) = 0$) et est asymptotiquement normalement distribué.

2.3. Les hypothèses d'application du modèle de régression de Cox

La modélisation de Cox repose sur deux hypothèses : la proportionnalité des rapports de risque et la log-linéarité entre la fonction de risque instantané et les covariables.

2.3.a. L'hypothèse des risques proportionnels

L'équation 6 se traduit par un rapport de risques entre deux individus qui doit être indépendant du temps. En effet, si on prend en considération deux individus de profils $X_a = (x_{a1}, x_{a2}, \dots, x_{ap})'$ et $X_b = (x_{b1}, x_{b2}, \dots, x_{bp})'$, alors le rapport des risques entre ces deux individus n'est pas fonction du temps t (équation 9).

Équation 9 : Rapport de risques entre deux individus de profils différents.

$$\frac{h(t, X_a)}{h(t, X_b)} = \frac{\lambda_0(t) \cdot \exp(\sum_{i=1}^p \beta_i \cdot X_{ai})}{\lambda_0(t) \cdot \exp(\sum_{i=1}^p \beta_i \cdot X_{bi})} = \exp\left(\sum_{i=1}^p \beta_i \cdot (X_{ai} - X_{bi})\right)$$

Le logarithme du rapport de risques entre deux individus s'exprime donc selon un terme linéaire. Les risques des individus, conditionnellement aux covariables introduites, sont censés connaître des évolutions parallèles au cours du temps. Plusieurs méthodes permettent d'appréhender l'hypothèse de proportionnalité des risques, qui doit être testée pour chaque covariable incorporée dans le modèle de régression de Cox.

Représentation graphique de $\ln(-\ln(S(t)))$ en fonction de $\ln(t)$

D'après les équations 5 et 6, on peut en déduire les équations successives suivantes (équation 10) :

Équation 10 : Expressions de $S(t)$ et de $\ln(-\ln(S(t)))$.

$$\begin{aligned} S(t, X) &= \exp\left[-\int_{u=0}^t h(u) du\right] \\ &= \exp\left[e^{\beta' \cdot X} \cdot \int_{u=0}^t -\lambda_0(u) du\right] \\ &= [S_0(t)] e^{\beta' \cdot X} \end{aligned}$$

On en déduit donc :

$$\begin{aligned}\ln(-\ln(S(t, X))) &= \ln(-\ln([S_0(t)]^{e^{\beta' \cdot X}})) \\ &= \beta' \cdot X + \ln(-\ln(S_0(t)))\end{aligned}$$

et que, pour deux modalités X_1 et X_2 de la variable X , on a l'expression :

$$\ln(-\ln(S(t, X_1))) - \ln(-\ln(S(t, X_2))) = \beta' \cdot (X_1 - X_2)$$

Ainsi, si la variable X respecte l'hypothèse de proportionnalité des risques, alors, pour chaque modalité X_m de la covariable X , les courbes $\ln(-\ln(S(t, X_m)))$, tracées en fonction du temps, doivent présenter un parallélisme important.

Comparaison entre les processus de martingale empirique et simulés

Cette vérification consiste à comparer les résidus de martingales (221) observés d'une covariable selon le temps (appelé processus de martingale empirique) et plusieurs distributions de résidus de martingale pour cette même covariable au cours du temps produites par simulation et ré-échantillonnages successifs (appelés processus de martingale simulés). Si l'écart entre le processus empirique et les processus simulés, pouvant être évalué statistiquement et graphiquement, est important alors l'hypothèse de proportionnalité des risques n'est pas vérifiée. Le test sous-jacent permettant de tester statistiquement l'écart entre les deux processus repose sur un test de type Kolmogorov-Smirnov.

Que faire lorsqu'une covariable ne respecte pas l'hypothèse de proportionnalité des risques

Lorsqu'une covariable introduite dans le modèle de Cox ne respecte pas la proportionnalité des risques, il convient alors soit de ne rien faire et considérer que les risques estimés sont moyennés sur la période d'observation, soit de prendre en compte la covariable comme dépendante du temps, soit d'effectuer la modélisation sur plusieurs intervalles de temps en estimant autant de coefficients de régression que d'intervalles, soit de stratifier le modèle selon la covariable ne respectant pas l'hypothèse de proportionnalité des risques en modélisant un risque instantané de base différent pour chaque modalité de la variable concernée (les coefficients issus de la régression sont uniques et identiques à travers les strates).

2.3.b. L'hypothèse de log-linéarité des covariables

Si la covariable X_j est incorporée en tant que variable continue dans le modèle, il faut vérifier que la fonction de risque instantané augmente log-linéairement avec le niveau de cette covariable (équation 11).

Équation 11 : Équation sous-jacente à l'hypothèse de log-linéarité d'une covariable.

$$h(t, X_j, X_{i, i \in [1, p], i \neq j}) = h_0(t) \cdot \exp\left(\beta_j \cdot X_j + \sum_{\substack{i=1 \\ i \neq j}}^p \beta_i \cdot X_i\right)$$

$$\text{Donc : } \ln\left(h(t, X_j, X_{i, i \in [1, n], i \neq j})\right) = \ln(\lambda_0(t)) + \beta_j \cdot X_j + \sum_{\substack{i=1 \\ i \neq j}}^n \beta_i \cdot X_i$$

Par exemple, si on considère comme covariable l'âge à la ménopause, incorporée en continu dans un modèle de régression de Cox, alors le risque associé à l'augmentation de 5 années pour l'âge à la ménopause, toutes choses égales par ailleurs (*i.e.* les valeurs de l'ensemble des autres covariables sont identiques entre les individus), est multiplié par $\exp(5 \cdot \beta_{\text{âge ménop}})$. Cette augmentation de risque doit être la même entre deux femmes ayant des âges à la ménopause de 45 et 50 ans, qu'entre deux femmes ayant des âges à la ménopause de 55 et 60 ans.

Pour vérifier cette hypothèse, il convient de modéliser le risque associé à la covariable concernée en la catégorisant selon ses quartiles de distribution. Si, en traçant les risques obtenus pour chaque quartile en fonction du centre de chaque classe de quartile, on obtient une droite, l'hypothèse de log-linéarité est respectée. Dans le cas contraire, il convient de ne pas incorporer la variable en continu dans le modèle et de la catégoriser en respectant les exigences médicales, la précision des données et la pertinence statistique.

2.4. Le rapport de risques ou *hazard ratio*

L'équation 9 permet d'évaluer le rapport de risques entre deux individus de profils de covariables différents. Ce rapport est appelé *hazard ratio** et peut être estimé pour chaque covariable. En effet, si on reprend l'équation 9 et que l'on considère que les profils des deux individus (Z_a et Z_b) ne diffèrent que par rapport à la variable X_1 , dont les valeurs pour les deux individus sont respectivement X_{a1} et X_{b1} , alors on peut estimer le rapport de risques, ou *hazard ratio* (HR), de la modalité X_{a1} par rapport à la modalité X_{b1} de la variable X_1 , toutes choses égales par ailleurs (*i.e.* $Z_a = Z_b$) :

Équation 12 : Définition de l'hazard ratio entre deux modalités d'une covariable.

$$\begin{aligned} HR(X_{a1} \text{ vs. } X_{b1}) &= \frac{h(t, X_{a1}, Z_a)}{h(t, X_{b1}, Z_b)} = \frac{h_0(t) \cdot \exp(\beta_1 \cdot X_{a1} + \sum_{i=1}^p \beta_i \cdot Z_{ai})}{h_0(t) \cdot \exp(\beta_1 \cdot X_{b1} + \sum_{i=1}^p \beta_i \cdot Z_{bi})} \\ &= \exp(\beta_1 \cdot (X_{a1} - X_{b1})) \end{aligned}$$

L'HR associé à la modalité X_{a1} par rapport à la modalité X_{b1} s'interprète comme le facteur multiplicatif caractérisant l'effet de l'exposition X_{a1} par rapport à l'effet de l'exposition X_{b1} sur le risque de survenue de l'événement d'intérêt.

2.5. La fonction de risque instantané de base

Dans le cas de l'analyse de l'effet de l'exposition à un facteur par rapport à la non-exposition à ce même facteur sur le risque de survenue de l'événement d'intérêt, seule l'estimation de l'HR est utile. Cependant, si on veut prédire le risque de survenue de l'événement d'intérêt, l'estimation de la fonction de risque instantané de base $h_0(t)$ est indispensable. L'estimation de $h_0(t)$ s'appuie sur l'estimateur de Breslow de la fonction de risque cumulée (222) (équation 13).

Équation 13 : Estimateur de Breslow pour la fonction de risque cumulée de base.

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{d_i}{\sum_{j \in T_{t_i}} \exp(\beta' \cdot X)}$$

Avec d_i le nombre d'événements survenus à l'instant t_i .

Cet estimateur permet d'obtenir les valeurs de la fonction de risque cumulée de base pour chaque temps d'événement rencontré dans la population d'analyse. Une modélisation de cette fonction a été faite par spline cubique (223). Le principe général des fonctions splines est de découper les valeurs d'une variable continue (ici $\hat{H}_0(t)$) en $(N+1)$ intervalles, en définissant N nœuds (les bornes des intervalles) que l'on notera n_n . Au sein de chaque intervalle, la relation entre la variable continue et la variable explicative (ici le temps t) est modélisée par un polynôme de degré 3 (pour un spline cubique) à l'aide d'un modèle de régression (ici linéaire) de façon à estimer les coefficients (notés ici β) qui ajustent le mieux la relation entre les deux variables (équation 14).

Équation 14 : Modélisation par spline de la fonction de risque cumulée de base.

$$\hat{H}_0(t) = \alpha + \sum_{i=1}^{N+2} \beta_i \cdot S_i(t)$$

$$\text{Avec } S_i(t) = \begin{cases} t & \text{si } i = 0 \\ t^2 & \text{si } i = 1 \\ t^3 & \text{si } i = 2 \\ (t - n_{i-2})^3 & \text{si } i > 2 \text{ et } t \geq n_{i-2} \\ 0 & \text{sinon} \end{cases}$$

La modélisation de la fonction de risque cumulée de base peut être obtenue à partir d'un modèle vide (sans variable explicative) ou à partir d'un modèle multi-ajusté. La fonction de risque instantané et la fonction de survie de base s'obtiennent ensuite à l'aide des formules fournies dans l'équation 5 à partir de la modélisation de la fonction de survie cumulée de base.

Généralement, un nombre N de nœuds inférieur à 10 est utilisé, choix qui dépend de la taille de l'échantillon et de la distribution de la variable à modéliser. Le nombre de nœuds optimal correspond à celui qui minimise le critère d'Akaike (AIC), défini par l'équation 15. Les nœuds sont souvent placés aux percentiles de la variable à modéliser par spline, ce qui permet d'équilibrer les effectifs dans les intervalles. Dans le cas de spline cubique restreint, il est recommandé d'écarter les nœuds extrêmes de façon à mieux modéliser la courbe pour les petites et grandes valeurs de la variable à modéliser (tableau 13) (224).

Équation 15 : Formulation du critère d'information d'Akaike.

$$AIC = 2 \cdot k - 2 \cdot \ln(L)$$

Avec k le nombre de paramètres du modèle et L la vraisemblance du modèle.

Tableau 13 : Définition du placement des nœuds pour la modélisation par spline restreint.

Nombre de nœuds	Percentiles où les nœuds sont placés						
3	10	50	90				
4	5	35	65	95			
5	5	27,5	50	72,5	95		
6	5	23	41	59	77	95	
7	7,5	18,33	34,17	50	65,83	81,67	97,5

2.6. Les variables dépendantes du temps

Un des avantages de la modélisation de Cox est la possibilité d'introduire des variables explicatives dont les valeurs évoluent dans le temps. Cette extension du modèle de régression de Cox permet une mise à jour de ces variables au cours du suivi à chaque temps d'événement. La formalisation de la modélisation de Cox présentée dans l'équation 6 est modifiée dans l'équation 16 pour prendre en compte cette extension.

Équation 16 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox en présence de variables explicatives dépendantes du temps.

$$h(t, X(t)) = h_0(t) \cdot \exp(\beta' \cdot X(t)) = h_0(t) \cdot \exp\left(\sum_{i=1}^p \beta_i \cdot X_i(t)\right)$$

2.7. L'âge comme échelle de temps

Dans un modèle de Cox, la variable correspondant au temps est un délai jusqu'à la survenue de l'événement d'intérêt, caractérisé par le début de l'exposition au risque, le début de l'observation et la fin de l'observation ou de l'exposition au risque.

Dans les études de cohorte, le début de l'exposition et le début de l'observation sont rarement identiques car, pour un grand nombre de maladies, les individus sont à risque de développer la maladie dès leur naissance, mais ne sont observés qu'à partir de l'inclusion dans l'étude. La considération de l'âge comme échelle de temps permet de fixer le début de l'exposition à la date de naissance de l'individu (âge nul) et le début de l'observation à l'âge d'inclusion dans l'étude (on parle alors d'entrée retardée).

Un autre avantage de cette considération est qu'aucune hypothèse paramétrique n'est imposée à la variable âge puisque sa modélisation est prise en compte dans la partie non-paramétrique de la fonction de risque instantané. En s'affranchissant de cette hypothèse paramétrique, on évite d'aboutir à des estimations erronées qui pourraient résulter de la non validation des hypothèses d'application du modèle de Cox (225;226). Plusieurs auteurs recommandent donc l'utilisation de l'âge en échelle de temps dans l'analyse des enquêtes de cohortes étudiant la survenue d'un événement dont l'incidence est fortement déterminée par l'âge (225;227).

2.8. Le test de tendance

Lorsqu'une variable continue est catégorisée, il est possible de tester l'existence d'une tendance pour l'association entre cette variable et le risque étudié. Les tests de tendance présentés dans ce travail sont issus de modèles incorporant la variable en continu (*i.e.* la variable d'origine avant qu'elle soit catégorisée).

2.9. L'effet cohorte

Le risque de développer un cancer n'est pas le même pour une femme selon son année de naissance, effet particulièrement marqué pour les cancers du sein, du poumon et du col de

l'utérus, et peut être observé tant en terme d'incidence que de mortalité (figure 33). Par exemple, le risque de développer un cancer du sein pour une femme née en 1945 (9,7 %) est pratiquement le double que pour une femme née en 1920 (5,8 %), mais le risque est plus faible pour une femme née en 1950 (9,1 %) (3). Ce phénomène, appelé « effet cohorte », doit être pris en compte dans la modélisation des risques liés au cancer ou à la mortalité en attribuant aux différentes générations des risques de base différents (équation 17).

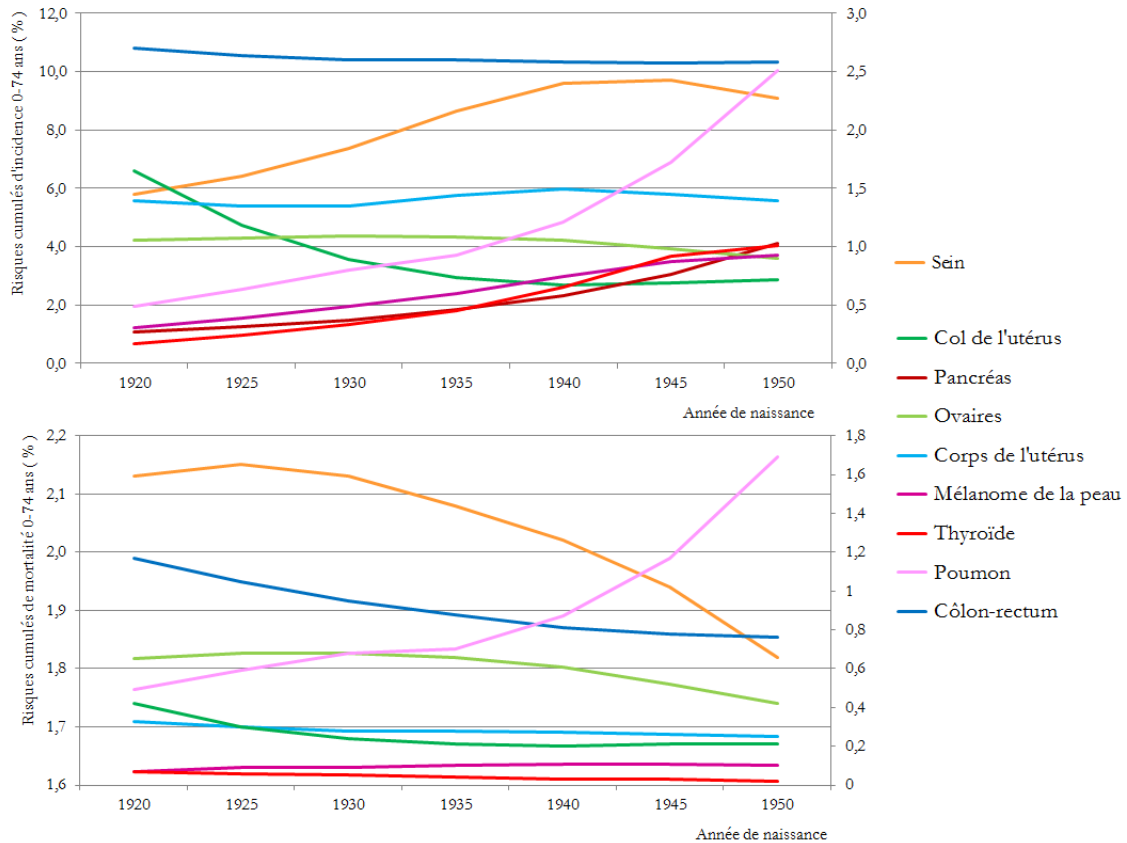


Figure 33 : Risques cumulés d'incidence (en haut) et de mortalité (en bas) chez les femmes âgées de 0 à 74 ans en France métropolitaine entre 1980 et 2012 selon la cohorte de naissance (1920 à 1950).

Note : Les axes des ordonnées de gauche correspondent aux risques cumulés d'incidence et de mortalité pour le cancer du sein. Les axes des ordonnées de droite correspondent aux risques cumulés d'incidence et de mortalité pour les autres localisations de cancer. Les données sur les lymphomes non hodgkinien ne permettent pas d'évaluer l'évolution des taux d'incidence et de mortalité à cause de l'hétérogénéité du recueil des données au cours du temps (4).

Source : Binder-Foucard *et al.* 2013 (3)

Équation 17 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox dans un modèle stratifié.

$$h_j(t, X) = h_{0j}(t) \cdot \exp(\beta' \cdot X) = h_{0j}(t) \cdot \exp\left(\sum_{i=1}^p \beta_i \cdot X_i\right)$$

Avec $h_j(t, X)$ et h_{0j} la fonction de risque instantané et la fonction de risque instantané de base dans la strate j .

2.10. L'analyse en risques compétitifs

Une situation à risques compétitifs survient lorsqu'un individu peut expérimenter plusieurs types d'événements dont l'occurrence de l'un empêche la survenue des autres. Par exemple, lorsque l'on étudie le risque de survenue d'un premier cancer selon les différentes localisations de la tumeur, on est en situation de risques compétitifs étant donné que la survenue d'un premier cancer localisé dans le sein empêche l'individu d'avoir un premier cancer localisé à un endroit différent.

Mathématiquement, la formalisation d'une situation à risques compétitifs impose la prise en considération d'une variable de censure C à autant de modalités qu'il y a de causes en compétition (alors que cette variable est binaire en absence de risque compétitif). En présence d'un nombre k de causes en compétition, la variable C vaudra 0 si l'individu est censuré (n'expérimente aucune des causes en compétition au cours de son suivi) ou la valeur k ($k \in [1; c]$) si l'individu expérimente la cause k au cours de son suivi. La définition de la fonction de risque instantané en présence de risques compétitifs est donnée dans l'équation 18.

Équation 18 : Définition de la fonction de risque instantané en présence de risques compétitifs.

$$h_k(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt, C = k | T \geq t)}{dt}$$

À ce jour, deux méthodes sont proposées pour modéliser un modèle de régression de Cox en présence de causes en compétition : modéliser séparément le risque de survenue de chaque cause i en censurant les autres causes en compétition (228), ou modéliser un seul et même modèle incorporant les différents types de censures (229). La seconde méthode impose une duplication des données autant de fois qu'il y a de types de censure différents. Pour la duplication k ($k \in [1; C]$), la variable de censure c a pour valeur k , et la valeur de chaque covariable X est remplacée par $\delta_c \cdot X$, avec $\delta_c = 1$ si $c = k$, 0 sinon. Le modèle de régression de Cox est stratifié sur la variable de censure, permettant une fonction de risque instantané de base $h_{0k}(t)$ différent selon le type de censure (équation 17). Cette méthode de duplication repose sur une hypothèse d'additivité des fonctions de risque instantané et d'indépendance des causes en compétition (équation 19). Cette méthode de duplication a pour avantage de limiter la surparamétrisation (229) et est celle utilisée dans le présent travail.

Équation 19 : Formalisation de la modélisation du risque instantané par le modèle de régression de Cox en présence de risques compétitifs.

$$h_k(t, X) = h_{0k}(t) \cdot \exp(\delta \cdot \beta' \cdot X) = h_{0k}(t) \cdot \exp\left(\sum_{c=1}^c \sum_{i=1}^p \delta_c^k \cdot \beta_{ic} \cdot X_i\right)$$

Avec $\delta_c^k = 1$ si $c = k$, $\forall i \in [1; p]$

Le rapport de risques entre les profils X_a et X_b , pour la cause k , s'exprime alors selon la formule présentée dans l'équation 20.

Équation 20 : Rapport de risques entre deux individus de profils différents en présence de risques compétitifs.

$$\begin{aligned} HR_k(X_a \text{ vs. } X_b) &= \frac{h_k(t, X_a)}{h_k(t, X_b)} = \frac{h_0(t) \cdot \exp(\sum_{c=1}^c \delta_c \cdot \beta_c \cdot X_a)}{h_0(t) \cdot \exp(\sum_{c=1}^c \delta_c \cdot \beta_c \cdot X_b)} \\ &= \exp\left(\sum_{c=1}^c \delta_c \cdot \beta_c \cdot (X_a - X_b)\right) \\ &= \exp(\beta_k \cdot (X_a - X_b)) \end{aligned}$$

Car, pour la cause $c = k$, $\delta_c = 1$, et $\forall c \neq k$, $\delta_c = 0$.

2.11. La prédiction à cinq ans du risque

2.11.a. Estimation du risque prédit à cinq ans

Le risque de survenue de l'événement d'intérêt dans la fenêtre de prédiction de cinq ans est défini comme la probabilité de survenue de l'événement d'intérêt dans les cinq ans suivant le début de suivi de l'individu, sachant que l'individu est sain au début du suivi.

En se basant sur une modélisation de Cox et sur la formulation présentée dans l'équation 5, il en découle la formulation de la survie à cinq ans, puis du risque à cinq ans (équation 21).

Équation 21 : Estimation de la survie et du risque à cinq ans.

$$\begin{aligned} S(5, X) &= \frac{[S_0(5)]^{\exp(\sum_{i=1}^p \beta_i \cdot X_i)}}{S_0(0)} \\ P(5, X) &= 1 - \frac{[S_0(5)]^{\exp(\sum_{i=1}^p \beta_i \cdot X_i)}}{S_0(0)} \\ &= 1 - \exp\left[\ln\left(\frac{S_0(5)}{S_0(0)}\right) \cdot \exp\left(\sum_{i=1}^p \beta_i \cdot X_i\right)\right] \end{aligned}$$

Par définition, étant donné qu'au début du suivi les individus sont sains, on a $S_0(0) = 1$. La valeur $S_0(5)$ peut être obtenue par une modélisation par spline de la fonction de survie de base (équation 14), estimée à partir de l'estimateur de Breslow de la fonction de risque cumulée de base (équation 13).

2.11.b. Validation d'un modèle prédictif

Une étape primordiale dans la construction d'un score prédictif est la validation de ce score afin de juger si celui-ci est correctement construit et est fidèle à la réalité. Deux mesures sont principalement évaluées pour caractériser la performance d'un modèle : la discrimination, qui estime la capacité du score à séparer les individus malades et non-malades, et la calibration, qui estime la capacité du score à prédire un niveau de risque proche du niveau de risque réel. Ainsi, la figure 34 illustre la façon dont des individus sont classés selon que le score est discriminant (les profils sont nettement séparés sur l'échelle du risque prédit), mais mal calibré (le risque réel n'augmente pas en fonction de la hausse du risque prédit) (figure du haut), mal discriminant (il n'existe pas de nette séparation en fonction du risque prédit), mais bien calibré (l'augmentation du risque évalué correspond à une augmentation du risque réel) (figure du milieu), ou discriminant et bien calibré (figure du bas).

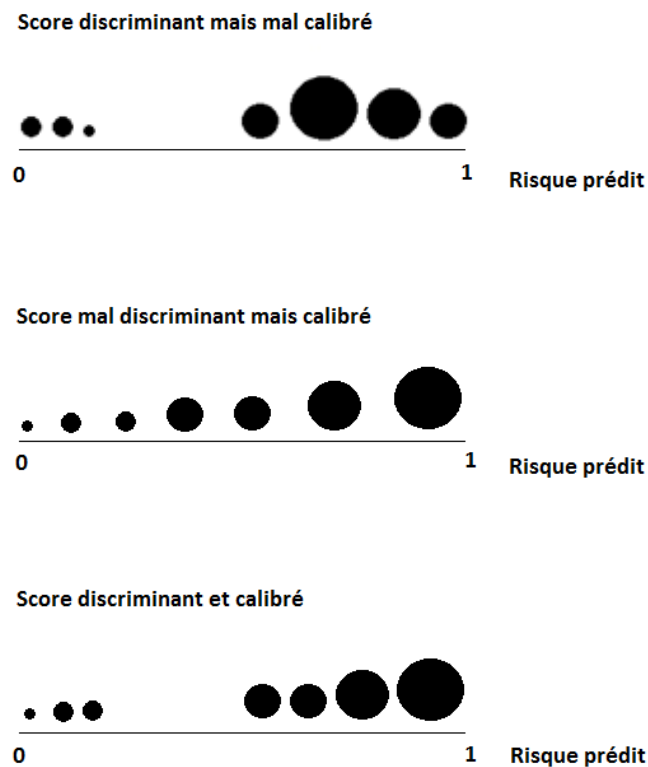


Figure 34 : Représentation de la discrimination et de la calibration d'un score prédictif.

Note : Le risque réel est représenté par la taille des cercles, plus le cercle est grand, plus le risque réel est élevé.

- **Définition des échantillons d'apprentissage et de test**

Le score se teste sur un échantillon différent de celui sur lequel il a été estimé. Pour évaluer la performance d'un score, il convient de le tester soit sur une population indépendante (on parle de validation externe), soit sur la même population en la subdivisant (on parle de validation interne). Dans le présent travail, il a été choisi de réaliser une validation interne du score prédictif à partir d'une double subdivision. La population d'analyse a été aléatoirement divisée en deux échantillons, un échantillon d'apprentissage et un échantillon de validation, avec le rapport 4:1 afin de pouvoir apprendre le score final sur l'ensemble de l'échantillon d'apprentissage et le valider sur l'échantillon de validation. Les variables à incorporer dans le score final sont sélectionnées à l'aide d'une méthode de validation croisée. Cette méthode consiste à scinder aléatoirement l'échantillon d'apprentissage en p partitions (ici $p = 5$) qui serviront successivement à tester le score par l'évaluation de ses performances. Chaque score est appris sur $\frac{p-1}{p}$ (ici $\frac{4}{5}$) de l'échantillon d'apprentissage et testé sur $\frac{1}{p}$ (ici $\frac{1}{5}$), et ce, cinq fois, de sorte que chacune des p partitions servira à tester le score une fois. On obtient alors p évaluations des performances du modèle que l'on moyennera pour obtenir la performance globale du score.

- **Estimation de la discrimination d'un modèle prédictif**

La discrimination d'un score prédictif estime sa capacité à séparer les individus malades et non-malades. Pour mesurer la discrimination d'un score prédictif, la mesure la plus couramment utilisée est l'aire sous la courbe de la fonction d'efficacité du récepteur (courbe ROC pour *receiver operating characteristic*), appelée AUC (*area under the ROC curve*).

Un score prédictif permet d'attribuer à chaque individu une classification prédite (malade / non-malade), qui sera alors comparée à la classification réelle (malade / non-malade) issue de l'observation, permettant de construire le tableau de contingence présenté dans le tableau 14.

Tableau 14 : Tableau de contingence entre la classe prédite et la classe réelle.

		Classe prédite	
		Malade	Non-malade
Classe réelle	Malade	Vrai positif	Faux Négatif
	Non-malade	Faux positif	Vrai Négatif

Ainsi, selon les valeurs de leurs classes prédite et réelle, chaque individu peut être classé comme malade, à tort (faux positif) ou à raison (vrai positif), ou comme non-malade, à tort (faux négatif) ou à raison (vrai négatif). Les effectifs de ces quatre classifications (tableau 14)

permettent ensuite de calculer la sensibilité (équation 22) et la spécificité (équation 23) associées au score prédictif.

Équation 22 : Définition de la sensibilité d'un score prédictif.

$$\begin{aligned} \text{Sensibilité} &= \frac{\text{Nombre de vrais positifs}}{\text{Nombre de positifs}} \\ &= \frac{\text{Nombre de vrais positifs}}{\text{Nombre de vrais positifs} + \text{Nombre de faux négatifs}} \end{aligned}$$

Équation 23 : Définition de la spécificité d'un score prédictif.

$$\begin{aligned} \text{Spécificité} &= \frac{\text{Nombre de vrais négatifs}}{\text{Nombre de négatifs}} \\ &= \frac{\text{Nombre de vrais négatifs}}{\text{Nombre de faux positifs} + \text{Nombre de vrais négatifs}} \end{aligned}$$

Une des difficultés d'un score prédictif réside dans la détermination d'une valeur seuil qui va décider si l'individu est classé malade ou non-malade. Si le seuil est trop bas, beaucoup de personnes seront classées comme malades, la sensibilité sera élevée, cependant un grand nombre le seront à tort (faux positifs), la spécificité sera alors très mauvaise. À l'inverse, un seuil trop élevé produira une bonne spécificité (peu d'individus non-malades seront diagnostiqués malades), mais une mauvaise sensibilité (le test laissera échapper un grand nombre de vrais malades). La meilleure valeur seuil est celle qui optimise la sensibilité et la spécificité du score prédictif pour trouver un compromis entre détecter un maximum de patients malades et écarter un maximum de patients non-malades.

La courbe ROC permet de définir la valeur de ce seuil en représentant la sensibilité en fonction de 1-spécificité pour l'ensemble des valeurs seuils possibles avec le jeu de données. L'AUC correspond au calcul de l'aire sous cette courbe ROC et estime la probabilité que le modèle prédictif attribue à un individu malade un score plus élevé qu'à un individu non-malade. Une performance parfaite est représentée par une AUC de 1, une performance équivalant à assigner la classe prédite au hasard par une AUC de 0,5.

Pour illustrer la construction d'une courbe ROC, le tableau 40 (en annexe page 219) présente le calcul de la sensibilité et de la spécificité pour un jeu de données simulées dans lequel une probabilité de survenue d'une maladie a été assignée à chacun des 32 individus malades ou non-malades. Les différentes statistiques sont calculées pour chaque valeur seuil possible, c'est-à-dire pour chaque valeur du score prédit observé dans le jeu de données. Le nombre de vrais positifs

est le nombre d'individus malades dont le score est supérieur (ou égal) au seuil, le nombre de faux positifs le nombre d'individus non-malades dont le score est supérieur (ou égal) au seuil, le nombre de faux négatifs le nombre d'individus malades dont le score prédit est inférieur (strictement) au seuil et le nombre de vrais négatifs le nombre d'individus non-malades dont le score prédit est inférieur (strictement) au seuil. La figure 35 présente la courbe ROC obtenue pour cet exemple.

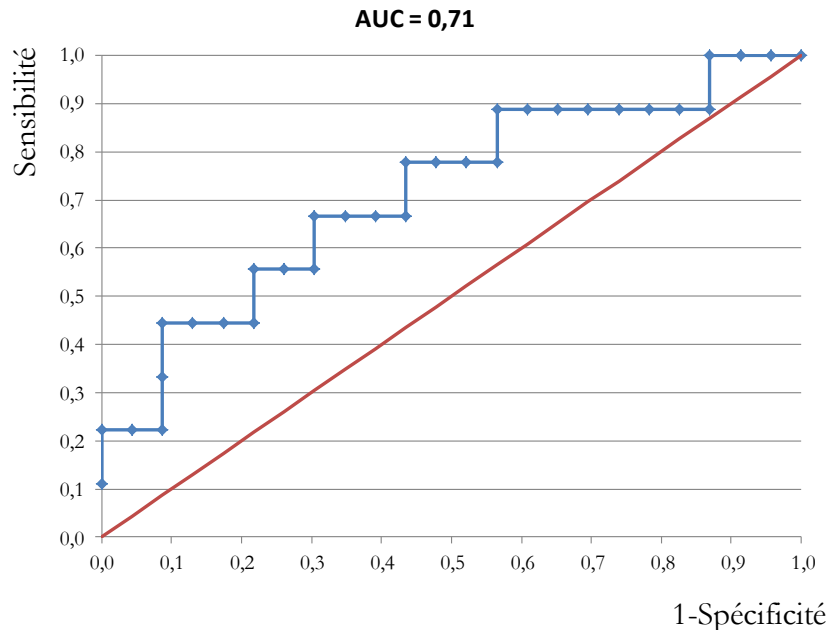


Figure 35 : Exemple de représentation d'une courbe ROC à partir d'un jeu de données simulées.

- **Estimation de la calibration d'un modèle prédictif**

La calibration d'un score prédictif est sa capacité à prédire un niveau de risque proche du niveau de risque réel. Pour mesurer la calibration d'un score prédictif, la mesure la plus couramment utilisée est le rapport du nombre estimé de cas de maladie sur le nombre observé de cas, mesure pouvant être, en complément, illustrée à l'aide de diagrammes de fiabilité.

Le nombre observé (O) est directement observable sur la population et représente le nombre d'individus qui ont été diagnostiqués avec la maladie dans la fenêtre de prédiction. Le nombre estimé (E) correspond à la somme des probabilités prédites de l'ensemble des individus de la population. Plus le rapport E/O est proche de 1, plus le nombre estimé de cas est proche du nombre observé de cas, et meilleure est la calibration. Pour avoir une vision plus précise de la calibration, le rapport E/O peut également être calculé par quantile de scores prédits afin d'estimer si la calibration est meilleure parmi les individus à faible risque ou parmi les individus à haut risque.

Une illustration du rapport E/O peut être obtenue par les diagrammes de calibration (également appelés diagrammes de fiabilité) qui représentent le niveau moyen des scores prédits en fonction de la proportion de cas dans plusieurs sous-groupes de la population, découpés selon les percentiles du score prédit. Plus les points s'éloignent de la diagonale, plus la différence entre le nombre de cas observés et le score prédit est importante.

Pour illustrer l'estimation du rapport E/O et la représentation des diagrammes de calibration, reprenons l'exemple présenté dans le tableau 40 (en annexe page 219). Le rapport E/O calculé pour cet exemple de score est de 0,86. Le tableau 15, présentant le calcul des rapports E/O par quintiles de score, permet de se rendre compte que le score a tendance à sous-estimer le risque réel de pratiquement toutes les sous-populations.

Tableau 15 : Exemple de calcul des rapports E/O par intervalles de quartiles du score prédit pour un jeu de données simulées.

Intervalles de score	Rapport E/O
[0,00 ; 0,12[0,590
[0,12 ; 0,23[0,641
[0,23 ; 0,36[1,223
[0,36 ; 1,00]	0,853

La figure 36 présente le diagramme de calibration pour le jeu de données du tableau 40 (en annexe page 219). Ce diagramme permet de se rendre compte que le risque prédit est proche du risque réel.

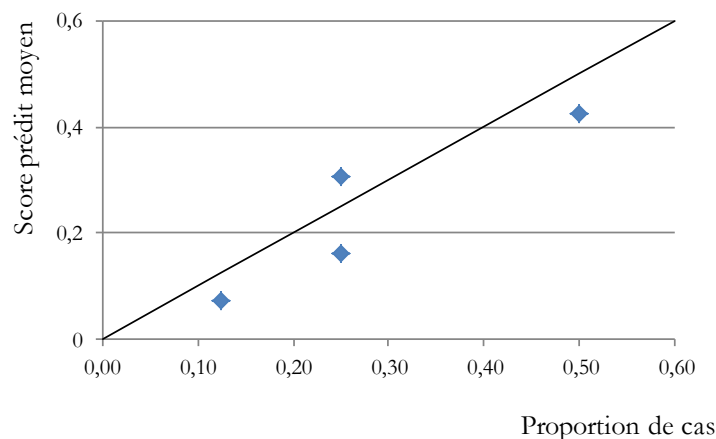


Figure 36 : Exemple d'un diagramme de calibration pour un jeu de données simulées.

Dans le cas du modèle de régression de Cox, la discrimination peut également être estimée en validant dans l'échantillon test, l'estimation effectuée dans l'échantillon d'apprentissage, de deux manières différentes. La première consiste à régresser le pronostic index ($\sum \beta'X$), dont les

coefficients de régression sont issus de l'estimation dans l'échantillon d'apprentissage, dans l'échantillon test ; plus le coefficient de régression associé au pronostic index est proche de 1, meilleure est la discrimination. La seconde méthode consiste à régresser les variables incluses dans le modèle final et le pronostic index pour qui le coefficient de régression a été contraint à 1 ; plus les coefficients de régression associés à chaque variable sont proches de zéro, meilleure est la discrimination.

Dans le cas particulier d'un score construit à l'aide d'un modèle de régression de Cox, comme la fonction de survie de base est modélisée il convient de valider la modélisation, ce qui peut être réalisé de deux manières différentes. La première consiste à évaluer, dans l'échantillon d'apprentissage, si la modélisation est proche des vraies valeurs. Ainsi, les valeurs estimées, ainsi que leur intervalle de confiance à 95 % obtenu par bootstrap (voir le paragraphe suivant), sont comparées aux valeurs observées dans l'échantillon d'apprentissage. La seconde méthode consiste à évaluer, dans l'échantillon de validation, si les vraies valeurs sont proches de la modélisation. Ainsi, les valeurs observées dans l'échantillon de validation, ainsi que leur intervalle de confiance à 95 % obtenu par bootstrap, sont comparés aux valeurs estimées obtenues en réaffectant la modélisation obtenue dans l'échantillon d'apprentissage au sein de l'échantillon de validation.

La méthode par bootstrap permet d'obtenir un intervalle de confiance pour la statistique d'intérêt lorsqu'aucune formulation analytique pour cet intervalle n'est disponible, ou n'est pas facilement estimable. Cette méthode consiste à ré-échantillonner l'échantillon d'origine, en sélectionnant aléatoirement, avec remise, un nouvel échantillon de même taille que l'échantillon d'origine. En général, l'échantillonnage est réalisé 100, 1 000 ou 10 000 fois. Les statistiques d'intérêt sont estimées dans chaque échantillon, ce qui permet d'en déduire leur distribution. L'intervalle de confiance à 95 % des statistiques d'intérêt correspond alors à l'intervalle défini par les percentiles à 2,5 % et 97,5 % de la distribution obtenue par l'ensemble des échantillons.

2.11.c. Méthodologie employée pour l'obtention du modèle de prédiction final

Pour prédire le risque de survenue à cinq ans de l'événement d'intérêt, il convient, dans un premier temps, de choisir les variables à inclure dans le modèle de prédiction. La méthodologie choisie dans le présent travail consiste à partir d'un ensemble de variables potentiellement prédictives (établi à partir de la littérature) et à tester plusieurs modèles à l'aide des valeurs de la discrimination (AUC) et de la calibration (rapport E/O) obtenues dans ces modèles. Les termes d'interactions entre chaque paire de variables sont également testés.

Une méthode de sélection ascendante a été mise en place en partant du modèle vide auquel ont été successivement ajoutées les variables permettant d'optimiser la discrimination et la calibration du modèle. Les 15 premières variables sélectionnées de cette manière sont considérées dans une seconde étape de sélection afin de s'assurer que l'ordre dans lequel les variables ont été incorporées dans le modèle n'influe pas. Les combinaisons de 1 à 15 variables des 15 variables sélectionnées dans la première étape sont alors testées afin de sélectionner le modèle optimisant la discrimination et la calibration. Lors de la sélection des variables, la fonction de survie de base est modélisée à l'aide de spline cubique à partir de l'estimateur de Breslow de la fonction de risque cumulée de base issue d'un modèle vide (*i.e.* sans variable explicative) et en prenant comme nœuds les quintiles de distribution de la fonction de risque cumulée de base.

Lorsque les variables à inclure dans le modèle final sont sélectionnées, la validation complète du modèle final peut être effectuée (section Méthodes statistiques, partie 2.11.b page 120). La modélisation de la fonction de survie de base est alors adaptée au modèle multivarié en testant plusieurs choix de nœuds pour la modélisation par splines cubiques ou par splines cubiques restreints. La modélisation qui est retenue est celle qui minimise l'AIC.

2.11.d. Comparaison de deux scores prédictifs

La comparaison des risques de survenue de l'événement d'intérêt fournis par des modèles de prédiction différents permet de mettre en relation les performances des modèles. Cette comparaison est principalement établie à partir de l'estimation de deux statistiques différentes, que sont le coefficient de corrélation de Pearson et la mesure de concordance.

- **Le coefficient de corrélation de Pearson**

Le coefficient de corrélation de Pearson est défini par la formule présentée dans l'équation 24. Il est fonction de la covariance entre deux statistiques X et Y , $cov(X, Y)$ et de leurs écarts-types respectifs σ_X et σ_Y .

Équation 24 : Définition du coefficient de corrélation de Pearson entre deux variables quantitatives.

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

La valeur du coefficient de corrélation est comprise entre -1 et 1. Si $\rho_{X,Y} = 0$, alors il n'existe aucune corrélation entre les valeurs des deux statistiques, les scores sont dissemblables. Si $\rho_{X,Y} > 0$, les valeurs prédites X élevées sont associées à des valeurs prédites Y élevées. Si $\rho_{X,Y} < 0$, les valeurs prédites X élevées sont associées à des valeurs prédites Y faibles.

• *La mesure de concordance*

La mesure de concordance permet de quantifier la comparaison entre les rangs assignés aux individus par deux modèles prédictifs en estimant le coefficient de concordance κ (kappa) défini par la formule présentée dans l'équation 25, établie à partir des effectifs du tableau de concordance de la distribution des prédictions obtenues par les deux modèles, présentés dans le tableau 16.

Tableau 16 : Tableau de concordance pour n individus classés par tertiles de valeurs de deux modèles prédictifs.

		Modèle 2			Total
		Tertile 1	Tertile 2	Tertile 3	
Modèle 1	Tertile 1	n_{11}	n_{12}	n_{13}	l_1
	Tertile 2	n_{21}	n_{22}	n_{23}	l_2
	Tertile 3	n_{31}	n_{32}	n_{33}	l_3
Total		c_1	c_2	c_3	n

Équation 25 : Définition du coefficient de concordance entre deux variables quantitatives.

$$\kappa_{X,Y} = \frac{C_O - C_A}{1 - C_A}$$

Avec $C_O = \frac{1}{n}(n_{11} + n_{22} + n_{33})$, la concordance observée

et $C_A = \frac{1}{n}\left(\frac{c_1 \cdot l_1}{n} + \frac{c_2 \cdot l_2}{n} + \frac{c_3 \cdot l_3}{n}\right)$, la concordance attendue

Les valeurs du coefficient de concordance sont comprises entre -1 et 1. Il représente le degré d'accord dans les classements obtenus par l'attribution des valeurs par les deux modèles prédictifs. Si $\kappa_{X,Y} = 0$, alors l'accord est dû au hasard. Si $\kappa_{X,Y} = 1$, l'accord est parfait entre les modèles prédictifs. Si $\kappa_{X,Y} = -1$, le désaccord est total entre les modèles prédictifs. En général, on considère que la concordance est très bonne pour une valeur du coefficient supérieure à 0,8, bonne pour une valeur comprise entre 0,6 et 0,8, moyenne pour une valeur entre 0,4 et 0,6, mauvaise en dessous de 0,4.

La méthode de calcul du coefficient de concordance présentée s'estime à partir du découpage des risques de survenue de l'événement d'intérêt selon les tertiles de distribution. La méthode est identique lorsque le découpage est effectué selon les quartiles ou les quintiles de distribution des risques.

2.11.e. Deux autres méthodes pour la construction d'un score prédictif : le modèle des plus proches voisins et le modèle de Gail.

- *L'algorithme des plus proches voisins*

L'algorithme des plus proches voisins repose sur une méthode exploratoire (230-232). Le principe de cet algorithme consiste à déterminer le risque associé à une maladie (ici le cancer du sein invasif) en fonction de la proportion d'individus similaires atteints par cette maladie dans la population d'étude. Les individus similaires à un individu constituent le voisinage de cet individu, et sont déterminés en estimant la distance séparant les individus deux à deux. Chaque individu est défini par les valeurs qu'ils présentent pour l'ensemble des facteurs constituant le score de prédiction. À chaque valeur possible d'une variable a été associée la proportion de cancers du sein diagnostiqués parmi l'ensemble des femmes présentant la même valeur pour cette variable, afin de pouvoir considérer simultanément des variables quantitatives et qualitatives. La distance entre deux individus est calculée à l'aide d'une mesure de distance Euclidienne dont la formule est présentée dans l'équation 26.

Équation 26 : Formule de la distance Euclidienne utilisée dans l'algorithme des plus proches voisins.

$$d(X, Y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

Avec $X = (x_1, x_2, \dots, x_n)'$ et $Y = (y_1, y_2, \dots, y_n)'$ les profils des deux individus pour lesquels la distance $d(X, Y)$ les séparant est estimée.

Plusieurs tailles de voisinage sont testées afin de déterminer le nombre d'individus similaires à sélectionner pour optimiser les performances du score prédictif. Il est à noter que la taille de voisinage choisie est la taille minimum du voisinage. En effet, étant donné que plusieurs individus peuvent être à même distance de l'individu pour lequel on cherche à construire le voisinage, la distance maximale à prendre en considération peut induire la sélection d'un nombre d'individus similaires supérieur à la taille fixée pour le voisinage (ici fixée à 20). Une fois le voisinage d'un individu établi, son niveau de risque de cancer du sein est estimé comme la proportion d'individus dans le voisinage qui ont été diagnostiqués avec un cancer du sein dans la fenêtre de prédiction (ici fixée à cinq ans).

Cette méthode a l'avantage d'être intuitive et plus compréhensible pour le grand public qu'un score établi à partir d'un modèle statistique.

- **Le modèle de Gail**

Le modèle de Gail repose sur une modélisation statistique, le modèle de régression logistique. Il a été construit à partir des données d'une étude cas-témoin nichée dans la cohorte BCDDP (*Breast cancer detection and demonstration project*) qui regroupe 2 852 cas de cancer du sein et 3 146 témoins, âgés de 35 à 79 ans. Les facteurs prédictifs pris en considération dans la construction du score sont l'âge (< 50 ans, ≥ 50 ans), l'âge à la puberté (< 12 , entre 12 et 13, ≥ 14), le nombre de biopsies pratiquées (0, 1, ≥ 2), l'âge à la première grossesse à terme (< 20 , entre 20 et 24, entre 25 et 29 ou nullipare, ≥ 30) et le nombre de parents au premier degré chez qui un cancer du sein a été diagnostiqué (0, 1, ≥ 2). Le modèle incorpore l'ensemble de ces variables individuellement, ainsi que les termes d'interaction entre le nombre de biopsies et l'âge et entre l'âge à la première grossesse à terme et le nombre d'antécédents familiaux de cancer du sein.

Le modèle de régression logistique repose sur la modélisation d'une variable binaire, définie comme la probabilité de l'occurrence d'une maladie (équation 27), en fonction de covariables X .

Équation 27 : Modélisation du modèle de régression logistique.

$$\text{logit}(P(\text{Malade}|X)) = \beta_0 + \beta'X$$

Avec $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$

La probabilité individuelle de développer un cancer du sein a été calculée à partir des risques relatifs, estimés à l'aide de la régression logistique (équation 28), associés à chaque modalité de chaque variable incorporée dans le score et de la probabilité de décès par une autre cause que le cancer du sein. L'âge étant un facteur dépendant du temps, sa variation dans le temps a été prise en compte dans le calcul du risque prédit de cancer du sein.

Équation 28 : Estimation du risque relatif entre deux profils à partir d'une modélisation par modèle de régression logistique.

$$RR(X_1 \text{ vs. } X_2) = \frac{\frac{P(\text{Malade}|X_1)}{1-P(\text{Malade}|X_1)}}{\frac{P(\text{Malade}|X_2)}{1-P(\text{Malade}|X_2)}} = \frac{\exp(\beta_0 + \beta'X_1)}{\exp(\beta_0 + \beta'X_2)} = e^{\beta'(X_1 - X_2)}$$

Partie 3 Estimation des fractions attribuables

3.1. Définition

La fraction attribuable (ou PAF pour *population attributable fraction*) est un concept statistique permettant de quantifier l'impact d'une exposition sur l'incidence d'une pathologie (ou sur la mortalité). La fraction attribuable estime le nombre (ou le pourcentage) de cas de maladie (ou de décès) attribuables à l'exposition au facteur de risque d'intérêt, ce qui correspond au nombre (ou pourcentage) de cas qui auraient pu théoriquement être évités si l'exposition avait pu être supprimée dans la population d'analyse. Cette statistique dépend à la fois de la force de l'association entre l'exposition et la pathologie étudiée (estimée à partir du risque relatif) et de la prévalence de l'exposition au sein de la population. Ainsi, la fraction attribuable permet de quantifier le poids, au sein d'une population, d'un facteur de risque sur la survenue d'une pathologie. Toutefois, il est à noter que l'estimation de la proportion de cas attribuables à une exposition repose sur une hypothèse forte de causalité entre le facteur d'exposition et la pathologie étudiée.

La définition numérique de la fraction attribuable est présentée dans l'équation 29. La fraction attribuable prend ses valeurs dans $]-\infty ; 1]$ et est généralement exprimée en pourcentage. Lorsque les valeurs sont négatives, la probabilité d'être malade est supérieure chez les non exposés que dans la population générale et la fraction attribuable exprime un pourcentage de cas supplémentaires qui seraient théoriquement diagnostiqués avec la suppression de l'exposition. Lorsque les valeurs sont positives, la probabilité d'être malade est supérieure chez les exposés que dans la population générale et la fraction attribuable exprime un pourcentage de cas qui seraient théoriquement évités avec la suppression de l'exposition.

Équation 29 : Définition de la fraction attribuable.

$$PAF = \frac{P(\text{Malade}) - P(\text{Malade} | \text{Non exposé})}{P(\text{Malade})}$$

3.2. Formalisation mathématique

La première formulation mathématique a été introduite en 1953 (équation 30) dans une analyse étudiant l'impact du tabagisme actif sur l'incidence de cancer du poumon chez les hommes (233). Cette formulation s'obtient directement de l'équation 29 en posant que le risque

$$\text{relatif}^* RR = \frac{P(\text{Malade} | \text{Exposé})}{P(\text{Malade} | \text{Non exposé})}$$

Équation 30 : Formulation mathématique de la fraction attribuable proposée par Levin, 1953.

$$PAF = \frac{P(\text{Exposé}) \cdot (RR - 1)}{1 + P(\text{Exposé}) \cdot (RR - 1)}$$

Néanmoins, cette formulation est valable pour une exposition dichotomique et produit des résultats biaisés lorsque le RR employé est un RR ajusté, issu de modèles multivariés (234). Pour contourner ces deux problèmes, deux autres définitions mathématiques ont été fournies en 1974 (équation 31) pour une exposition dichotomique (235), puis en 1985 (équation 32) pour une exposition polytomique (130;236) donnant des résultats non biaisés avec l'utilisation de RR ajustés (234).

Équation 31 : Formulation mathématique de la fraction attribuable proposée par Miettinen, 1974.

$$PAF = P(\text{Exposé} | \text{Malade}) \cdot \frac{(RR - 1)}{RR}$$

Équation 32 : Formulation mathématique de la fraction attribuable proposée par Bruzzi, 1985.

$$PAF = 1 - \frac{\sum_{a=1}^A p_{.a} \cdot RR_{2a}}{\sum_{i=1}^I \sum_{a=1}^A p_{ia} \cdot RR_{1i} \cdot RR_{2a}}$$

Avec : p_{ia} la prévalence de la modalité d'exposition i pour le niveau d'ajustement a

$$p_{.a} = \sum_{i=1}^I p_{ia}$$

RR_{1i} le risque relatif de la modalité i du facteur d'exposition d'intérêt par rapport à la référence ($RR_{11} = 1$)

RR_{2a} le risque relatif de la modalité a du facteur d'ajustement par rapport à la référence ($RR_{21} = 1$)

Les intervalles de confiance pour les fractions attribuables sont très peu estimés dans la littérature. À ce jour, les méthodes les plus utilisées pour leur estimation sont la méthode par bootstrap (237) et la delta-méthode (236). Les intervalles de confiance pour les fractions attribuables présentées dans ce travail sont issus de la delta-méthode (236).

3.3. Erreurs à éviter

L'erreur la plus répandue dans la littérature est l'utilisation de la formule de Levin (équation 30) pour le calcul de fractions attribuables multi-ajustés (234). Une autre erreur à éviter est d'oublier que les fractions attribuables ne sont pas sommables : la fraction attribuable associée à plusieurs expositions n'est pas égale à la somme des fractions attribuables associées à chaque exposition individuelle.

Enfin, étant donné que la formule de Miettinen (235;236) n'accepte pas de variables d'intérêt et d'ajustement continues, il est important que la catégorisation pour chaque variable incluse dans le modèle soit choisie *a priori* et établie à partir de la littérature.

3.4. Interprétation

La fraction attribuable prend en compte à la fois la force de l'association entre l'événement d'intérêt et l'exposition analysée et la prévalence de l'exposition au sein de la population d'analyse. Cette statistique permet d'estimer l'impact de l'exposition au sein d'une population et d'aider à cibler des stratégies de prévention (234). Les fractions attribuables estiment la proportion de cas attribuables à une ou plusieurs expositions, dont la suppression pourrait potentiellement prévenir ces cas.

Enfin, il ne faut pas oublier que l'estimation des fractions attribuables repose sur une hypothèse forte de causalité entre l'exposition et l'événement d'intérêt. Les résultats doivent donc être interprétés avec prudence.

Partie 4 Gestion des valeurs manquantes

4.1. Gestion des valeurs manquantes de l'exposition principale

Lorsqu'une exposition principale est analysée, les individus de la population d'analyse présentant une valeur manquante pour cette variable d'exposition ont été exclus des analyses.

4.2. Gestion des valeurs manquantes des variables d'ajustement

Lorsqu'une variable d'ajustement présentait moins de 5 % de valeurs manquantes dans la population d'analyse, celles-ci ont été imputées par la médiane si la variable d'ajustement était quantitative ou par la classe modale si la variable d'ajustement était qualitative. Si, au contraire, les valeurs manquantes représentaient plus de 5 % des observations, une catégorie « Manquant » a été créée et intégrée dans la modélisation.

RÉSULTATS

Partie 1 Influence du mode de vie sur le risque de cancer

1.1. Introduction

Comme précisé dans l'introduction (partie 1.2 page 33), le cancer est la deuxième cause de décès chez les femmes en France, depuis plusieurs décennies (93). Des recommandations nationales de santé publique ont été établies pour alerter la population sur l'amélioration de l'espérance de vie et la diminution du risque de cancer (174;238;239) qu'apporte l'adoption d'un mode de vie sain combinant des comportements de santé tels que le fait de ne pas fumer, le maintien d'un poids normal (IMC entre 18,5 et 25), la limitation de la consommation d'alcool, l'adoption d'habitudes alimentaires spécifiques (comme une consommation élevée de fruits et légumes) et la pratique régulière d'une activité physique (section Introduction, partie 3.1 page 57).

Le but de cette partie est d'estimer la proportion de cancers, toutes localisations confondues et par localisations spécifiques, attribuables à un mode de vie malsain, établi à partir de cinq caractéristiques comportementales.

1.2. Matériel et méthodes

1.2.a. Population d'étude

Les femmes sont suivies de la date de réponse au questionnaire alimentaire Q3 (1993) jusqu'au premier événement observé entre la date de diagnostic de n'importe quel cancer, autre que le cancer cutané basal, la date de réponse au dernier questionnaire rempli et la date d'envoi du dernier questionnaire disponible (Q9, 06/2008). Les femmes non-répondantes au questionnaire alimentaire (N = 24 473), celles dont un cancer, autre qu'un cancer cutané basal, a été diagnostiqué avant le début de suivi (N = 4 705), celles sans suivi après Q3 (N = 769), les cas de cancer sans date de diagnostic (N = 54) et les femmes dont l'information sur au moins un des cinq comportements de mode de vie d'intérêt est manquante (N = 4 262) ont été exclues des analyses. La population d'analyse comprend ainsi 64 732 femmes, âgées de 43 à 68 ans au début du suivi.

1.2.b. Collecte des données et construction de l'indice de santé

La collecte des données sur les cinq comportements de mode de vie analysés a été décrite dans la section Matériel, partie 4 page 94.

Les cinq comportements de mode de vie analysés, à savoir le tabagisme actif, l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique ont été combinés pour créer un indice de santé, construit *a priori*. Cet indice de santé a été établi à partir des recommandations de santé publique émises par le PNNS (174), l'ANSES (240) et l'OMS (239). Chaque score individuel associé à chaque recommandation prend comme valeurs 0 (non adhésion à la recommandation), 0,5 (adhésion partielle à la recommandation) ou 1 (adhésion complète à la recommandation). Les seuils choisis pour chaque score individuel sont présentés dans le tableau 17. Par exemple, les femmes qui adhèrent aux cinq recommandations de santé publique n'ont jamais fumé, ont un IMC entre 18,5 et 25, consomment moins d'un verre d'alcool par jour, consomment au moins cinq portions de fruits et légumes par jour et ont un niveau d'activité physique récréationnelle atteignant 20 METs-h/semaine (*i.e.* 300 minutes d'activité modérée ou 150 minutes d'activité intense par semaine) (239).

Tableau 17 : Construction de l'indice de santé à partir de cinq facteurs de mode de vie.

	Adhésion aux recommandations de santé publique		
	Complète (1 point)	Partielle (0,5 point)	Limitée (0 point)
Recommandations de santé			
Tabagisme actif	N'a jamais fumé	Ex-fumeuse	Fumeuse
IMC (kg.m ⁻²)	[18,5 ; 25[(normal)	[25 ; 30[(surpoids) [16 ; 18,5[(maigreur)	≥ 30 (obésité) < 16 (dénutrition)
Consommation d'alcool (verres ^a /jour)	< 1	[1 ; 2[≥ 2
Consommation de fruits et légumes (portions ^b /jour)	≥ 5	[3,5 ; 5[< 3,5
Niveau d'activité physique récréationnelle (METs-h/semaine)	≥ 20	[10 ; 20[< 10

a : Un verre d'alcool correspond à 10 grammes d'éthanol.

b : Une portion de fruits et légumes correspond à 80 grammes de fruits et légumes, qui incluent les légumes crus et cuits, la salade et les fruits frais.

Un indice de santé est alors assigné à chaque femme de la population d'analyse, défini comme la somme des cinq scores individuels. Par construction, l'indice de santé est compris entre 0 (mode de vie le moins sain) et 5 (mode de vie le plus sain), par intervalles de 0,5 point.

1.2.c. Définition des cas de cancer

L'analyse s'intéresse à l'association entre l'indice de santé et le risque d'occurrence d'un premier cancer invasif, quelle que soit sa localisation. La collecte et la validation de l'information concernant les cas de cancer ont été décrits dans la section Matériel, partie 2.1 page 79. Les différentes localisations considérées sont les cancers du sein, de la peau (à l'exclusion du cancer cutané basal), du côlon-rectum, digestifs (autre que côlon-rectum), de l'endomètre, de la thyroïde, des ovaires et du poumon, ainsi que les hémopathies malignes, dont la classification est présentée dans le tableau 9 page 83. Les cancers de la vessie, du rein, ORL, du cerveau, du col de l'utérus, des autres localisations ou de site inconnu n'ont pas été analysés à cause d'un trop faible effectif. Le nombre de cas de cancer dans la population d'analyse est présenté dans le tableau 18 page 136.

1.2.d. Méthodologie statistique

L'analyse de l'association entre chaque comportement de mode de vie ou l'indice de santé et le risque de cancer a été réalisée à l'aide de modèles de régression de Cox, standards et à risque compétitif, en utilisant l'âge comme échelle de temps (section Méthodes statistiques, partie 2.7 page 116). Les HR et les intervalles de confiance à 95 % pour le risque de cancer, toutes localisations confondues et par site spécifique, comparant les femmes qui adhèrent le mieux aux femmes qui adhèrent le moins bien aux recommandations de santé publique, ont été estimés.

Les analyses ont été ajustées sur le niveau d'éducation (inférieur à Bac, Bac-Bac+2, supérieur à Bac+2), la zone de résidence (sud, centre, nord), les antécédents familiaux de cancer chez les parents au premier degré (au moins un, aucun), l'activité professionnelle (active, retraitée ou n'a jamais travaillé), l'utilisation de contraceptifs oraux (déjà, jamais), le statut ménopausique et l'utilisation de THM (préménopause, postménopause avec THM, postménopause sans THM), l'âge à la puberté (en quartiles), le nombre d'enfants et l'âge à la première grossesse à terme (nullipare, un enfant avant 30 ans, un enfant après 30 ans, plus d'un enfant dont le premier avant 30 ans, plus d'un enfant dont le premier après 30 ans). L'information sur le statut ménopausique et l'utilisation de THM a été mise à jour à tous les questionnaires ; toutes les autres variables d'ajustement ont été évaluées au début du suivi (section Matériel, partie 4 page 94). Les analyses ont également été stratifiées sur l'année de naissance ([1925-1930], [1931-1935], [1936-1945], [1946-1950]) pour prendre en compte l'effet cohorte (section Méthodes statistiques, partie 2.9 page 116).

Les proportions attribuables aux cinq comportements individuels et à leur combinaison ont été estimées d'après la méthodologie présentée dans la section Méthodes statistiques, partie 3 page 130.

1.3. Résultats

1.3.a. Caractéristiques de la population d'analyse

Les 64 732 femmes de la population d'analyse ont été suivies pendant un total de 840 097 femmes-années (avec un suivi médian de 8 ans pour les cas de cancer invasif et de 15 pour les non-cas). Parmi ces femmes, 6 938 ont été diagnostiquées avec un cancer, dont la distribution des localisations est présentée dans le tableau 18.

Tableau 18 : Localisations des cancers diagnostiqués dans la population d'analyse, cohorte E3N (N = 64 732).

Localisation	Effectif	Pourcentage
Sein en préménopause	609	8,78
Sein en postménopause	2 874	41,42
Cutané non basocellulaire	686	9,89
Côlon-rectum	481	6,93
Hémopathies malignes ^a	411	5,92
Digestifs, autres que côlon-rectum ^b	274	3,95
Endomètre	270	3,89
Thyroïde	267	3,85
Ovaires	248	3,57
Poumon	213	3,07

a : Les hémopathies malignes incluent les tumeurs malignes de la rate et des tissus lymphoïde, hématopoïétique et apparentés, la maladie de Hodgkin, les lymphomes non hodgkiniens, les maladies immunoprolifératives malignes, les myélomes multiples et les leucémies.

b : Les cancers digestifs incluent les tumeurs malignes de l'œsophage, de l'estomac, de l'intestin grêle, de l'appendice, du côlon sans précision, du foie, des voies biliaires, de la vésicule biliaire et du tractus intestinal et les lésions à localisations contiguës de l'appareil digestif.

Les cancers les plus fréquemment diagnostiqués dans la population d'analyse au cours du suivi sont les cancers du sein (en préménopause ou en postménopause pour un total de 50,2 % de cas de cancer du sein), suivi du cancer cutané non basocellulaire (9,9 %), du cancer du côlon-rectum (6,9 %) et des hémopathies malignes (5,9 %).

La distribution des cinq scores individuels dans la population d'analyse est illustrée dans la figure 37.

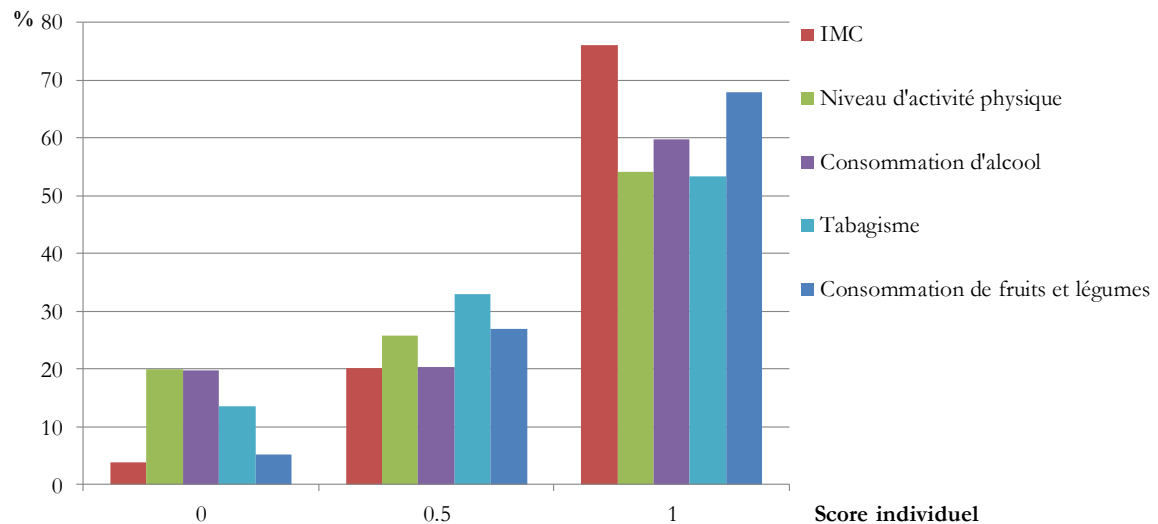


Figure 37 : Distribution des cinq scores individuels dans la population d'analyse, cohorte E3N (N = 64 732).

La majorité des femmes de la population d'analyse adhère à chacune des cinq recommandations analysées, dont les plus respectées sont : avoir un IMC dans les normes de recommandation de l'OMS (*i.e.* entre 18,5 et 25) et consommer la quantité recommandée de fruits et légumes (*i.e.* au moins cinq portions par jour).

Les caractéristiques des femmes de la population d'analyse selon la catégorie de l'indice de santé sont présentées dans le tableau 41 (en annexe page 220).

Les femmes de la population d'analyse adhèrent bien aux recommandations de santé publique : 26,2 % d'entre elles ont un indice de santé compris entre 4,5 et 5 (*i.e.* adhèrent totalement à quatre des recommandations et totalement ou partiellement à la cinquième) et seulement 6,0 % de la population d'analyse adhèrent à 2 recommandations ou moins. Les populations composées des femmes ayant un indice compris entre 0 et 2 et des femmes ayant un indice compris entre 4,5 et 5 présentent des différences statistiquement significatives (tableau 41 en annexe page 220). En particulier, on observe qu'une proportion plus élevée de femmes qui adhèrent le mieux aux recommandations de santé publique résident dans le sud de la France (31,8 %), comparées aux femmes qui y adhèrent le moins bien (24,4 %), de même que pour la proportion de femmes retraitées ou n'ayant jamais travaillé (35,2 % *vs.* 20,9 %), de femmes ménopausées (58,4 % *vs.* 49,6 %), de femmes ménopausées ayant récemment utilisé un THM (38,1 % *vs.* 34,7 %) et de femmes ayant eu au moins un enfant (89,5 % *vs.* 86,1 %). Ces différences peuvent être conséquentes d'un effet âge : les femmes qui ont un indice de santé compris entre 4,5 et 5 sont en moyenne plus âgées que les femmes qui ont un indice compris entre 0 et 2 (53,3 *vs.* 51,4 ans).

RÉSULTATS - PARTIE 1

Pour évaluer la corrélation entre les cinq recommandations de santé publique analysées dans cette partie, les coefficients de corrélations de Pearson estimés pour chaque paire de scores individuels (*i.e.* les variables renseignant pour chaque recommandation de santé si la femme y adhère (totalement ou partiellement) ou pas, prenant les valeurs 1, 0,5 ou 0) sont présentés dans le tableau 19.

Tableau 19 : Coefficients de corrélation de Pearson entre les scores individuels, deux à deux, cohorte E3N (N = 64 732).

	Tabagisme actif	IMC	Consommation d'alcool	Consommation de fruits et légumes	Niveau d'activité physique
Tabagisme actif	1	-0,00557	0,17957 ^a	0,05515 ^a	-0,00870 ^b
IMC		1	-0,01827 ^a	-0,01829 ^a	0,08793 ^a
Consommation d'alcool			1	0,05885 ^a	-0,04013 ^a
Consommation de fruits et légumes				1	0,08809
Niveau d'activité physique					1

a : p-valeur < 0,001.

b : p-valeur < 0,01.

Les estimations suggèrent une corrélation statistiquement significative entre chaque paire de comportements (p-valeur < 0,01), sauf entre les scores du tabagisme actif et de l'IMC (p-valeur = 0,16). Du tableau, on déduit que le tabagisme actif est associé à un poids sain ($c = -0,00557$), à une consommation élevée d'alcool ($c = 0,17957$), à une faible consommation de fruits et légumes ($c = 0,05515$) et à un faible niveau d'activité physique ($c = -0,0087$). De plus, un niveau élevé d'activité physique est associé à un poids sain ($c = 0,08793$) et à une consommation élevée d'alcool ($c = -0,04013$). Ces corrélations sont cohérentes avec les résultats de la littérature exposés dans la section Introduction, partie 3.1.f page 60.

1.3.b. Association entre les comportements individuels, leur combinaison et le risque de cancer, toutes localisations confondues et par site spécifique

Les résultats concernant l'estimation du risque de cancer associé à chaque score individuel sont présentés dans le tableau 20.

Tableau 20 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de cancer, toutes localisations confondues et par site spécifique, cohorte E3N (N = 64 732).

Localisation ^a Score	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Cancer, toutes localisations confondues (N = 6 938)										
0	966 (13,92)	1,00	302 (4,36)	1,00	1 492 (21,50)	1,00	924 (13,32)	1,00	1 367 (19,70)	1,00
0,5	2 264 (32,63)	0,93 [0,86 ; 1,00]	1 450 (20,90)	0,88 [0,78 ; 1,00]	1 455 (20,97)	0,95 [0,88 ; 1,02]	1 301 (18,75)	0,96 [0,88 ; 1,04]	1 781 (25,67)	1,00 [0,93 ; 1,07]
1	3 708 (53,44)	0,92 [0,86 ; 0,99]	5 186 (74,75)	0,84 [0,75 ; 0,95]	3 991 (57,52)	0,90 [0,85 ; 0,95]	4 713 (67,93)	0,93 [0,87 ; 1,00]	3 790 (54,63)	1,00 [0,94 ; 1,06]
P-valeur de tendance		0,0771		<0,0001		0,0001		0,1037		0,4925
Cancer du sein en préménopause (N = 609)										
0	92 (15,11)	1,00	17 (2,79)	1,00	117 (19,21)	1,00	105 (17,24)	1,00	135 (22,17)	1,00
0,5	240 (39,41)	1,10 [0,86 ; 1,40]	94 (15,44)	1,01 [0,60 ; 1,70]	129 (21,18)	1,04 [0,81 ; 1,33]	136 (22,33)	0,95 [0,74 ; 1,23]	165 (27,09)	0,95 [0,76 ; 1,20]
1	277 (45,48)	0,96 [0,75 ; 1,22]	498 (81,77)	1,17 [0,72 ; 1,90]	363 (59,61)	1,01 [0,81 ; 1,24]	368 (60,43)	0,86 [0,69 ; 1,08]	309 (50,74)	0,94 [0,77 ; 1,15]
P-valeur de tendance		0,3990		0,6410		0,8761		0,0310		0,1878
Cancer du sein en postménopause (N = 2 874)										
0	358 (12,46)	1,00	112 (3,89)	1,00	661 (23,00)	1,00	376 (13,08)	1,00	571 (19,87)	1,00
0,5	932 (32,43)	1,05 [0,93 ; 1,19]	568 (19,76)	0,93 [0,76 ; 1,14]	603 (20,98)	0,88 [0,79 ; 0,99]	544 (18,93)	0,97 [0,85 ; 1,11]	728 (25,33)	0,97 [0,87 ; 1,08]
1	1 584 (55,11)	1,10 [0,97 ; 1,23]	2 194 (76,34)	0,97 [0,80 ; 1,18]	1 610 (56,02)	0,81 [0,74 ; 0,89]	1 954 (67,99)	0,92 [0,82 ; 1,03]	1 575 (54,80)	0,98 [0,89 ; 1,08]
P-valeur de tendance		0,1134		0,0560		<0,0001		0,3874		0,1852
Cancer cutané non basal (N = 686)										
0	82 (11,95)	1,00	19 (2,77)	1,00	121 (17,64)	1,00	71 (10,35)	1,00	95 (13,85)	1,00
0,5	209 (30,47)	0,95 [0,74 ; 1,23]	133 (19,39)	1,22 [0,75 ; 1,97]	141 (20,55)	1,10 [0,86 ; 1,41]	133 (19,39)	1,22 [0,92 ; 1,63]	174 (25,36)	1,36 [1,06 ; 1,75]
1	395 (57,58)	1,06 [0,83 ; 1,35]	534 (77,84)	1,30 [0,82 ; 2,06]	424 (61,81)	1,15 [0,94 ; 1,41]	482 (70,26)	1,16 [0,90 ; 1,49]	417 (60,79)	1,53 [1,22 ; 1,91]
P-valeur de tendance		0,3765		0,1306		0,5636		0,2296		0,0964

Suite page suivante

Localisation ^a	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Cancer du côlon-rectum (N = 481)										
0	67 (13,93)	1,00	18 (3,74)	1,00	124 (25,78)	1,00	65 (13,51)	1,00	88 (18,30)	1,00
0,5	163 (33,89)	0,96 [0,72 ; 1,27]	109 (22,66)	1,14 [0,69 ; 1,88]	92 (19,13)	0,73 [0,56 ; 0,96]	93 (19,33)	0,96 [0,70 ; 1,31]	112 (23,28)	0,97 [0,73 ; 1,28]
1	251 (52,18)	0,84 [0,63 ; 1,11]	354 (73,60)	1,07 [0,66 ; 1,73]	265 (55,09)	0,74 [0,59 ; 0,92]	323 (67,15)	0,84 [0,64 ; 1,10]	281 (58,42)	1,10 [0,86 ; 1,40]
P-valeur de tendance		0,1287		0,5604		0,0141		0,2654		0,9600
Hémopathies malignes (N = 411)										
0	45 (10,95)	1,00	19 (4,63)	1,00	84 (20,44)	1,00	46 (11,19)	1,00	79 (19,22)	1,00
0,5	134 (32,60)	1,10 [0,78 ; 1,54]	111 (27,01)	1,06 [0,65 ; 1,73]	77 (18,73)	0,87 [0,64 ; 1,19]	64 (15,57)	0,93 [0,64 ; 1,36]	99 (24,09)	0,97 [0,72 ; 1,30]
1	232 (56,45)	1,09 [0,78 ; 1,51]	281 (68,37)	0,75 [0,47 ; 1,20]	250 (60,85)	0,96 [0,75 ; 1,23]	301 (73,24)	1,11 [0,81 ; 1,52]	233 (56,69)	1,04 [0,81 ; 1,35]
P-valeur de tendance		0,7378		0,7817		0,5289		0,0785		0,9068
Cancer digestif, autre que côlon-rectum (N = 274)										
0	45 (16,42)	1,00	21 (7,66)	1,00	49 (17,88)	1,00	45 (16,42)	1,00	51 (18,61)	1,00
0,5	83 (30,29)	0,71 [0,49 ; 1,02]	61 (22,26)	0,54 [0,33 ; 0,88]	68 (24,82)	1,39 [0,96 ; 2,01]	40 (14,60)	0,60 [0,39 ; 0,92]	76 (27,74)	1,16 [0,81 ; 1,66]
1	146 (53,28)	0,73 [0,52 ; 1,03]	192 (70,07)	0,48 [0,30 ; 0,76]	157 (57,30)	1,12 [0,81 ; 1,56]	189 (68,98)	0,72 [0,52 ; 1,01]	147 (53,65)	1,05 [0,76 ; 1,45]
P-valeur de tendance		0,6899		0,0003		0,9660		0,2757		0,5737
Cancer de l'endomètre (N = 270)										
0	37 (13,70)	1,00	24 (8,89)	1,00	60 (22,22)	1,00	42 (15,56)	1,00	54 (20,00)	1,00
0,5	84 (31,11)	0,86 [0,58 ; 1,27]	74 (27,41)	0,57 [0,36 ; 0,90]	64 (23,70)	1,06 [0,74 ; 1,51]	40 (14,81)	0,64 [0,41 ; 0,98]	73 (27,04)	1,07 [0,75 ; 1,52]
1	149 (55,19)	0,89 [0,61 ; 1,29]	172 (63,70)	0,37 [0,24 ; 0,58]	146 (54,07)	0,81 [0,60 ; 1,10]	188 (69,63)	0,75 [0,53 ; 1,05]	143 (52,96)	0,98 [0,71 ; 1,34]
P-valeur de tendance		0,1766		<0,0001		0,4944		0,7444		0,4356
Cancer de la thyroïde (N = 267)										
0	30 (11,24)	1,00	17 (6,37)	1,00	45 (16,85)	1,00	28 (10,49)	1,00	59 (22,10)	1,00
0,5	89 (33,33)	1,18 [0,78 ; 1,79]	55 (20,60)	0,64 [0,37 ; 1,11]	46 (17,23)	0,97 [0,64 ; 1,46]	61 (22,85)	1,52 [0,97 ; 2,38]	72 (26,97)	0,97 [0,69 ; 1,37]
1	148 (55,43)	1,23 [0,82 ; 1,83]	195 (73,03)	0,58 [0,35 ; 0,97]	176 (65,92)	1,21 [0,87 ; 1,69]	178 (66,67)	1,26 [0,84 ; 1,88]	136 (50,94)	0,92 [0,68 ; 1,26]
P-valeur de tendance		0,3636		0,0007		0,1275		0,9719		0,1902

Suite page suivante

Localisation ^a	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Cancer de l'ovaire (N = 248)										
0	33 (13,31)	1,00	17 (6,37)	1,00	40 (16,13)	1,00	35 (14,11)	1,00	50 (20,16)	1,00
0,5	82 (33,06)	0,93 [0,62 ; 1,39]	55 (20,60)	1,07 [0,58 ; 1,99]	59 (23,79)	1,43 [0,96 ; 2,14]	42 (16,94)	0,82 [0,52 ; 1,28]	70 (28,23)	1,09 [0,76 ; 1,58]
1	133 (53,63)	0,86 [0,58 ; 1,27]	171 (68,95)	0,80 [0,44 ; 1,45]	149 (60,08)	1,24 [0,87 ; 1,76]	171 (68,95)	0,88 [0,61 ; 1,27]	128 (51,61)	0,95 [0,68 ; 1,33]
P-valeur de tendance		0,2351		0,2925		0,1494		0,4032		0,3863
Cancer du poumon (N = 213)										
0	79 (37,09)	1,00	302 (4,36)	1,00	53 (24,88)	1,00	32 (15,02)	1,00	59 (27,70)	1,00
0,5	62 (29,11)	0,30 [0,21 ; 0,41]	1450 (20,90)	1,09 [0,48 ; 2,44]	45 (21,13)	0,93 [0,62 ; 1,38]	45 (21,13)	1,06 [0,67 ; 1,67]	61 (28,64)	0,76 [0,53 ; 1,09]
1	72 (33,80)	0,20 [0,14 ; 0,27]	168 (78,87)	1,38 [0,64 ; 2,96]	115 (53,99)	0,94 [0,67 ; 1,32]	136 (63,85)	0,89 [0,60 ; 1,31]	93 (43,66)	0,53 [0,38 ; 0,74]
P-valeur de tendance		<0,0001		0,0035		0,4983		0,2813		0,0446

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes localisations de cancer sont définies dans le tableau 9 page 83.

b : Ajusté sur le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, l'utilisation de contraceptif oral, l'âge à la puberté, le nombre d'enfants, l'âge à la première grossesse à terme, le statut ménopausique, l'utilisation récente de THM et les quatre autres comportements de mode de vie.

Nous pouvons observer une diminution du risque de tout cancer chez les femmes n'ayant jamais fumé (HR = 0,92 [0,86 ; 0,99]) comparées aux femmes fumeuses, diminution qui ne se retrouve que pour le cancer du poumon, de façon plus prononcée (HR = 0,20 [0,14 ; 0,27]). De même le risque de tout cancer est diminué chez les femmes ayant un IMC compris entre 18,5 et 25 (HR = 0,84 [0,75 ; 0,95], comparées aux femmes obèses ou en dénutrition). Cette diminution de risque se retrouve pour les cancers digestifs, autres que côlon-rectum (HR = 0,48 [0,30 ; 0,76]), pour le cancer de l'endomètre (HR = 0,37 [0,24 ; 0,58]) et pour le cancer de la thyroïde (HR = 0,58 [0,35 ; 0,97]). Le risque de cancer est également diminué chez les femmes consommant moins d'un verre d'alcool par jour (HR = 0,90 [0,85 ; 0,95], comparées aux femmes en consommant plus de deux), résultat qui se retrouve pour les cancers du sein en postménopause (HR = 0,81 [0,74 ; 0,89]) et du côlon-rectum (HR = 0,74 [0,59 ; 0,92]). Une consommation d'au moins 5 portions de fruits et légumes est associée à une diminution du risque de cancer à la limite de la significativité statistique (HR = 0,93 [0,87 ; 1,00], comparée à une consommation de moins de 3,5 portions par jour). Aucune association statistiquement significative n'est observée entre la consommation de fruits et légumes et le risque de cancer de quelque localisation, que ce soit. Néanmoins, une tendance statistiquement significative est observée pour le cancer du sein en préménopause (p-valeur = 0,0310). Un niveau d'activité physique récréationnelle d'au moins 20 METs-h/semaine n'est pas statistiquement associé au risque de cancer, toutes localisations confondues, comparé à un niveau de moins de 10 METs-h/semaine, mais est associé à une diminution du risque de cancer du poumon (HR = 0,53 [0,38 ; 0,74]). *A contrario*, un niveau élevé d'activité physique est associé à une augmentation du risque de cancer cutané, autre que basocellulaire (HR = 1,53 [1,22 ; 1,91]). Cette augmentation de risque se retrouve à la fois pour les mélanomes (N = 304, HR = 1,78 [1,26 ; 2,52]) et les cancers cutanés non basal non mélanome (N = 382, HR = 1,34 [1,01 ; 1,82]) (tableau 46 en annexe page 231).

Les résultats concernant l'association entre le risque de cancer et l'indice de santé sont présentés dans le tableau 21.

Tableau 21 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de cancer, toutes localisations confondues et par site spécifique, cohorte E3N (N = 64 732).

	Catégories de l'indice de santé			
	[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Cancer, toutes localisations confondues (N = 6 938)				
N (%)	450 (6,48)	1 554 (22,40)	2 974 (42,87)	1 960 (28,25)
HR [IC 95 %] ^c	1,00	0,88 [0,79 ; 0,98]	0,86 [0,78 ; 0,95]	0,81 [0,73 ; 0,89]
HR (augmentation de 1 point)		0,93 [0,91 ; 0,96]		
P-valeur de tendance		<0,001		
Cancer du sein en préménopause (N = 609)				
N (%)	47 (7,72)	141 (23,15)	268 (44,01)	153 (25,12)
HR [IC 95 %] ^a	1,00	0,81 [0,59 ; 1,13]	0,90 [0,66 ; 1,22]	0,80 [0,58 ; 1,12]
HR (augmentation de 1 point)		0,97 [0,89 ; 1,06]		
P-valeur de tendance		0,288		
Cancer du sein en postménopause (N = 2 874)				
N (%)	169 (5,88)	623 (21,68)	1 284 (44,68)	798 (27,76)
HR [IC 95 %] ^b	1,00	0,95 (0,80 ; 1,12)	0,99 (0,85 ; 1,17)	0,87 (0,74 ; 1,03)
HR (augmentation de 1 point)		0,96 [0,92 ; 1,00]		
P-valeur de tendance		0,016		
Cancer cutané non basocellulaire (N = 686)				
N (%)	26 (3,79)	125 (18,22)	283 (41,25)	252 (36,74)
HR [IC 95 %] ^c	1,00	1,22 [0,80 ; 1,87]	1,40 [0,94 ; 2,09]	1,75 [1,17 ; 2,62]
HR (augmentation de 1 point)		1,19 [1,10 ; 1,30]		
P-valeur de tendance		<0,001		
Cancer du côlon-rectum (N = 481)				
N (%)	33 (6,86)	110 (22,87)	207 (43,04)	131 (27,23)
HR [IC 95 %] ^c	1,00	0,83 [0,56 ; 1,22]	0,76 [0,53 ; 1,10]	0,66 [0,45 ; 0,97]
HR (augmentation de 1 point)		0,85 [0,78 ; 0,94]		
P-valeur de tendance		0,013		
Hémopathies malignes (N = 411)				
N (%)	18 (4,38)	97 (23,60)	160 (38,93)	136 (33,09)
HR [IC 95 %] ^c	1,00	1,33 [0,80 ; 2,20]	1,08 [0,66 ; 1,76]	1,27 [0,77 ; 2,08]
HR (augmentation de 1 point)		0,99 [0,88 ; 1,11]		
P-valeur de tendance		0,894		
Cancers digestifs, autres que côlon-rectum (N = 274)				
N (%)	22 (8,03)	63 (22,99)	115 (41,97)	74 (27,01)
HR [IC 95 %] ^c	1,00	0,72 [0,44 ; 1,16]	0,65 [0,41 ; 1,04]	0,59 [0,36 ; 0,95]
HR (augmentation de 1 point)		0,86 [0,75 ; 0,99]		
P-valeur de tendance		0,036		

Suite page suivante

RÉSULTATS - PARTIE 1

	Catégories de l'indice de santé			
	[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Cancer de l'endomètre (N = 270)				
N (%)	26 (9,63)	66 (24,44)	105 (38,89)	73 (27,04)
HR [IC 95 %] ^c	1,00	0,61 [0,39 ; 0,97]	0,48 [0,31 ; 0,73]	0,45 [0,29 ; 0,71]
HR (augmentation de 1 point)		0,79 [0,69 ; 0,90]		
P-valeur de tendance		<0,001		
Cancer de la thyroïde (N = 267)				
N (%)	16 (5,99)	65 (24,34)	103 (38,58)	83 (31,09)
HR [IC 95 %] ^c	1,00	1,08 [0,62 ; 1,87]	0,92 [0,54 ; 1,56]	1,08 [0,63 ; 1,86]
HR (augmentation de 1 point)		1,04 [0,90 ; 1,20]		
P-valeur de tendance		0,573		
Cancer des ovaires (N = 248)				
N (%)	15 (6,05)	59 (23,79)	101 (40,73)	73 (29,43)
HR [IC 95 %] ^c	1,00	0,99 [0,56 ; 1,75]	0,87 [0,51 ; 1,51]	0,89 [0,51 ; 1,55]
HR (augmentation de 1 point)		0,93 [0,81 ; 1,07]		
P-valeur de tendance		0,297		
Cancer du poumon (N = 213)				
N (%)	32 (15,02)	61 (28,64)	85 (39,90)	35 (16,44)
HR [IC 95 %] ^c	1,00	0,47 [0,31 ; 0,72]	0,33 [0,22 ; 0,49]	0,19 [0,11 ; 0,30]
HR (augmentation de 1 point)		0,60 [0,52 ; 0,69]		
P-valeur de tendance		<0,001		

Note : Les classifications utilisées pour repérer les différentes localisations de cancer sont définies dans le tableau 9 page 83.

a : Ajusté sur le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, l'utilisation de contraceptif oral, l'âge à la puberté, le nombre d'enfants et l'âge à la première grossesse à terme.

b : Ajusté sur (a) + l'utilisation récente de THM.

c : Ajusté sur (a) + le statut ménopausique et l'utilisation récente de THM.

Comparées aux femmes qui adhèrent le moins aux recommandations de santé publique (indice de santé entre 0 et 2), les femmes qui y adhèrent le plus (indice de santé entre 4,5 et 5) sont à risque diminué de cancer, toutes localisations confondues (HR = 0,81 [0,73 ; 0,89], p-valeur de tendance < 0,001). La diminution du risque est observée dès l'adhésion à une recommandation de santé publique supplémentaire (HR = 0,93 [0,91 ; 0,96]). L'association entre l'indice de santé et le risque de cancer est observée pour plusieurs localisations spécifiques de tumeur : cancer du poumon (HR = 0,19 [0,11 ; 0,30], p < 0,001), cancer de l'endomètre (HR = 0,45 [0,29 ; 0,71], p < 0,001), cancer digestif autre que côlon-rectum (HR = 0,59 [0,36 ; 0,95], p = 0,036), et cancer du côlon-rectum (HR = 0,66 [0,45 ; 0,97], p = 0,013). Une tendance statistiquement significative est observée pour le cancer du sein en postménopause (p = 0,016), mais la réduction de risque pour les femmes qui adhèrent le mieux aux recommandations de santé comparées à celles qui y adhèrent le moins est à la limite de la significativité statistique (HR = 0,87 [0,74 ; 1,03],

$p = 0,016$). *A contrario*, une augmentation de risque de cancer cutané autre que basocellulaire est observée chez les femmes adhérant le mieux aux recommandations (HR = 1,75 [1,17 ; 2,62], $p < 0,001$). Cette augmentation de risque est observée pour les non-mélanomes (HR = 2,16 [1,17 ; 4,00]), mais seule la tendance reste statistiquement significative pour les mélanomes ($p = 0,017$) (tableau 48 en annexe page 232). Les associations entre l'adhésion aux recommandations de santé et le risque de cancer des autres localisations (sein en préménopause, hémopathies malignes, thyroïde et ovaires) ne sont pas statistiquement significatives.

1.3.c. Proportions de cas de cancer attribuables aux comportements individuels de mode de vie ou à leur combinaison

Les proportions de cas de cancer (toutes localisations confondues ou par site spécifique) attribuables à la non-adhésion à chacune des cinq recommandations de santé publique ou à un mode de vie malsain (combinant les recommandations) et leur intervalle de confiance à 95 % sont présentées dans le tableau 22.

Tableau 22 : Proportions [IC 95 %] de cas de cancer, toutes localisations confondues et par site spécifique, attribuables à la non-adhésion aux recommandations individuelles de santé publique et à leur combinaison, cohorte E3N (N = 64 732).

	Proportions de cancers attribuables aux comportements de mode de vie ^a					Proportions de cancers attribuables à un mode de vie non sain ^b
	Tabagisme actif	IMC	Consommation d'alcool	Consommation de fruits et légumes	Niveau d'activité physique récréationnelle	
Cancer, toutes localisations confondues (N = 6 938)	1,3 [-1,8 ; 4,4]	1,4 [-0,5 ; 3,4]	3,2 [0,7 ; 5,7]	1,5 [-0,8 ; 3,8]	0,0 [-2,8 ; 2,8]	6,3 [2,2 ; 10,3]
Cancer du sein en préménopause (N = 609)	2,0 [-5,2 ; 9,3]	-6,3 [-9,9 ; -2,6]	2,1 [-3,4 ; 7,6]	2,5 [-3,0 ; 8,0]	-1,0 [-7,4 ; 5,3]	-2,3 [-12,0 ; 7,5]
Cancer du sein en postménopause (N = 2 874)	-2,6 [-7,0 ; 1,8]	-0,3 [-3,2 ; 2,6]	5,2 [1,6 ; 8,9]	2,4 [-0,5 ; 5,6]	-0,8 [-4,8 ; 3,3]	6,3 [0,5 ; 12,1]
Cancer de la peau non basal (N = 686)	-4,5 [-14,1 ; 5,3]	-2,0 [-8,0 ; 4,1]	-3,6 [-11,5 ; 4,3]	-0,6 [-7,8 ; 6,5]	-10,6 [-19,4 ; -1,7]	-21,5 [-34,6 ; -7,5]
Cancer du côlon-rectum (N = 481)	7,0 [-5,0 ; 18,8]	1,0 [-6,3 ; 8,2]	6,6 [-2,9 ; 16,0]	5,1 [-4,1 ; 14,1]	-5,0 [-15,5 ; 5,6]	12,8 [-2,4 ; 27,4]
Hémopathies malignes (N = 411)	-0,7 [-13,2 ; 11,8]	8,6 [0,2 ; 16,9]	-1,0 [-10,9 ; 9,0]	-4,6 [-13,3 ; 4,0]	-2,9 [-14,3 ; 8,6]	-6,7 [-23,6 ; 10,6]
Cancers digestifs, autre que côlon-rectum (N = 274)	3,9 [-11,7 ; 19,3]	6,1 [-3,8 ; 15,9]	2,5 [-10,6 ; 15,4]	1,7 [-9,5 ; 12,8]	2,0 [-12,3 ; 16,3]	12,1 [-8,4 ; 31,6]
Cancer de l'endomètre (N = 270)	0,7 [-15,6 ; 17,0]	13,3 [3,7 ; 22,5]	9,8 [-3,1 ; 22,3]	1,6 [-10,3 ; 13,5]	2,9 [-11,9 ; 17,6]	15,0 [-5,8 ; 34,6]
Cancer de la thyroïde (N = 267)	-3,6 [-18,1 ; 11,0]	4,6 [-6,0 ; 15,1]	-7,8 [-19,7 ; 4,2]	1,2 [-10,0 ; 12,4]	3,0 [-11,3 ; 17,1]	-7,5 [-28,7 ; 14,4]
Cancer de l'ovaire (N = 248)	4,5 [-11,6 ; 20,3]	7,2 [-3,5 ; 17,8]	-0,6 [-13,9 ; 12,8]	0,5 [-11,4 ; 12,3]	4,7 [-10,4 ; 19,6]	2,9 [-18,8 ; 24,3]
Cancer du poumon (N = 213)	41,8 [24,8 ; 56,3]	-5,7 [-16,6 ; 5,4]	1,2 [-14,2 ; 16,5]	5,4 [-8,5 ; 19,1]	22,5 [6,3 ; 37,5]	47,5 [26,8 ; 64,1]

Note : Les classifications utilisées pour repérer les différentes localisations de cancer sont définies dans le tableau 9 page 83.

a : Pourcentage de cas de cancer attribuables à la non-adhésion aux recommandations individuelles de santé publique (i.e. ne jamais avoir fumé, avoir un IMC compris entre 18,5 et 25, consommer moins d'un verre d'alcool par jour, consommer au moins 5 portions de fruits et légumes par jour et pratiquer l'équivalent d'au moins 20 METs-h/semaine). Les analyses concernant un des comportements sont ajustées sur le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, l'utilisation de contraceptif oral, l'âge à la puberté, le nombre d'enfants et l'âge à la première grossesse à terme et les quatre autres comportements de mode de vie. Les analyses concernant le cancer du sein en préménopause ne sont ajustées ni sur le statut ménopausique ni sur l'utilisation récente de THM, et celles concernant le cancer du sein en postménopause ne sont pas ajustées sur le statut ménopausique.

b : Pourcentage de cas de cancer attribuables à un mode de vie malsain (i.e. un indice de santé strictement inférieur à 4,5), ajustés sur (a), à l'exception des comportements individuels de mode de vie.

Les proportions de cas de cancer attribuables à un mode de vie malsain (*i.e.* un indice de santé strictement inférieur à 4,5) sont de 6,3 % [2,2 % ; 10,3 %] pour l'ensemble des localisations de cancer, 6,3 % [0,5 % ; 12,1 %] pour le cancer du sein en postménopause et 47,5 % [26,8 % ; 64,1 %] pour le cancer du poumon. À l'inverse, on observe le diagnostic de 21,5 % [7,5 % ; 34,6 %] de cas de cancer cutané non basocellulaire supplémentaires pour l'adhésion aux recommandations de santé publique. Cette proportion est similaire pour les mélanomes et les non-mélanomes (tableau 47 en annexe page 231). Les proportions de cas de cancer d'autres localisations attribuables à la non-adhésion aux recommandations de santé publique ne sont pas statistiquement significatives.

Les proportions de cas de cancer attribuables à la non-adhésion à chacune des recommandations individuelles de santé publique varient selon la localisation du cancer. On observe que 41,8 % [24,8 % ; 56,3 %] des cas de cancer du poumon seraient attribuables au tabagisme actif (actuel et passé), que 13,3 % [3,7 % ; 22,5 %] des cas de cancer de l'endomètre et 8,6 % [0,2 % ; 16,9 %] des cas d'hémopathies malignes seraient attribuables à un IMC inférieur à 18,5 ou supérieur à 25, alors que 6,3 % [2,6 % ; 9,9 %] de cas supplémentaires de cancer du sein en préménopause seraient diagnostiqués si les femmes avaient toutes eu un IMC entre 18,5 et 25. On dénombre également 3,2 % [0,7 % ; 5,7 %] des cas de cancer, toutes localisations confondues et 5,2 % [1,6 % ; 8,9 %] des cas de cancer du sein en postménopause attribuables à une consommation d'alcool supérieure à 1 verre par jour. Enfin, 22,5 % [6,3 % ; 37,5 %] des cas de cancer du poumon seraient attribuables à un niveau d'activité physique inadéquat (inférieur à 10 METs-h/semaine), alors que 10,6 % [1,7 % ; 19,4 %] de cas supplémentaires de cancer cutané non basocellulaire seraient diagnostiqués si les femmes pratiquaient toutes une activité physique équivalente à au moins 20 METs-h/semaine. Cette dernière proportion est plus marquée pour les mélanomes (14,9 % [1,5 % ; 27,8 %]) que pour les non-mélanomes (7,3 [-4,7 % ; 19,1 %]) (tableau 47 en annexe page 231). Les autres fractions attribuables n'atteignent pas la significativité statistique.

1.4. Discussion

Dans cette étude portant sur l'association entre l'adhésion à cinq recommandations de santé publique et le risque de cancer, toutes localisations confondues et par site spécifique, nous avons pu observer que l'adhésion à une simple recommandation a un impact modéré sur la proportion de cas de cancer attribuables à la non-adhésion à ces recommandations. Cependant, l'adhésion à un mode de vie sain, impliquant l'adhésion conjointe aux recommandations de santé publique

concernant le tabagisme actif, l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique récréationnelle, pourrait permettre d'éviter jusqu'à 6,3 % de cas de cancer. Cette proportion varie selon la localisation du cancer et atteint 47,5 % pour le cancer du poumon. Les comportements associés à la proportion de cas de cancer attribuables à la non-adhésion à une seule recommandation de santé publique la plus élevée sont l'IMC pour le cancer de l'endomètre et les hémopathies malignes, la consommation d'alcool pour le cancer du sein en postménopause, le tabagisme actif pour le cancer du poumon et le niveau d'activité physique récréationnelle pour le cancer cutané non basocellulaire. Le cancer est une maladie multifactorielle, dont les facteurs de risque incluent de multiples causes endogènes (génétique, hormones, système immunitaire, métabolisme) et exogènes (mode de vie, environnement). Ainsi, les comportements de mode de vie analysés représentent cinq facteurs parmi une multitude de facteurs de risque associés aux cancers.

1.4.a. Confrontation à la littérature

Les quelques études qui ont quantifié l'impact du mode de vie sur le risque de cancer ont rapporté une diminution du risque de cancer, toutes localisations confondues (117-119;121-123), du côlon-rectum (119;120;123;124), du sein (119;123;125), de l'endomètre (119;123), de l'appareil digestif (116;119) et du poumon (119), chez les femmes qui adhèrent le mieux aux recommandations de santé publique comparées aux femmes qui y adhèrent le moins. Il a été estimé que 11 à 31 % des cas de cancer, toutes localisations confondues (117-119), et 6 % des cas de cancer du côlon-rectum (120), seraient attribuables à la non-adhésion aux recommandations de santé publique. La variation des expositions à travers le monde et l'utilisation de différentes méthodes d'estimation des proportions de cas attribuables peuvent contribuer à la variabilité des résultats. De plus, la formule de Levin (241), largement utilisée dans la littérature, produit des résultats biaisés lorsqu'estimés à partir de risques relatifs multi-ajustés (234), alors que la formule de Spiegelman (236), utilisée dans le présent travail, produit des estimateurs de fractions attribuables mieux adaptés.

Les facteurs de mode de vie sain analysés dans cette partie semblent jouer un rôle mineur dans l'étiologie du cancer de la thyroïde, pour lequel l'exposition aux radiations est le facteur de risque le plus clairement établi dans la littérature (242). En complément, l'HR obtenu pour l'association entre l'IMC et le risque de cancer de la thyroïde est statistiquement significatif, résultat en accord avec une récente méta-analyse sur le sujet concluant à une augmentation statistiquement significative du risque de cancer de la thyroïde avec le surpoids et l'obésité chez les femmes (243). Cependant, l'association entre l'IMC et le risque de cancer de la thyroïde serait due à une carence

en iode, fréquente chez les personnes obèses (47). La forte association observée entre le tabagisme actif et le cancer du poumon est en accord avec une récente méta-analyse sur le sujet (244), et souligne la nature causale de cette association. D'un autre côté, il est surprenant de ne pas observer une association statistiquement significative entre le niveau d'activité physique récréationnelle et le risque de cancer du côlon-rectum ou du sein en postménopause alors que des suggestions ont été avancées sur ces deux associations (47).

L'augmentation du nombre de cancers cutané autre que basocellulaire diagnostiqués si les femmes avaient toutes eu un niveau d'activité physique supérieur à 20 METs-h/semaine persiste lorsque les analyses sont ajustées sur le niveau de doses de radiations ultraviolettes journalières du département de résidence, moyennées sur l'année (PAF = -9,2 [-17,0 ; -1,2]). Ceci suggère un effet résiduel confondant de l'exposition au soleil individuelle, due aux activités extérieures, qui a été associée à une augmentation du risque de cancer de la peau (245). Ce résultat souligne l'importance de recommander à la population d'éviter les coups de soleil et l'exposition excessive au soleil sans protection (246).

Nous avons observé qu'approximativement 22 % des cas de cancer du poumon étaient attribuables à un niveau d'activité physique inadéquat, résultat en accord avec la relation inverse entre le risque de cancer du poumon et le niveau d'activité physique observée dans deux méta-analyses (247;248). Cependant, l'association pourrait refléter un possible effet causal inverse car les personnes atteintes de maladies chroniques bronchiques, qui sont à risque augmenté de cancer du poumon, seraient moins enclines à pratiquer une activité physique (47).

En accord avec les effets individuels des facteurs de mode de vie observés, des méta-analyses précédemment publiées concluent que les femmes devraient limiter leur consommation d'alcool afin de prévenir le risque de cancer du sein (249;250), que l'IMC est inversement associé au risque de cancer du sein en préménopause (251) et qu'un IMC en dessous de 25 serait associé à un risque diminué de cancer de l'endomètre (252). Un IMC inférieur à 25 a également été associé à un risque diminué de lymphome non-hodgkinien (253), de leucémie (254) et de myélome multiple (255), des hémopathies malignes dont peu de facteurs de risque ont été établis dans la littérature (256). Enfin, en accord avec les résultats observés dans cette partie, plusieurs études ont observé que le maintien d'un poids sain, la pratique d'une activité physique, la consommation d'alcool et le tabagisme expliqueraient la plupart des associations observées entre le mode de vie et le risque de cancer (117;119;122;123).

1.4.b. Limites de l'étude

La catégorisation de l'indice de santé pourrait être discutée, étant donné que l'indice est fortement dépendant des seuils choisis. L'approche *a priori* qui a été choisie permet de s'assurer que les résultats obtenus n'ont pas influencé notre jugement et que l'indice construit est le plus proche possible des recommandations de santé publique. Cependant, afin d'évaluer l'influence du choix des seuils, des analyses de sensibilité ont été effectuées en modifiant chacun des cinq seuils et les résultats obtenus étaient similaires aux analyses principales.

Comme toute étude de cohorte, E3N est encline à l'effet de « cohorte saine ». Cependant, étant donné que les proportions de cas de cancer attribuables au mode de vie ont été évaluées à partir des prévalences des expositions et de la force de l'association, les résultats obtenus seraient plus accentués dans la population générale, dont les prévalences des expositions sont plus élevées.

1.4.c. Conclusion de l'étude

Plusieurs études, précédemment publiées, ont analysé de manière indépendante l'association d'expositions liées au mode de vie sur le risque de cancer avec des critères d'évaluation similaires, pourtant très peu ont quantifié l'impact du mode de vie sur la population en utilisant une méthodologie adéquate. Alors que l'adhésion à une seule recommandation de santé semble avoir un impact modéré sur la proportion de cas de cancer potentiellement évitables, la combinaison de plusieurs recommandations – être non-fumeurs, modérer sa consommation d'alcool, consommer la quantité recommandée de fruits et légumes, être physiquement actif et avoir un IMC dans la norme de recommandation de l'OMS – pourrait diminuer substantiellement le nombre de cas de cancer diagnostiqués, particulièrement le nombre de cas de cancer du poumon et du sein en postménopause. Ces résultats devraient encourager à renforcer les recommandations de santé publique concernant ces cinq comportements afin d'améliorer la prévention du cancer.

Partie 2 Estimation du poids relatif de différents facteurs de risque établis de cancer du sein

2.1. Introduction

Le cancer du sein est la tumeur maligne la plus fréquemment diagnostiquée chez les femmes, avec un taux d'incidence atteignant 88,0 pour 100 000 femmes-années en 2012 en France (section Introduction, partie 1.2.a page 33) (2). Les facteurs de risque de cancer du sein ont été largement explorés, et inclus des facteurs reproductifs, génétiques, anthropométriques et de mode de vie (section Introduction, partie 1.2.d page 38). Ceux-ci peuvent être comportementaux (comme la consommation d'alcool) ou non-comportementaux (comme les antécédents familiaux de cancer du sein) (257-259). Des analyses évaluant la proportion de cas de cancer du sein attribuables à la combinaison de plusieurs facteurs de risque ont été précédemment publiées chez des femmes de divers pays, mais très peu d'entre elles ont estimé l'effet combiné de plus de dix facteurs de risque simultanément, se fondent sur des études prospectives ou prennent en compte le statut ménopausique des femmes (section Introduction, partie 3.2.c page 62).

L'objectif de cette partie est d'estimer les fractions attribuables de plusieurs types de cancers du sein associées aux facteurs de risque comportementaux et non-comportementaux, de manière individuelle et combinée.

2.2. Matériel et méthodes

2.2.a. Population d'étude

Les femmes sont suivies de la date de remplissage du questionnaire alimentaire Q3 (1993) jusqu'au premier événement observé entre la date de diagnostic d'un cancer autre que le cancer cutané basal, la date de réponse au dernier questionnaire rempli et la date d'envoi du dernier questionnaire disponible (Q9, 06/2008). Les femmes non-répondantes au questionnaire alimentaire (N = 24 473), celles qui ont été diagnostiquées avec un cancer, autre qu'un cancer cutané basal, avant le début du suivi (N = 4 705), les femmes sans suivi après Q3 (N = 769), les cas de cancer sans date de diagnostic (N = 54) et les femmes présentant des valeurs extrêmes pour le rapport apport énergétique observé sur apport énergétique requis (premier et dernier percentile, N = 1 360) ont été exclues des analyses. La population d'analyse comporte ainsi 67 634 femmes, âgées de 42 à 72 ans au début du suivi.

2.2.b. Facteurs de risque comportementaux et non-comportementaux

Les facteurs de risque non-comportementaux concernent les facteurs sur lesquels il n'est pas possible d'agir directement et les facteurs de risque comportementaux tous les autres. La collecte de l'ensemble des facteurs comportementaux (niveau d'activité physique total, tabagisme actif, nombre d'enfants, âge à la première grossesse à terme, durée d'allaitement cumulée, silhouette au moment de la puberté, profil alimentaire, consommation d'alcool, vitamine D alimentaire et exposition aux UV solaires, IMC et utilisation de contraceptifs oraux, progestatifs seuls et THM) et non-comportementaux (antécédents personnels de maladie bénigne du sein, antécédents familiaux de cancer du sein dans la famille au premier degré, niveau d'étude, taille à l'âge adulte, âge à la puberté, poids à la naissance, âge à la ménopause et statut ménopausique) a été décrite dans la section Matériel, partie 4 page 94. Les catégorisations et les références utilisées pour l'ensemble de ces variables sont présentées dans le tableau 42 (en annexe page 222) et ont été établies à partir de la littérature (257-259).

2.2.c. Méthodologie statistique

Les associations entre les variables décrites dans le paragraphe précédent et le risque de cancer du sein ont été estimées à l'aide de modèles de régression de Cox, en utilisant l'âge comme échelle de temps (227), et différenciées selon le statut ménopausique des femmes. Les analyses ont ensuite été réalisées séparément pour les différents types de cancer du sein suivants : HER2+, HER2-, RO+, RO-, HER2+/RO+, HER2+/RO-, HER2-/RO+ et HER2-/RO-. Pour ces analyses selon les types de cancer du sein, les cas de cancer pour lesquels l'expression de la protéine HER2 et/ou le statut RO est(sont) inconnu(s) ont été exclus (tableau 23). Les informations sur le statut ménopausique, le statut tabagique, l'utilisation de THM, l'IMC, les antécédents personnels de maladie bénigne du sein et l'utilisation de pilule contraceptive et de progestatif ont été mises à jour à chaque questionnaire, tandis que tous les autres facteurs de risque ont été estimés au début du suivi ou avant, sauf pour le poids de naissance qui a été évalué à Q7 (2002). Les analyses ont également été ajustées sur les antécédents de mammographie, mises à jour à chaque questionnaire, et ont été stratifiées sur la génération de naissance (section Méthodes statistiques, la partie 2.9 page 116). Les proportions de cas attribuables à une seule recommandation ou à la combinaison des recommandations ont été estimées à partir de la formule de Spiegelman (section Méthodes statistiques, partie 3 page 130) (236). Étant donné que l'estimation des fractions attribuables est sensible aux choix des seuils (260), les différentes variables ont été catégorisées afin d'avoir, dans la mesure du possible, au moins 20 % de la population d'analyse dans la catégorie de référence.

Comme les valeurs manquantes pour les variables sur les antécédents familiaux de cancer du sein et le poids de naissance représentaient plus de 5 % de la population d'analyse, des catégories manquantes ont été créées pour ces deux variables. Pour l'ensemble des autres facteurs inclus dans les analyses, les valeurs manquantes ont été imputées à la médiane pour les variables quantitatives ou à la classe modale pour les variables qualitatives.

2.3. Résultats

2.3.a. Caractéristiques de la population d'analyse

Les 67 634 femmes de la population d'analyse ont été suivies pendant un total de 876 468 femmes-années (avec un suivi médian de 7 ans pour les cas de cancer invasif et de 13 pour les non-cas). Parmi ces femmes, 3 635 ont été diagnostiquées avec un cancer du sein, dont 497 n'étaient pas encore ménopausées au moment du diagnostic et 3 138 l'étaient. La distribution des différents types de cancer du sein diagnostiqués dans la population d'analyse est présentée dans le tableau 23.

Tableau 23 : Types de cancer du sein diagnostiqués dans la population d'analyse, cohorte E3N (N = 67 634).

Types de cancer du sein	Effectif	Pourcentage
Sein en préménopause	497	13,67
Sein en postménopause	3 138	86,33
Sein en préménopause RO+	314	76,40
Sein en préménopause RO-	97	23,60
Sein en postménopause RO+	2 050	81,48
Sein en postménopause RO-	466	18,52
Sein en préménopause HER2+	8	29,63
Sein en préménopause HER2-	19	70,37
Sein en postménopause HER2+	178	19,24
Sein en postménopause HER2-	747	80,76
Sein en postménopause RO+/HER2+	113	12,39
Sein en postménopause RO+/HER2-	629	68,97
Sein en postménopause RO-/HER2+	63	6,91
Sein en postménopause RO-/HER2-	107	11,73

Plus de la moitié des cancers du sein présentent des récepteurs à l'œstrogène, que ce soit en préménopause ou en postménopause. L'information concernant l'expression de la protéine HER2 est connue pour seulement 27 cas de cancer en préménopause, ce qui empêche d'effectuer des analyses spécifiques selon la surexpression de cette protéine. Parmi les cancers en postménopause dont l'information sur l'expression de la protéine HER2 est disponible, la majorité ne présente pas de surexpression de cette protéine.

Les caractéristiques des femmes de la population d'analyse selon leur statut cas/non-cas sont présentées dans le tableau 42 (en annexe page 222).

Nous pouvons observer que les cas de cancers présentent plus fréquemment au moins un antécédent familial de cancer du sein chez les parents au premier degré, ont plus souvent été précédemment diagnostiqués avec une maladie bénigne du sein, ont davantage tendance à avoir eu leurs premières règles précocement (en particulier pour les cas de cancer du sein en préménopause), ont eu leur premier enfant plus tardivement, pratiquent moins d'activité physique, consomment plus d'alcool (en particulier pour les cas de cancer du sein en postménopause) et ont un IMC plus élevé (en particulier les cas de cancer du sein en postménopause) que les non-cas de cancer du sein.

2.3.b. Proportions de cas de cancer du sein attribuables à un facteur de risque

Le tableau 42 (en annexe page 222) présente les HR (avec les IC à 95 %) du risque de cancer du sein en préménopause et en postménopause, associés à chaque facteur comportemental et non-comportemental pris en considération pour le calcul des fractions attribuables individuelles et combinées. La figure 38 présente les proportions de cas de cancers du sein en préménopause et en postménopause attribuables à un facteur de risque et à la combinaison des facteurs de risque, exprimées en pourcentage.

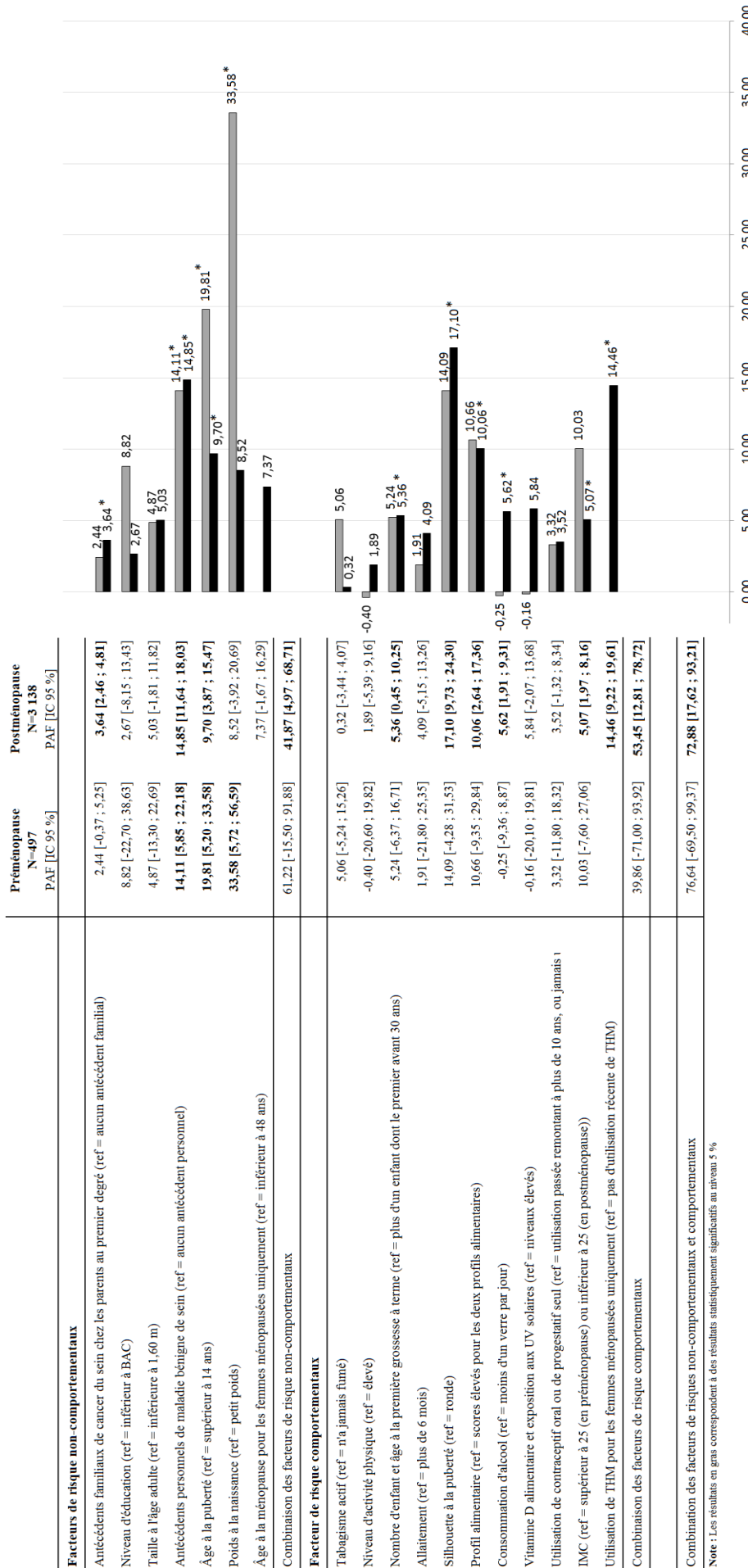


Figure 38 : Proportions de cas de cancer du sein en préménopause et en postménopause attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634).

- ***Proportions de cas de cancer du sein en préménopause attribuables à un facteur de risque***

Les estimations atteignant la significativité statistique sont les proportions de cas attribuables à la présence d'antécédents de maladie bénigne du sein (14,1 %), à un âge à la puberté précoce (19,8 %) et à un poids normal ou un gros poids de naissance (33,6 %). Ainsi environ un cas de cancer sur, respectivement, 7, 5 et 3 est attribuable à ces trois facteurs de risque.

- ***Proportions de cas de cancer du sein en postménopause attribuables à un facteur de risque***

Les estimations atteignant la significativité statistique sont les proportions de cas attribuables à la présence d'antécédents familiaux de cancer du sein (3,6 %), à un âge à la puberté précoce (9,7 %) et à la présence d'antécédents de maladie bénigne du sein (14,9 %) en ce qui concerne les facteurs de risque non-comportementaux ; et attribuables à un IMC inférieur à 25 (5,1 %), à la nulliparité ou un âge à la première grossesse à terme tardif (5,4 %), à une consommation élevée d'alcool (5,6 %), à un régime alimentaire déséquilibré (10,1 %), à l'utilisation récente de THM (14,5 %) et à une fine silhouette à la puberté (17,1 %) en ce qui concerne les facteurs de risque comportementaux.

2.3.c. Proportions de cas de cancer du sein attribuables à la combinaison des facteurs de risque

La proportion de cas de cancer du sein en préménopause attribuables à la combinaison des variables incluses dans les analyses a été estimée à 76,6 % (-69,5 % ; 99,4 %). Cette proportion est de 61,2 % (-15,5 % ; 91,9 %) pour la combinaison des facteurs de risque non-comportementaux et de 39,9 % (-71,0 % ; 93,9 %) pour la combinaison des facteurs de risque comportementaux (figure 38).

La proportion de cas de cancer du sein en postménopause attribuables à la combinaison des variables incluses dans les analyses a été estimée à 72,9 % (17,6 % ; 93,2 %). Cette proportion est de 41,9 % (5,0 % ; 68,7 %) pour la combinaison des facteurs de risque non-comportementaux et de 53,5 % (12,8 % ; 78,7 %) pour la combinaison des facteurs de risque comportementaux (figure 38).

2.3.d. Proportions de cas de cancer du sein en préménopause attribuables aux facteurs de risque selon le statut RO

Les estimations des proportions de cas de cancer du sein en préménopause attribuables à chaque facteur de risque individuel et à leur combinaison, selon le statut RO des cancers, sont présentées dans le tableau 49 (en annexe page 233).

Au total, trois-quarts des cas de cancers en préménopause seraient attribuables aux facteurs de risque inclus dans les analyses dans chaque strate de statut de récepteur hormonal. Ce résultat provient majoritairement des fractions attribuables individuelles suivantes : le surpoids et l'obésité (25,0 %) et la présence d'antécédents personnels de maladie bénigne du sein (16,2 %) pour les cancers du sein RO+, et un poids normal ou un gros poids de naissance (61,0 %) et un âge à la puberté précoce (38,6 %) pour les cancers du sein RO-.

2.3.e. Proportions de cas de cancer du sein en postménopause attribuables aux facteurs de risque selon le statut RO et l'expression de la protéine HER2

Les estimations des proportions de cas de cancer du sein en postménopause attribuables à chaque facteur de risque et à leur combinaison selon le statut RO ou l'expression de la protéine HER2 des cancers sont présentées dans le tableau 50 (en annexe page 234) ; les estimations selon la combinaison de ces deux informations sont présentées dans le tableau 51 (en annexe page 235).

Un total de 80,2 % (20,3 % ; 96,4 %) des cancers du sein RO+ et 53,6 % (-91,6 % ; 99,2 %) des cancers du sein RO- en postménopause sont attribuables à la combinaison de l'ensemble des facteurs de risque inclus dans les analyses. Les proportions attribuables à un seul facteur de risque qui atteignent la significativité statistique pour les analyses portant sur l'ensemble des cancers du sein en postménopause sont également statistiquement significatives pour les cancers du sein RO+. Les proportions attribuables à une grande taille adulte (8,7 %) et à un âge à la ménopause tardif (12,8 %) deviennent statistiquement significatives pour les cancers en postménopause RO+. En ce qui concerne le cancer du sein RO-, seule la proportion de cas attribuables à la présence d'antécédents de maladie bénigne du sein atteint la significativité statistique (15,3 %).

En ce qui concerne la stratification selon le statut RO et l'expression de la protéine HER2, conjointement, les proportions de cas de cancers du sein en postménopause attribuables à l'ensemble des facteurs de risque considérés dans les analyses sont comprises entre 57,8 % pour les cancers RO-/HER2- et 87,1 % pour les cancers RO+/HER2-. Les résultats concernant les

cancers HER2- et RO+/HER2- sont similaires aux analyses principales, exceptions faites des proportions de cas attribuables à un âge à la puberté précoce, à la nulliparité ou un âge tardif à la première grossesse à terme et à la consommation élevée d'alcool, qui ne sont plus statistiquement significatives. En ce qui concerne les cancers HER2+, seule la proportion de cas attribuables à la présence d'antécédents de maladie bénigne du sein reste statistiquement significative (11,7 %). Enfin, nous pouvons observer des estimations statistiquement significatives concernant les proportions de cas attribuables à un niveau d'étude élevé pour les cancers HER2- (25,5 %) et pour les cancers RO+/HER2- (27,1 %) et à un âge à la ménopause tardif pour les cancers HER2+ (39,1 %).

2.4. Discussion

Cet article permet d'ajouter des arguments à la littérature en faveur d'une proportion élevée de cas de cancer du sein évitables, que ce soit globalement ou selon le type de cancer du sein. À notre connaissance, il s'agit de la première étude estimant des fractions attribuables combinées associées aux cancers du sein en préménopause et en postménopause incluant un vaste nombre de facteurs de risque. Un total de 61 % de cas de cancer en préménopause diagnostiqués après 40 ans serait attribuable aux facteurs non-comportementaux et 40 % aux facteurs comportementaux. *A contrario*, après la ménopause, l'impact des facteurs comportementaux est plus élevé que celui des facteurs non-comportementaux (42 % contre 54 %, respectivement).

2.4.a. Confrontation à la littérature

À notre connaissance, il s'agit de la première étude estimant les proportions de cas attribuables aux facteurs anthropométriques autour de la puberté et au poids de naissance en différenciant les cancers du sein en préménopause et en postménopause. Une étude (261) a estimé les proportions de cas de cancer du sein attribuables à un IMC faible autour de la puberté à 15 % et au poids de naissance à 7 %. Les proportions de cas attribuables à la silhouette autour de la puberté dans notre étude sont en accord avec ces résultats en postménopause (cependant, en préménopause, la significativité statistique n'est pas atteinte). Une silhouette ronde autour de la puberté a été précédemment associée à une diminution du risque relatif de cancer du sein (60;262;263). Nos résultats concernant le poids de naissance sont en accord avec deux méta-analyses précédemment publiées concluant à une association positive, plus prononcée chez les femmes préménopausées que chez les femmes ménopausées (59;264).

En ce qui concerne les cancers du sein en préménopause, la proportion de cas attribuables à une puberté tardive, statistiquement significative, est en accord avec une estimation récemment publiée s'élevant à 22,2 % (132).

En ce qui concerne les cancers du sein en postménopause, la proportion de cas attribuables à l'utilisation de THM observée dans notre étude est comprise dans l'étendue des résultats publiés (entre 2,4 % à 19,4 %) (131-133;136;265-268). En effet, les résultats de notre étude suggèrent qu'une proportion statistiquement significative de cas de cancer du sein en postménopause serait attribuable à une récente utilisation de THM (14,5 % [9,2 % ; 19,6 %]), restreinte à l'utilisation d'un traitement œstro-progestatif (13,8 % [10,1 % ; 17,4 %], résultats non tabulés) (219). La proportion de cas attribuables à un IMC inférieur à 25 est comprise dans l'étendue des résultats précédemment publiés (entre 0,2 % à 24,8 %) (132-135;267-269), de même pour une consommation d'alcool élevée (entre -7,6 % à 9,1 %) (131-133;135;268), un âge à la ménopause tardif (entre 5,9 % à 14,5 %) (131-133;136) et un âge à la puberté précoce (entre 7,7 % à 18,8 %) (130-133). La proportion de cas attribuables à la présence d'antécédents familiaux de cancer du sein chez les parents au premier degré est plus faible que les estimations que l'on peut trouver dans la littérature (de 5,7 % à 15,7 %) (130-134;136) à cause d'une plus faible prévalence de l'exposition dans notre population d'analyse que dans les autres études publiées (jusqu'à 20,4 % dans la littérature contre 7,3 % dans notre population d'analyse) et de plus faibles risques relatifs (de 1,49 à 6,29 dans la littérature contre 1,37 et 1,50 chez les femmes préménopausées et ménopausées dans notre population d'analyse).

Quelques études ont quantifié l'influence conjointe de facteurs comportementaux et/ou non-comportementaux sur le risque de cancer du sein en préménopause (135;136), en postménopause (131;133-136) ou sans distinction sur le statut ménopausique (130;132;137;138). Parmi elles, très peu ont analysé l'impact conjoint de facteurs comportementaux uniquement (131;133;135) ou de facteurs non-comportementaux uniquement (131;133;136). Les études précédemment publiées ont estimé des proportions de cas de cancer du sein en préménopause attribuables à la combinaison de certains facteurs à hauteur de 42,8 %, en combinant une consommation élevée d'alcool et un faible niveau d'activité physique (135), et de 52,0 %, en combinant un faible niveau d'éducation, la nulliparité, un âge à la première grossesse à terme tardif et la présence d'antécédents familiaux de cancer du sein (136). Pour les cas de cancer du sein en postménopause, différentes combinaisons de facteurs de risque comportementaux ont été associées à une proportion de cas attribuables de 26,3 % (en combinant un faible niveau d'activité physique, l'utilisation de THM, un IMC élevé et une consommation élevée d'alcool) (133), 40,7 %

(en combinant un faible niveau d'activité physique, l'utilisation de THM, une consommation élevée d'alcool et un gain de poids depuis l'âge de 18 ans) (131) et de 12,4 % à 40,7 % en combinant deux facteurs parmi une consommation élevée d'alcool, un faible niveau d'activité physique et un IMC élevé (135). Dans notre étude, la proportion de cas attribuables à la combinaison des facteurs de risque comportementaux (42,3 %) est plus élevée que celles précédemment publiées. En ce qui concerne les facteurs de risque non-comportementaux, les différentes estimations précédemment publiées s'élèvent à 37,2 % (en combinant un âge tardif à la ménopause, la présence d'antécédents personnels de maladie bénigne du sein, un âge à la puberté précoce et la présence d'antécédents familiaux de cancer du sein chez les parents au premier degré) (133) et de 57,3 % (en ajoutant à ces facteurs un âge tardif à la première grossesse à terme et une haute taille adulte) (131). Dans notre étude, nous observons une proportion de cas attribuables aux facteurs de risque non-comportementaux de 51,7 %. Les différences trouvées pour les estimations des proportions de cas attribuables aux facteurs comportementaux et non-comportementaux pourraient s'expliquer par un plus grand nombre de facteurs de risque pris en considération dans nos analyses.

Nos résultats concernant la différenciation selon le type de cancer du sein sont en accord avec ceux de l'unique publication évaluant les proportions de cas attribuables différenciées selon le statut hormonal de la tumeur maligne (133). Dans cette étude, les proportions de cas attribuables à l'utilisation de THM, à un âge tardif à la ménopause et à un âge précoce à la puberté sont plus élevées pour les cancers RO+ que pour les cancers RO- (133). De plus, de précédentes études poolées ou méta-analyses ont conclu qu'un IMC élevé était associé à une diminution du risque de cancer du sein en préménopause RO+/RP+ en termes d'HR (270;271) et que la diminution de risque de cancer du sein en postménopause pour une augmentation de 5 points d'IMC était limitée aux cancers du sein en postménopause RO+/RP+ (271). Bien que non statistiquement significatif, une étude a observé qu'une puberté tardive serait associée à une diminution du risque de cancer du sein en préménopause RO- et à une augmentation du risque de cancer du sein en préménopause RO+ (272). Un âge tardif à la puberté a également été associé à une diminution de risque de cancer du sein RO+ (273-276). Un IMC élevé autour de la puberté a été associé à une diminution du risque de cancer en postménopause RO- (277), RO+/RP- (278) et RO+/RP+ (60). Enfin, l'utilisation de THM a été associée à une augmentation du risque de cancer en postménopause RO+/RP+ (274;279), RO+ (275;280) et HER2- (279;281).

2.4.b. *Limites de l'étude*

Les seuils choisis pour définir les niveaux d'exposition ont une influence sur l'estimation des fractions attribuables (260). Afin de limiter les biais, les catégorisations effectuées ont été déterminées *a priori* en se référant aux résultats de la littérature et aux recommandations de santé publique.

L'âge à la puberté et la silhouette à la puberté ont été rétrospectivement collectés auprès de femmes âgées de 40 à 65 ans, ce qui peut générer des erreurs de mesures. Cependant, une étude de validation indépendante a montré que l'information était fiable (les corrélations obtenues entre les véritables données et celles déclarées sont de 0,8 pour l'âge à la puberté et de 0,6 pour la silhouette à la puberté) (282). La nécessité de regrouper les cinq catégories les plus élevées de silhouettes à la puberté indique un manque de puissance statistique pour les femmes qui étaient les plus corpulentes à cette période. Il serait intéressant de réitérer les analyses concernant ce facteur dans une population présentant une plus grande variabilité de silhouettes à la puberté.

Comme toute étude de cohorte, E3N n'est pas représentative de la population générale française et est encline à un effet de « cohorte saine ». Cependant, estimer les fractions attribuables dans la population générale n'est pas faisable en raison d'un manque de données de prévalence en France.

Enfin, les résultats concernant le cancer du sein en préménopause devraient être interprétés avec précaution à cause du nombre limité de cas dans E3N, en raison de la sélection initiale de femmes de plus de 40 ans ; de même pour les résultats différenciant les cancers du sein en postménopause selon l'expression de la protéine HER2.

2.4.c. *Conclusion*

Des modifications comportementales pourraient avoir un impact important sur le risque de cancer du sein après 40 ans ; et de manière encore plus prononcée après la ménopause. Les agences de santé publique devraient se focaliser sur les facteurs comportementaux ayant le plus d'impact sur la population, à savoir la consommation d'alcool, l'alimentation, l'obésité après la ménopause et l'utilisation de THM, et encourager des études de recherche sociologique permettant de comprendre comment modifier les habitudes personnelles des populations.

Partie 3 Prédiction à cinq ans du risque de cancer du sein

3.1. Introduction

Le cancer du sein demeure un problème majeur de santé publique mondial, atteignant une incidence de 43,3 pour 100 000 femmes-années dans le monde en 2012 (2). Malgré une diminution constante de la mortalité par cancer du sein dans plusieurs pays (section Introduction, partie 2.3.b page 54), l'incidence globale continue à augmenter, avec des variations différentes selon les pays (section Introduction, partie 1.2.b page 35) (2;283). Le modèle de prédiction est un outil statistique permettant d'estimer la probabilité qu'un individu sain, ayant un certain profil de caractéristiques, développe l'événement d'intérêt dans une période de temps définie. Les prédictions de risque permettent d'estimer le risque absolu de la survenue d'une maladie et sont à l'origine de recommandations d'interventions préventives et de thérapies personnalisées pour les professionnels de santé et donnent la possibilité aux patients de se situer par rapport au risque moyen (284) (section Introduction, partie 3.3 page 63). Le score prédictif de la survenue d'un cancer du sein le plus utilisé dans la littérature est le modèle de Gail (section Méthodes statistiques, partie 2.11.e page 128) (146).

L'objectif de cette partie est d'identifier les facteurs prédictifs du risque à cinq ans de cancer du sein, selon le statut ménopausique des femmes, en utilisant un modèle de régression de Cox. Les modèles retenus seront comparés à deux autres méthodes de construction d'un score prédictif : une méthode de fouille de données, appelée méthode des plus proches voisins, et le modèle de Gail établi à partir d'un modèle de régression logistique.

3.2. Matériel et méthodes

3.2.a. Population d'étude

Dans cette partie, les femmes sont suivies de la date de réponse au questionnaire alimentaire Q3 (1993) jusqu'à la date de diagnostic de n'importe quel cancer autre que le cancer cutané basal, ou jusqu'à cinq ans après le début de l'étude. Les femmes non-répondantes au questionnaire alimentaire (N = 24 473) et celles dont un cancer, autre qu'un cancer cutané basal, a été diagnostiqué avant le début de suivi (N = 4 705) ont été exclues des analyses, de même que les femmes sans suivi après Q3 (N = 769) et les cas de cancer sans date de diagnostic (N = 54). Par ailleurs, les femmes avec une durée de suivi inférieure à 5 ans (N = 1 505), les femmes diagnostiquées avec un cancer *in situ* (N = 270) ou un cancer invasif, autre qu'un cancer cutané

basal (N = 864) dans les cinq ans suivant le début de l'étude ont été exclues des analyses. La population d'analyse comprend ainsi 64 995 femmes, âgées de 42 à 72 ans au début du suivi.

3.2.b. Collecte des données

La définition du statut ménopausique et la collecte des données nécessaires aux analyses ont été décrites dans la section Matériel, partie 4 page 94. Les variables testées pour être incluses dans la construction des scores de risque sont les antécédents familiaux de cancer du sein chez les parents au premier degré, la taille à l'âge adulte, les antécédents de maladie bénigne du sein, l'âge à la puberté, l'âge à la ménopause, le nombre d'enfants, l'âge à la première grossesse à terme, la consommation d'alcool, l'IMC et l'utilisation récente de THM. Étant donné que, pour chaque variable, les valeurs manquantes représentent moins de 5 % de la population d'analyse, celles-ci ont été imputées à la médiane pour les variables quantitatives et au mode pour les variables qualitatives. Les variables ont été incluses dans les modèles de régression de Cox en continu, excepté l'âge à la première grossesse qui a été catégorisée (≤ 20 ,]20 ; 25],]25 ; 30],]30 ; 35], > 30 et nullipare) en raison de la présence d'une référence commune avec la variable renseignant le nombre d'enfants.

3.2.c. Méthodologie statistique

Des analyses séparées selon le statut ménopausique des femmes ont été réalisées. La population d'analyse a été aléatoirement divisée (sans remise) en deux échantillons : un échantillon d'apprentissage (N = 51 821) sur lequel les modèles de prédiction sont testés pour identifier les variables ayant le plus fort pouvoir prédictif par la méthode de validation croisée et un échantillon de validation (N = 13 174) sur lequel les modèles de prédiction sélectionnés sont validés. Les performances des modèles sont estimées par le calcul de la discrimination et de la calibration (section Méthodes statistiques, partie 2.11.b page 120).

La formule de l'estimation du risque de survenue de cancer du sein à cinq ans est présentée dans l'équation 21 page 119. L'estimation du risque est réalisée à l'aide de modèles de régression semi-paramétrique de Cox, en utilisant le temps comme échelle de temps. La fonction de survie de base est modélisée par des fonctions splines (section Méthodes statistiques, partie 2.5 page 114).

3.3. Résultats

3.3.a. Caractéristiques de la population d'analyse

- **Caractéristiques de l'échantillon d'apprentissage**

Les 51 821 femmes de la population d'analyse ont été suivies pendant un total de 256 843 femmes-années (avec un suivi médian de 2,6 ans pour les cas de cancer du sein invasif et de 5,0 pour les non-cas). Parmi ces femmes, 930 ont été diagnostiquées avec un cancer du sein, 393 n'étaient pas ménopausées au moment du diagnostic et 537 l'étaient.

Les caractéristiques au début de l'étude des femmes de la population d'analyse, incluses dans l'échantillon d'apprentissage, selon leur statut cas/non-cas sont présentées dans le tableau 43 (en annexe page 225).

- **Caractéristiques de l'échantillon de validation**

Les 13 174 femmes de la population d'analyse ont été suivies pendant un total de 65 474 femmes-années (avec un suivi médian de 2,9 ans pour les cas de cancer du sein invasif et de 5,0 pour les non-cas). Parmi ces femmes, 176 ont été diagnostiquées avec un cancer du sein : 78 n'étaient pas ménopausées au moment du diagnostic et 98 l'étaient.

Les caractéristiques au début de l'étude des femmes de la population d'analyse incluses dans l'échantillon de validation, selon leur statut cas/non-cas sont présentées dans le tableau 44 (en annexe page 227).

- **Comparaison des échantillons d'apprentissage et de validation**

Les différences entre les moyennes observées dans les échantillons d'apprentissage et de validation ont été estimées en utilisant la méthode décrite dans la section Méthodes statistiques, partie 1.2 page 107 et sont présentées dans le tableau 24.

Tableau 24 : Différences de moyennes et de pourcentages, pour les variables analysées, observées entre les échantillons d'apprentissage et de validation, cohorte E3N (N = 64 995).

Variable	P-valeur	Moyenne	
		Échantillon d'apprentissage N = 51 821	Échantillon de validation N = 28 459
Âge à la première grossesse à terme	0,5579	24,9033	24,9274
Âge	0,0660	52,6864	52,8054
Consommation d'alcool	0,5801	11,5780	11,5032
Antécédents personnels de maladie bénigne du sein	0,7602	0,2930	0,2916
IMC	0,9356	22,9370	22,9396
Antécédents familiaux de cancer du sein chez les parents au premier degré	0,3463	0,1266	0,1300
Âge à la ménopause	0,7522	49,4715	49,4539
Nombre d'enfants	0,0129	2,0187	1,9910
Âge à la puberté	0,7580	12,8134	12,8092
Taille à l'âge adulte	0,9022	161,6	161,6
Variable ^a	P-valeur	Pourcentage	
		Échantillon d'apprentissage N = 51 821	Échantillon de validation N = 28 459
Pourcentage de cas de cancers du sein invasifs	0,0003	0,0179	0,0134
Âge ∈ [45 ; 50[0,0209	0,3200	0,3095
Au moins deux cancers du sein chez les parents au premier degré	0,0209	0,0092	0,0114
Au moins trois enfants	0,0145	0,2942	0,2834

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Seules les modalités présentant une différence statistiquement significative entre les pourcentages observés dans les échantillons d'apprentissage et de validation sont présentées dans le tableau.

Une différence faible, mais statistiquement significative, entre les deux échantillons est observée concernant la moyenne du nombre d'enfants (2,02 enfants dans l'échantillon d'apprentissage contre 1,99 enfant dans l'échantillon de validation). De plus, des différences statistiquement significatives ont été observées entre les deux échantillons concernant la proportion de femmes diagnostiquées avec un cancer du sein (respectivement 1,79 % et 1,34 %), âgées entre 45 et 50 ans (respectivement 32,00 % et 30,95 %), ayant au moins deux antécédents familiaux de cancer du sein chez les parents au premier degré (respectivement 0,92 % et 1,14 %) et ayant au moins trois enfants (respectivement 29,42 % et 28,34 %).

3.3.b. Sélection des modèles de prédiction du risque de cancer du sein en préménopause et en postménopause

Les performances des modèles à une variable, obtenues par la méthode de validation croisée au sein de l'échantillon d'apprentissage, sont présentées dans le tableau 25.

Tableau 25 : Performances des modèles à un seul prédicteur pour la prédiction du risque de cancer du sein invasif en préménopause et en postménopause, cohorte E3N (N = 51 821).

Variables	AUC	E/O
Préménopause (N = 23 362)		
Âge	0,5777	1,0022
Antécédents personnels de maladie bénigne du sein	0,5541	0,9978
Âge à la puberté	0,5326	0,9740
Antécédents familiaux de cancer du sein chez les parents au premier degré	0,5313	0,9825
IMC	0,5246	0,9746
Nombre d'enfants	0,5225	0,9757
Consommation d'alcool	0,5212	0,9741
Âge à la première grossesse à terme	0,5196	0,9868
Taille à l'âge adulte	0,5186	0,9736
Postménopause (N = 28 459)		
Antécédents personnels de maladie bénigne du sein	0,5518	0,9890
Âge à la puberté	0,5365	0,9677
Nombre d'enfants	0,5365	0,9709
Âge à la ménopause	0,5351	0,9794
Taille à l'âge adulte	0,5347	0,9696
Antécédents familiaux de cancer du sein chez les parents au premier degré	0,5323	0,9796
Consommation d'alcool	0,5290	0,9708
Utilisation récente de THM	0,5254	1,0064
Âge	0,5242	0,9701
IMC	0,5209	0,9688
Âge à la première grossesse à terme	0,4980	0,9717

Les meilleurs prédicteurs en univarié pour la prédiction du risque de cancer du sein invasif sont l'âge pour le cancer du sein en préménopause et les antécédents de maladie bénigne du sein en postménopause.

Plusieurs combinaisons de variables, parmi celles qui maximisent le mieux l'AUC et le coefficient de régression du PI (section Méthodes statistiques, partie 2.11.b page 120), sont présentées dans le tableau 26 (la sélection des variables pour les modèles prédictifs en préménopause et postménopause est décrite dans la section Méthodes statistiques, partie 2.11.c page 125).

Tableau 26 : Meilleures combinaisons optimisant l'AUC et le coefficient de régression associé au PI, échantillon d'apprentissage, cohorte E3N (N = 51 821).

Meilleures combinaisons de prédicteurs	Nombre de variables	AUC	Coefficient de régression associé au PI	Rapport E/O
Préménopause (N = 23 362)				
âge, MBS, famille (avec le terme d'interaction entre MBS et famille)	3	0,603	0,686	1,04
âge, MBS, famille, nbenf (avec les termes d'interaction entre famille et nbenf et entre famille et MBS)	4	0,603	0,664	1,04
âge, MBS, famille, IMC (avec les termes d'interaction entre âge et IMC)	4	0,603	0,582	1,04
âge, MBS, famille, IMC (avec les termes d'interaction entre famille et MBS)	4	0,603	0,683	1,04
âge, MBS, famille, nbenf (avec les termes d'interaction entre MBS et nbenf et entre famille et MBS)	4	0,603	0,646	1,04
âge, MBS, famille, nbenf, taille (avec les termes d'interaction entre famille et nbenf et entre famille et MBS)	5	0,603	0,663	1,04
Postménopause (N = 28,459)				
MBS, agemeno, famille, nbenf, THM, âge (avec les termes d'interaction entre agemeno et THM et entre MBS et THM)	6	0,602	0,544	1,04
MBS, agemeno, famille, nbenf, THM, âge, alcool (avec les termes d'interaction entre agemeno et MBS et entre MBS et THM)	7	0,601	0,555	1,05
MBS, agemeno, famille, nbenf, THM, âge, alcool, IMC (avec les termes d'interaction entre MBS et THM, et entre agemeno et MBS)	8	0,601	0,555	1,05
MBS, agemeno, nbenf, THM, âge, alcool, IMC, taille (avec les termes d'interaction entre agemeno et MBS, entre agemeno et taille, entre âge et agemeno, entre MBS et THM et entre agemeno et alcool)	8	0,600	0,578	1,05
MBS, agemeno, famille, nbenf, THM, âge, alcool, taille (avec les termes d'interaction entre âge et agemeno, et entre agemeno et MBS)	8	0,601	0,560	1,05
MBS, agemeno, famille, nbenf, THM, âge, alcool, taille (avec les termes d'interaction entre agemeno et MBS, entre agemeno et taille et entre âge et agemeno)	8	0,602	0,555	1,05
MBS, agemeno, famille, nbenf, THM, âge, alcool, IMC (avec les termes d'interaction entre âge et agemeno et entre agemeno et MBS)	8	0,602	0,561	1,05

Note : Les modèles en gras correspondent aux modèles retenus en se basant sur la maximisation des estimateurs de la discrimination (AUC et coefficient de régression associé au PI).

Abréviations : MBS = antécédents personnels de maladie bénigne du sein, agemeno = âge à la ménopause, famille = antécédents familiaux de cancer du sein chez les parents au premier degré, alcool = consommation d'alcool, nbenf = nombre d'enfants, IMC = indice de masse corporelle, THM = utilisation récente de traitement hormonaux de la ménopause, âge = âge au début du suivi, taille = taille à l'âge adulte.

Le modèle prédictif retenu pour prédire le risque de cancer du sein invasif en préménopause inclut trois variables : l'âge, les antécédents familiaux de cancer du sein chez les parents au premier degré et les antécédents personnels de maladie bénigne du sein, en complément du terme

d'interaction entre les antécédents familiaux de cancer du sein chez les parents au premier degré et les antécédents personnels de maladie bénigne du sein. Les performances obtenues au sein de l'échantillon d'apprentissage grâce à la méthode de validation croisée sont une discrimination modeste avec un AUC de 0,603 et un coefficient de régression associé au PI de 0,686 et une calibration correcte avec un rapport E/O de 1,04.

Le modèle prédictif retenu pour prédire le risque de cancer du sein invasif en postménopause inclut huit variables : les antécédents personnels de maladie bénigne du sein, l'âge à la ménopause, les antécédents familiaux de cancer du sein chez les parents au premier degré, la consommation d'alcool, l'utilisation récente de THM, l'âge, le nombre d'enfants et l'IMC, en complément des termes d'interaction entre l'âge à la ménopause et l'âge, et entre l'âge à la ménopause et les antécédents personnels de maladie bénigne du sein. Les performances obtenues au sein de l'échantillon d'apprentissage grâce à la méthode de validation croisée sont une discrimination modeste avec un AUC de 0,602 et un coefficient de régression associé au PI de 0,561 et une calibration correcte avec un rapport E/O de 1,05.

3.3.c. Validation des meilleurs modèles de prédiction du risque de cancer du sein en préménopause et en postménopause

- **Estimation des performances des modèles retenus dans l'échantillon de validation**

Les performances des modèles retenus pour la prédiction du risque de cancer du sein en préménopause et en postménopause, construits à l'aide de l'ensemble des observations de l'échantillon d'apprentissage, et testés chez les femmes préménopausées et postménopausées dans l'échantillon de validation sont présentées dans le tableau 27.

Tableau 27 : Performances des modèles sélectionnés, échantillon de validation, cohorte E3N (N = 13 174).

	Score appliqué aux femmes préménopausées N = 5 765			Score appliqué aux femmes postménopausées N = 7 233		
	AUC [IC 95 %]	Coefficients de régression sur le PI	Rapport E/O	AUC [IC 95 %]	Coefficients de régression sur le PI	Rapport E/O
Préménopause^a	0,565 [0,500 ; 0,631]	0,422	1,28	0,544 [0,487 ; 0,602]	0,235	1,25
Postménopause^b	0,581 [0,514 ; 0,649]	0,008	5,67	0,617 [0,561 ; 0,673]	0,592	1,40

a : Modèle retenu en préménopause : âge, MBS, famille (avec le terme d'interaction entre MBS et famille).

b : Modèle retenu en postménopause : MBS, agemeno, famille, nbenf, THM, âge, alcool, IMC (avec les termes d'interaction entre âge et agemeno et entre agemeno et MBS).

Lorsqu'appliqués à la population ciblée (*i.e.* le score en préménopause appliqué aux femmes préménopausées et le score en postménopause appliqué aux femmes postménopausées), la discrimination des modèles retenus en préménopause et en postménopause est modeste si l'on considère l'estimation de l'AUC (respectivement 0,565 et 0,617) et limitée si on s'intéresse au coefficient de régression associé au PI (respectivement 0,422 et 0,592). L'estimation du rapport E/O permet de se rendre compte que les deux modèles ont tendance à surestimer le nombre de cas de cancer du sein diagnostiqués (respectivement 1,28 et 1,40).

Lorsqu'appliqué à la population non ciblée (*i.e.* le score en préménopause appliqué aux femmes postménopausées et le score en postménopause appliqué aux femmes préménopausées), les modèles retenus sont moins performants que lorsqu'ils sont appliqués à la population cible : les valeurs de l'AUC sont respectivement de 0,544 et 0,581, celles du coefficient de régression sur le PI de 0,235 et de 0,008 et le rapport E/O de 1,25 et 5,67.

• **Description par quartiles de score prédit**

Les associations entre les scores prédictifs, catégorisés selon les quartiles de distribution dans les échantillons d'apprentissage et de validation, et le risque de cancer du sein invasif sont présentées dans le tableau 28, de même que les estimations des rapports E/O par quartiles.

Tableau 28 : Associations entre les scores prédictifs préménopausique et postménopausique, catégorisés selon les quartiles de distribution, et le risque de cancer du sein invasif, cohorte E3N (N = 64 995).

		Premier quartile	Deuxième quartile	Troisième quartile	Quatrième quartile
Préménopause^a					
Échantillon d'apprentissage (N = 51 821)	Intervalle	[0 ; 0,0117]	[0,0117 ; 0,0157]	[0,0157 ; 0,0204]	[0,0204 ; 1,0000]
	HR [IC 95 %]	1,00	1,70 [1,21 ; 2,39]	1,97 [1,41 ; 2,75]	2,93 [2,14 ; 4,01]
	E/O	1,17	0,88	1,02	1,00
Échantillon de validation (N = 13 174)	Intervalle	[0 ; 0,0118]	[0,0118 ; 0,0158]	[0,0158 ; 0,0206]	[0,0206 ; 1,0000]
	HR [IC 95 %]	1,00	0,94 [0,46 ; 1,89]	1,13 [0,57 ; 2,21]	1,82 [0,99 ; 3,35]
	E/O	0,96	1,31	1,46	1,33
Postménopause^b					
Échantillon d'apprentissage (N = 51 821)	Intervalle	[0 ; 0,0137]	[0,0137 ; 0,0174]	[0,0174 ; 0,0225]	[0,0225 ; 1,0000]
	HR [IC 95 %]	1,00	1,03 [0,77 ; 1,38]	1,61 [1,24 ; 2,09]	2,28 [1,78 ; 2,92]
	E/O	0,83	1,17	0,96	1,00
Échantillon de validation (N = 13 174)	Intervalle	[0 ; 0,0134]	[0,0134 ; 0,0173]	[0,0173 ; 0,0226]	[0,0226 ; 1,0000]
	HR [IC 95 %]	1,00	1,13 [0,57 ; 2,27]	2,01 [1,08 ; 3,73]	2,41 [1,32 ; 4,04]
	E/O	1,28	1,66	1,20	1,49

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Modèle retenu en préménopause : âge, famille, MBS (avec le terme d'interaction entre MBS et famille).

b : Modèle retenu en postménopause : MBS, agemeno, famille, nbenf, THM, âge, alcool, IMC (avec les termes d'interaction entre âge et agemeno et entre agemeno et MBS).

Pour le score préménopausique, le dernier quartile est associé à une augmentation statistiquement significative du risque de cancer du sein dans l'échantillon d'apprentissage (HR = 2,93 [2,14 ; 4,01]), alors que cette association est à la limite de la significativité statistique dans l'échantillon de validation (HR = 1,82 [0,99 ; 3,35]). Pour le score postménopausique, le dernier quartile est associé à une augmentation statistiquement significative du risque de cancer du sein dans les deux échantillons (respectivement, HR = 2,28 [1,78 ; 2,92] et HR = 2,41 [1,32 ; 4,04]).

L'estimation de la calibration des scores prédictifs, *via* le rapport E/O, selon les quartiles de distribution des scores prédictifs préménopausique et postménopausique permet d'observer que les scores prédictifs sont mieux calibrés dans l'échantillon d'apprentissage (estimations obtenues par la méthode de validation-croisée) que dans l'échantillon de validation. Dans l'échantillon d'apprentissage, les scores possèdent une meilleure calibration pour les femmes à haut risque de cancer du sein que pour les femmes à plus faible risque. Dans l'échantillon de validation, les deux scores surestiment le nombre de cas de cancer du sein dans chaque catégorie (sauf pour les femmes ayant les plus faibles scores en préménopause).

Les diagrammes de calibration, représentant la proportion de cas observés en fonction du score prédit moyen dans chaque classe de score, catégorisé selon les quartiles de distribution, sont présentés dans les figures 39 et 40.

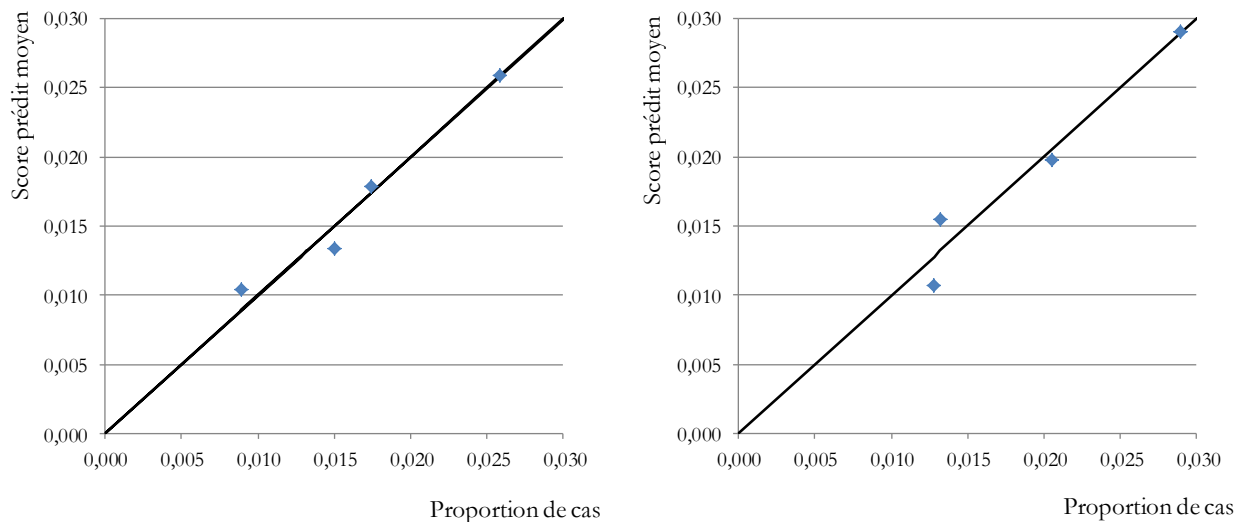


Figure 39 : Diagrammes de calibration des scores préménopausique et postménopausique dans l'échantillon d'apprentissage, cohorte E3N (N = 51 821).

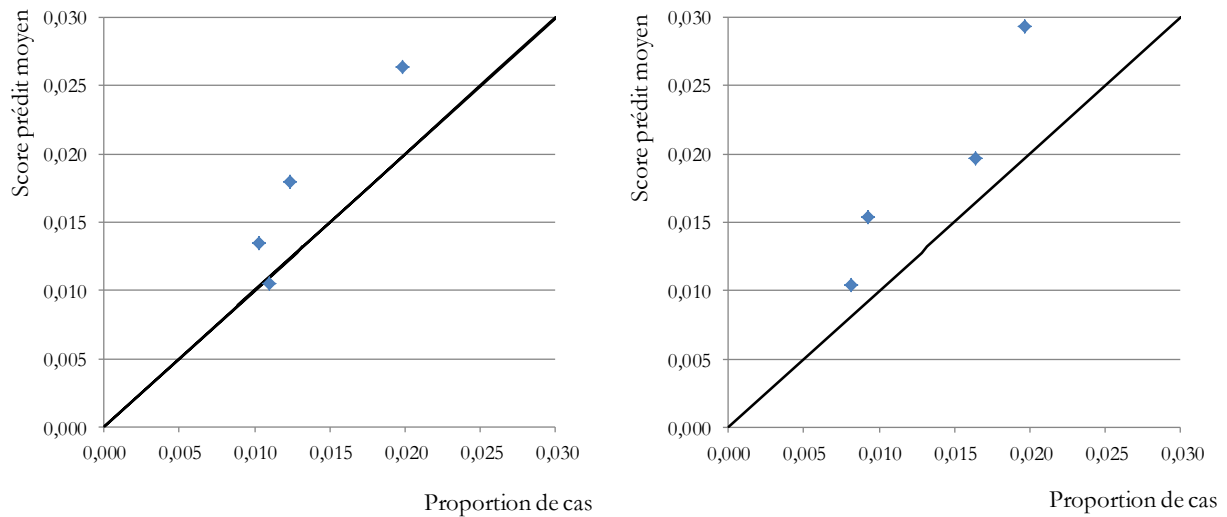


Figure 40 : Diagrammes de calibration des scores préménopausique et postménopausique dans l'échantillon de validation, cohorte E3N (N = 13 174).

Dans l'échantillon d'apprentissage, les deux scores prédictifs apparaissent bien calibrés, alors que dans l'échantillon de validation, les diagrammes confirment que les scores surestiment le nombre de cas de cancer du sein, que ce soit en préménopause ou en postménopause.

- **Validation de la modélisation de la fonction de survie de base**

Les équations obtenues pour la modélisation de la fonction de survie de base sont présentées dans l'note : dans les équations 33 et 34, « suivi » désigne la durée de suivi des individus, centrée et réduite à partir de la moyenne et de la variance observées dans l'échantillon d'apprentissage.

équation 33 pour le score en préménopause et dans l'équation 34 pour le score en postménopause.

Note : Dans les équations 33 et 34, « suivi » désigne la durée de suivi des individus, centrée et réduite à partir de la moyenne et de la variance observées dans l'échantillon d'apprentissage.

Équation 33 : Modélisation de la fonction de survie de base pour le score préménopausique.

$$\begin{aligned} \ln(H(t)) &= 24,81 + 80,21 \cdot \text{suivi} + 70,63 \cdot \text{suivi}^2 + 20,92 \cdot \text{suivi}^3 - 23,94 \cdot (\text{suivi} + 1,09)^3 \\ &\quad + 4,17 \cdot (\text{suivi} + 0,63)^3 - 1,51 \cdot (\text{suivi} + 0,01)^3 + 0,62 \cdot (\text{suivi} - 0,62)^3 \\ &\quad - 0,76 \cdot (\text{suivi} - 1,09)^3 + 24,55 \cdot (\text{suivi} - 1,58)^3 \\ S(t) &= \exp\left(-\exp\left(\ln(H(t))\right)\right) \end{aligned}$$

Équation 34 : Modélisation de la fonction de survie de base pour le score postménopausique.

$$\begin{aligned} \ln(H(t)) &= 14,49 + 51,95.\text{suivi} + 45,19.\text{suivi}^2 + 13,36.\text{suivi}^3 - 14,24.(\text{suivi} + 1,11)^3 \\ &\quad + 1,39.(\text{suivi} + 0,55)^3 - 0,65.(\text{suivi} + 0,01)^3 + 0,36.(\text{suivi} - 0,58)^3 \\ &\quad - 0,45.(\text{suivi} - 1,10)^3 - 10,73.(\text{suivi} - 1,56)^3 \\ S(t) &= \exp\left(-\exp\left(\ln(H(t))\right)\right) \end{aligned}$$

La validation de la modélisation de la fonction de survie de base au sein de l'échantillon d'apprentissage (section Méthodes statistiques, partie 2.11.b page 120) est présentée dans la figure 41 pour le score en préménopause et dans la figure 42 pour le score en postménopause.

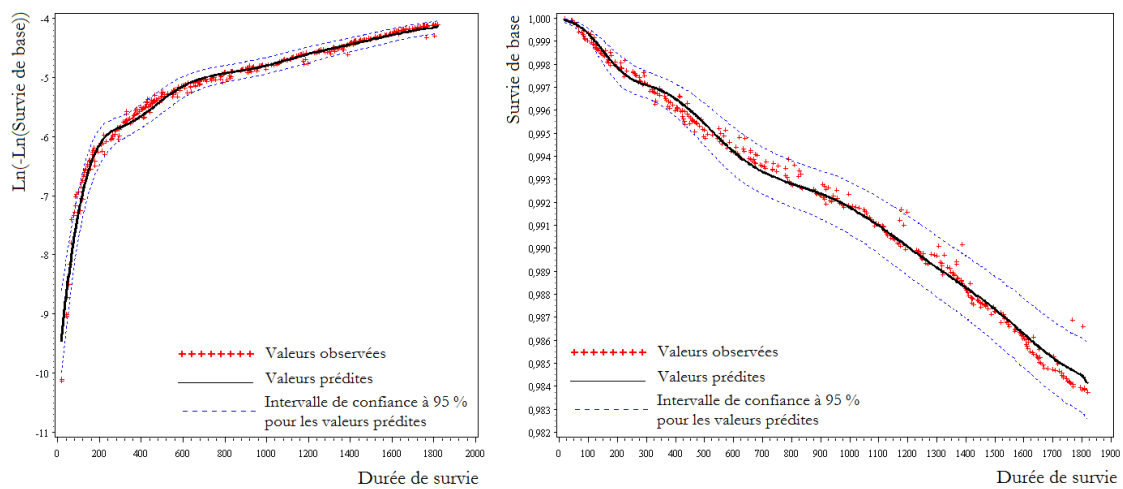


Figure 41 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en préménopause dans l'échantillon d'apprentissage, cohorte E3N (N = 23 362).

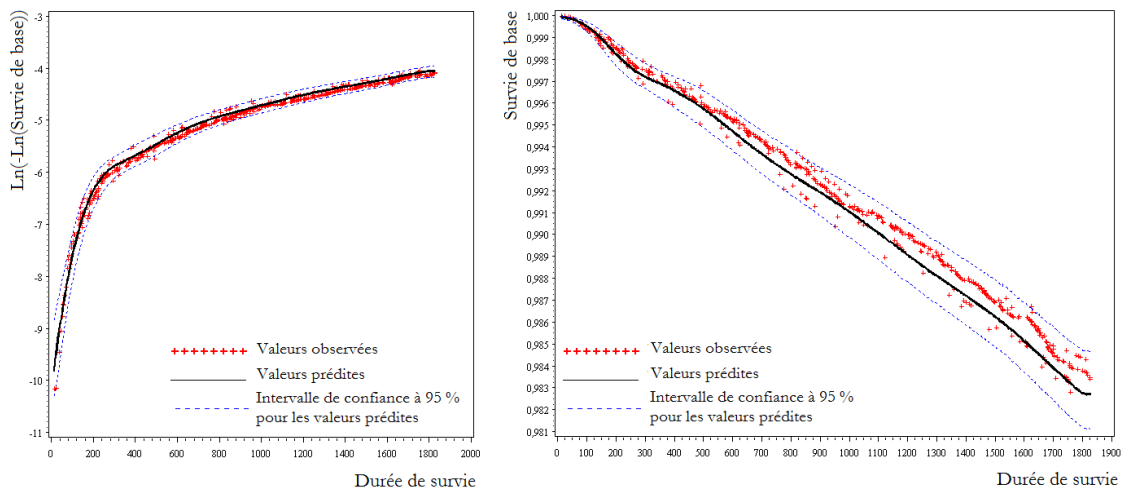


Figure 42 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en postménopause dans l'échantillon d'apprentissage, cohorte E3N (N = 28 459).

Que ce soit pour le score en préménopause ou le score en postménopause, les valeurs de survie observées dans l'échantillon d'apprentissage sont comprises à l'intérieur de l'intervalle de

confiance à 95 % des valeurs prédites dans l'échantillon d'apprentissage par la modélisation adoptée, ce qui permet de conclure à une prédiction adéquate de la fonction de survie de base dans cet échantillon.

La validation de la modélisation de la fonction de survie de base au sein de l'échantillon de validation (section Méthodes statistiques, partie 2.11.b page 120) est présentée dans la figure 43 pour le score en préménopause et dans la figure 44 pour le score en postménopause.

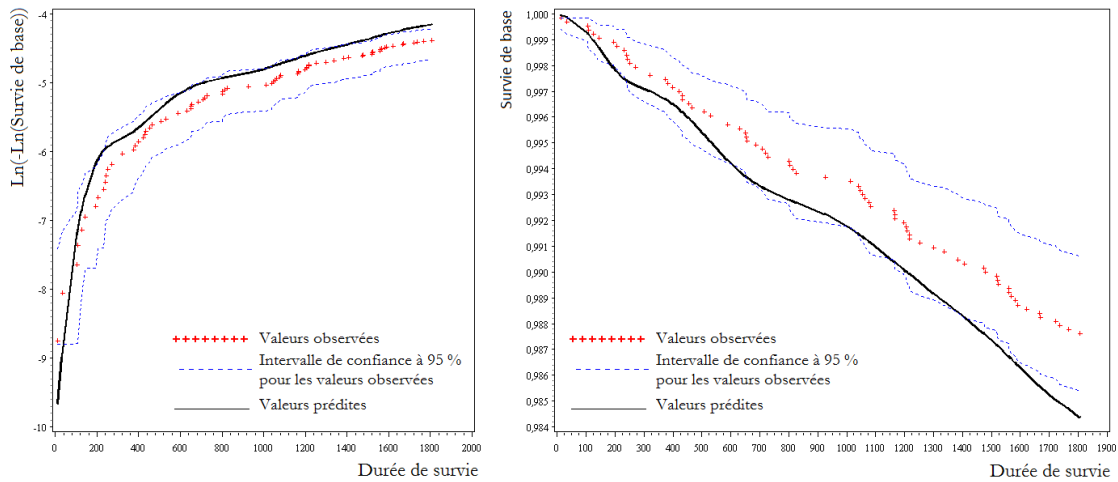


Figure 43 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en préménopause dans l'échantillon de validation, cohorte E3N (N = 5 843).

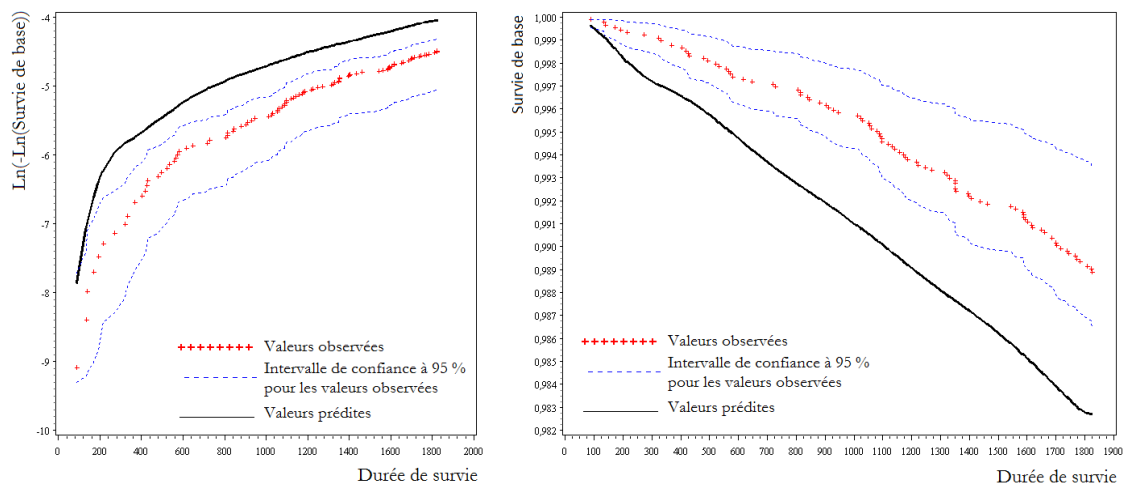


Figure 44 : Validation de la modélisation des fonctions $\ln(-\ln(S(t)))$ et $S(t)$ de base pour le score en postménopause dans l'échantillon de validation, cohorte E3N (N = 7 331).

Lorsque la modélisation de la fonction de survie de base (présentée dans les équations 33 et 34 et déterminée à partir des données de l'échantillon d'apprentissage) est injectée dans l'échantillon de validation, les valeurs prédites sont à la limite d'une des deux bornes de l'intervalle de confiance à 95 % des valeurs observées pour le score en préménopause et à l'extérieur pour le score en postménopause. Ainsi, la modélisation a tendance à sous-estimer les valeurs de la survie

de base, que ce soit en préménopause ou en postménopause, ce qui a pour conséquence une surestimation du risque de cancer.

3.3.d. Comparaison avec d'autres modèles de prédiction

Afin d'évaluer comment se situent les performances des scores construits par rapport à d'autres méthodologies de construction de scores prédictifs, celles-ci ont été comparées aux performances obtenues pour des scores construits à l'aide d'une méthode exploratoire appelée « Algorithme des plus proches voisins » (231;232) et au score de prédiction de cancer du sein le plus utilisé dans le monde, le modèle de Gail (section Méthodes statistiques, partie 2.11.e page 128).

Le tableau 29 présente les variables prises en compte dans la construction des différents scores prédictifs du risque de cancer du sein par les trois méthodes comparées.

Tableau 29 : Sélection des prédicteurs entrant dans la construction du score de prédiction du risque de cancer du sein pour le modèle de Cox, l'algorithme des plus proches voisins et le modèle logistique de Gail.

	Modèle de Cox		Algorithme des plus proches voisins		Modèle de Gail ¹
	Préménopause	Postménopause	Préménopause	Postménopause	
Âge	X	X	X	X	X
Antécédents familiaux de cancer du sein chez les parents au premier degré	X	X	X	X	X
Antécédents personnels de maladie bénigne du sein	X	X	X	X	
Antécédents de biopsie mammaire					X
Âge à la puberté			X		X
Nombre d'enfants		X			
Âge à la première grossesse à terme					X
Consommation d'alcool		X			
IMC		X			
Âge à la ménopause		X		X	
Utilisation récente de THM		X		X	

1 : Le modèle de Gail ne différencie pas le risque de cancer du sein selon le statut ménopausique des femmes.

Nous pouvons remarquer que l'âge et les antécédents familiaux de cancer du sein chez les parents au premier degré sont sélectionnés dans tous les modèles. Les antécédents personnels de maladie bénigne du sein sont choisis dans les modèles de Cox et des plus proches voisins, alors que le modèle de Gail a sélectionné, à la place, le nombre de biopsies mammaires pratiquées. Les deux modèles en postménopause construits par le modèle de Cox et des plus proches voisins incluent un nombre plus important de variables que les deux modèles en préménopause et que le modèle de Gail.

Les prédictions à cinq ans du risque pour 10 femmes hypothétiques avec différents scénarios de facteurs de risque sont présentées dans le tableau 30.

Tableau 30 : Prédiction à cinq ans du risque de cancer du sein invasif, à partir du modèle de Cox, de l'algorithme des plus proches voisins et du modèle logistique de Gail, pour dix femmes hypothétiques.

	Prédiction du risque de cancer du sein invasif à 5 ans (%)		
	Modèle de régression de Cox	Méthode des plus proches voisins	Modèle Logistique de Gail
Statut ménopausique			
Non ménopausée	15	45	0
Non ménopausée	11	50	1
Non ménopausée	9	55	2
Âge à la puberté (années)			
Âge au début du suivi (années)			
Antécédents familiaux de cancer du sein chez les parents au premier degré			
Antécédents personnels de maladie bénigne du sein			
Âge à la ménopause (années)			
Consommation d'alcool (g/jour)			
IMC (kg.m ⁻²)			
Nombre d'enfants			
Utilisation récente de THM			
Non ménopausée	1,06	1,23	0,54
Non ménopausée	2,75	2,65	2,29
Non ménopausée	4,19	3,80	7,78
Ménopausée	0,49	1,01	0,54
Ménopausée	0,73	1,25	0,63
Ménopausée	2,18	1,88	1,08
Ménopausée	3,18	3,45	2,93
Ménopausée	5,30	3,60	3,38
Ménopausée	7,75	4,18	10,78
Ménopausée	10,42	4,18	11,20

a : Les autres types de progestatif incluent l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, le diénoest, la drospirinone, le diacétate d'éthinodiol, le gestodène, le lévonorgestrel, le lynestrénol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de mégestrol, l'acétate de nomégestrol, l'acétate de noréthistérone ou la promégestone.

b : Les autres THM incluent la tibolone, les THM de type non spécifié, les traitements injectables et les THM associant un œstrogène et un androgène.

Nous pouvons observer que les trois scores évoluent dans le même sens. La prédiction à cinq ans du risque pour une femme en préménopause, âgée de 45 à 55 ans, est comprise entre 0,5 % et 7,8 % ; pour une femme ménopausée âgée de 45 à 70 ans, celui-ci est compris entre 0,5 % et 11,2 %. Les valeurs de prédiction du risque les plus extrêmes sont fournies, que ce soit en préménopause ou en postménopause, par le modèle de Gail. Le modèle des plus proches voisins fournit des estimations du risque moins étendues que les deux autres méthodes.

Afin de comparer statistiquement les scores obtenus par les différentes méthodes, les coefficients de corrélation entre chaque paire de score et la concordance des scores, catégorisés en quartiles, ont été estimés et sont présentés dans le tableau 31 ; de même que les performances de chaque modèle évaluées dans l'échantillon de validation, présentées dans le tableau 32.

Tableau 31 : Comparaisons des scores de prédiction à cinq ans du risque de cancer du sein invasif, construits à partir du modèle de Cox, de l'algorithme des plus proches voisins et du modèle logistique de Gail, échantillon de validation, cohorte E3N (N = 13 174).

Coefficient de corrélation de Pearson Concordance	Modèle de Cox	Méthode des plus proches voisins	Modèle logistique de Gail
Score préménopausique			
Modèle de Cox	1	0,7716 ^a 0,4937	0,5363 ^a 0,1564
Méthode des plus proches voisins	-	1	0,5200 ^a , 0,2053
Modèle logistique de Gail	-	-	1
Score postménopausique			
Modèle de Cox	1	0,6922 ^a 0,3204	0,4976 ^a 0,1419
Méthode des plus proches voisins	-	1	0,4126 ^a 0,1183
Modèle logistique de Gail	-	-	1

a : p-valeur < 0,001

Tableau 32 : Performances des scores de prédiction à cinq ans du risque de cancer du sein invasif, construits à partir du modèle de Cox, de l'algorithme des plus proches voisins et du modèle logistique de Gail, échantillon de validation, cohorte E3N (N = 13 174).

	AUC			E/O		
	kNN	Cox	Gail	kNN	Cox	Gail
Score préménopausique	0,583 [0,520 ; 0,646]	0,565 [0,500 ; 0,631]	0,614 [0,554 ; 0,675]	1,26	1,28	0,78
Score postménopausique	0,602 [0,546 ; 0,659]	0,617 [0,561 ; 0,673]	0,549 [0,495 ; 0,604]	1,43	1,40	1,12

Les corrélations des scores obtenus avec le modèle de Cox et le modèle des plus proches voisins sont correctes (0,77 pour le score préménopausique et 0,69 pour le score

postménopausique). En revanche, le score obtenu par le modèle de Gail, qui n'a pas été construit à partir des données de la cohorte E3N mais a été appliqué sur ces données, est modestement corrélé aux deux autres scores (les coefficients de corrélation sont compris entre 0,41 et 0,54). En ce qui concerne la concordance, elle est moyenne entre les scores obtenus par le modèle de Cox et l'algorithme des plus proches voisins (0,49 et 0,32), et mauvaise entre ces deux scores et celui obtenu par le modèle de Gail (entre 0,12 et 0,21). Les performances des modèles de prédiction sont proches si on s'intéresse à l'AUC. Le modèle le plus discriminant en préménopause est le score obtenu par le modèle de Gail et en postménopause celui obtenu par le modèle de Cox. Exception faite du score de Gail en préménopause, les scores ont tous tendance à surestimer le nombre de cas de cancer du sein diagnostiqués, surestimation plus faible pour le score fourni par le modèle de Gail en postménopause.

3.4. Discussion

Les combinaisons des facteurs de risque retenues par la méthode du modèle de Cox dans la construction des scores de prédiction du risque de cancer du sein diffèrent selon le statut ménopausique des femmes. Le score de risque de cancer du sein préménopausique ne comporte que des facteurs de risque non comportementaux (âge, antécédents familiaux de cancer du sein et antécédents personnels de maladie bénigne du sein). Le score de risque de cancer du sein postménopausique, quant à lui, comporte à la fois des facteurs de risque non-comportementaux (les trois mêmes et âge à la ménopause) et des facteurs comportementaux (consommation d'alcool, nombre d'enfants, utilisation récente de THM et IMC). Appliqués à la population non ciblée (*i.e.* le score préménopausique appliqué aux femmes ménopausées ou le score postménopausique appliqué aux femmes non-ménopausées), les modèles sont moins performant que lorsqu'appliqués à la population cible. Les performances des modèles obtenues par les méthodes comparées dans cette partie sont très proches les unes des autres, que ce soit pour la discrimination ou pour la calibration des scores, malgré des facteurs retenus différents.

3.4.a. Confrontation à la littérature

Avant la ménopause, les facteurs comportementaux semblent avoir un impact moins important que les facteurs non-comportementaux pour la prédiction à cinq ans du risque de cancer du sein invasif. Ce résultat est en accord avec les deux précédentes parties de ce travail de doctorat (285). Les scores de risque de cancer du sein précédemment publiés fournissent des scores avec une discrimination estimée entre 0,542 et 0,680 (148;153-161;163;165;166;274;286-288) et une calibration estimée entre 0,85 et 1,10 (156-161;163;286;287). Les performances des

scores obtenus par le modèle de Cox et l'algorithme des plus proches voisins se situent dans l'éventail des résultats de la littérature lorsque les performances sont estimées dans l'échantillon d'apprentissage par la méthode de validation croisée. Cependant les rapports E/O estimés dans l'échantillon de validation sont supérieurs à ceux de la littérature. Cette différence de résultat dans l'estimation de la calibration entre les deux échantillons pourrait provenir d'un nombre de cas de cancer significativement plus élevé dans l'échantillon d'apprentissage que dans l'échantillon de validation. Le nombre de prédicteurs sélectionnés dans les scores de la littérature se situe entre 3 et 15 (147-161;163-166;274;286-292). À notre connaissance, seulement deux études ont construit des scores différents selon le statut ménopausique des femmes (149;153) et tous deux ont retenu un nombre plus élevé de facteurs pour la prédiction du risque de cancer du sein en postménopause qu'en préménopause. Les performances des scores de prédiction pourraient être améliorées par la considération d'autres facteurs potentiellement prédictifs ; en particulier la densité mammaire. En effet, les scores de risque précédemment publiés utilisant cette information produisent des estimations d'AUC plus élevées (entre 0,62 et 0,68) (153;165;286;288).

3.4.b. Limites de l'étude

La population constituée des femmes E3N n'est pas représentative de la population générale française. Ainsi, les résultats obtenus concernant les valeurs prédites du risque de cancer du sein ne peuvent pas être généralisés. Cependant, la comparaison des scores reste valable.

Nous aurions pu considérer, au lieu du nombre d'enfants, le nombre de grossesses à terme qui est plus fortement associé au risque de cancer du sein (ces deux variables sont différentes pour 1 146 femmes de la population d'analyse en raison de grossesses multiples). Cependant, des analyses complémentaires en remplaçant le nombre d'enfants par le nombre de grossesses à terme produisent des résultats similaires pour le score postménopausique (l'AUC était de 0,618 et le rapport E/O de 1,39).

De même, les antécédents de biopsie mammaire pratiquée auraient pu être considérés à la place des antécédents de maladie bénigne du sein, comme ce qui est fait dans la construction du score de Gail (146). Les résultats obtenus en faisant cette considération étaient également très similaires, que ce soit pour le score préménopausique (l'AUC était de 0,568 et le rapport E/O de 1,28) ou pour le score postménopausique (l'AUC était de 0,617 et le rapport E/O de 1,40).

Il aurait été intéressant de tester un nombre plus élevé de prédicteurs potentiels. En effet, lorsque l'on teste l'ajout dans les modèles sélectionnés de certaines variables, associées dans la

littérature à des augmentations modestes du risque de cancer du sein (antécédent d'alcool, silhouette à la puberté, nombre de naevi, allaitement, poids à la naissance) les performances des modèles sont légèrement modifiées (AUC entre 0,558 et 0,573 pour le score préménopausique et entre 0,553 et 0,621 pour le score postménopausique). Des combinaisons de variables différentes auraient pu être sélectionnées si on avait inclus dès le départ un plus large éventail de facteurs de risque. De plus, des informations complémentaires concernant certains facteurs associés à une augmentation de risque de cancer du sein, qui ne sont pas (ou pas encore) disponibles dans la cohorte E3N auraient pu permettre l'obtention de scores prédictifs plus performants, comme par exemple la densité mammaire ou les facteurs génétiques.

Enfin, une plus grande puissance statistique aurait pu permettre l'investigation de la prédiction à cinq ans du risque de cancer du sein invasif selon le type histologique de la tumeur.

3.4.c. Conclusion

Cette étude apporte des preuves supplémentaires concernant la différence d'étiologie entre les cancers du sein en préménopause et en postménopause. Les scores de prédiction du risque de cancer du sein devraient prendre en compte cette différence lors de leur construction. Une grande variété de facteurs de risque devrait également être prise en compte, en incluant les facteurs associés à une augmentation, même modeste, du risque de cancer du sein.

La comparaison des différentes méthodes permet de souligner que l'algorithme des plus proches voisins, un outil innovant, intuitif et d'une meilleure compréhensibilité pour le grand public, possède des performances proches des modèles construits à l'aide de méthodes statistiques classiques et couramment utilisées.

Partie 4 Influence du mode de vie sur le risque de mortalité

4.1. Introduction

Comme précisé dans la section Résultats, partie 1.1 page 133 les facteurs de mode de vie, comme le fait de ne pas fumer, avoir un poids sain, la limitation de la consommation d'alcool, l'adoption d'habitudes alimentaires spécifiques, comme une consommation élevée de fruits et légumes, et la pratique régulière d'une activité physique ont été régulièrement associés à une diminution de la mortalité (47) et du risque de maladie chronique (47;293-295). Ils ont ainsi été reconnus comme des comportements sains et des recommandations nationales de santé publique sont régulièrement diffusées (174;239;296) afin d'informer la population sur l'amélioration de l'espérance de vie qu'apporte l'adoption d'un mode de vie sain combinant ces comportements de santé.

Les effets individuels de ces facteurs de mode de vie sur le risque de mortalité ont été largement explorés dans la littérature (section Introduction, partie 3.4 page 64). La corrélation de ces facteurs (section Introduction, partie 3.4.f page 67) incite néanmoins à analyser leur effet conjoint sur le risque de mortalité. Une récente méta-analyse sur le sujet a estimé que l'adoption d'au moins quatre comportements sains entraînerait une réduction du risque de mortalité de 78 %. Néanmoins, peu d'études ont estimé la proportion de décès attribuables (section Méthodes statistiques, partie 3 page 130) à un comportement malsain (section Introduction, partie 3.4.f page 67).

Le but de cette partie est d'estimer la proportion de décès, globalement et par cause spécifique, attribuables à un mode de vie sain, établi à partir de cinq caractéristiques comportementales.

4.2. Matériel et méthodes

4.2.a. Population d'étude

Dans cette partie, les femmes sont suivies de la date de réponse au questionnaire alimentaire Q3 (1993) jusqu'au premier événement observé entre la date de diagnostic de n'importe quel cancer autre que le cancer cutané basal, la date de réponse au dernier questionnaire rempli ou la date d'envoi du dernier questionnaire disponible (Q9, 06/2008). Les femmes non-répondantes au questionnaire alimentaire (N = 24 473), celles sans suivi après Q3 (N = 936), les décès sans date de décès spécifiée (N = 8), les femmes dont l'information sur au moins un parmi les cinq comportements de mode de vie d'intérêt est manquante (N = 5 583) et celles présentant une

valeur de dépense énergétique extrême (N = 1 274) ont été exclues des analyses. La population d'analyse comprend ainsi 66 721 femmes, âgées de 42 à 72 ans au début du suivi. Pour les analyses concernant la mortalité par cause spécifique, les femmes décédées d'une cause inconnue (tableau 33 page 184) ont été exclues (N = 161), portant la population d'analyse à 66 560.

4.2.b. Collecte des données et construction de l'indice de santé

Les variables d'exposition principales prises en considération dans les analyses sont les mêmes que celles définies dans la section Méthodes statistiques, partie 1 page 133, à savoir le tabagisme actif, l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique récréationnelle. La construction de l'indice de santé est explicitée dans le tableau 17 page 134. Les seuils utilisés pour la variable tabac ont été modifiés en séparant les anciennes fumeuses en deux catégories : les anciennes fumeuses régulières (*i.e.* fumant plus d'une cigarette par jour) qui ont arrêté de fumer depuis au moins 10 ans et les anciennes fumeuses occasionnelles ont été considérées comme adhérentes partiellement aux recommandations de santé publique (0,5 point) et les anciennes fumeuses régulières qui ont arrêté de fumer depuis moins de 10 ans ont été considérées comme n'adhérant pas aux recommandations de santé publique (0 point).

4.2.c. Méthodologie statistique

L'analyse de l'association entre chaque comportement de mode de vie, l'indice de santé et le risque de mortalité a été réalisée à l'aide de modèles de régression de Cox, standards et à risque compétitif, en utilisant l'âge comme échelle de temps (section Méthodes statistiques, partie 2.7 page 116). Les HR et leur intervalle de confiance à 95 % pour le risque de mortalité, toutes causes confondues et par cause spécifique, ont été estimés en comparant les femmes qui adhèrent le mieux aux recommandations de santé publique aux femmes qui y adhèrent le moins bien.

Les variables d'ajustement prises en considération dans les différentes analyses effectuées sont la dépense énergétique totale (découpée en quartiles), le statut marital (mariée/en couple, célibataire/veuve/séparée/divorcée), le niveau d'éducation (inférieur à Bac, Bac-Bac+2, supérieur à BAC+2), les antécédents familiaux de cancer chez les parents au premier degré (au moins un, aucun), la zone de résidence (sud, centre, nord), l'activité professionnelle (active, retraitée/n'a jamais travaillé), le statut ménopausique et l'utilisation récente de THM (préménopause, postménopause avec THM, postménopause sans THM). L'hypothèse de log-linéarité de la seule variable continue considérée (la dépense énergétique totale) a été testée. L'hypothèse n'ayant pas été vérifiée, la variable a été catégorisée selon ses quartiles de distribution dans la population

d'analyse. L'hypothèse de proportionnalité des risques a également été vérifiée pour l'ensemble des variables d'ajustement. Seule la variable renseignant la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète, sur laquelle il était initialement prévu d'ajuster, ne respectait pas cette hypothèse. Il a été choisi de stratifier sur cette variable (au moins un, aucun), en complément de l'année de naissance ([1925-1930], [1931-1935], [1936-1945], [1946-1950]) permettant de prendre en compte l'effet cohorte (section Méthodes statistiques, partie 2.9 page 116), en estimant un effet de base différent selon les différentes modalités de ces deux variables. L'information sur le statut ménopausique et l'utilisation de THM a été mise à jour à tous les questionnaires, alors que toutes les autres variables d'ajustement ont été évaluées au début du suivi (voir la section Matériel, partie 4 page 94 pour la collecte des informations sur ces données).

Les fractions attribuables aux cinq comportements individuels et à leur combinaison ont été estimées d'après la méthodologie présentée dans la section Méthodes statistiques, partie 3 page 130. Les fractions attribuables individuelles expriment le pourcentage de décès attribuables à la non-adhésion à l'une des cinq recommandations de santé publique, toutes choses égales par ailleurs. La fraction attribuable à la combinaison de ces cinq comportements exprime le pourcentage de décès attribuables à un indice de santé non maximal.

4.3. Résultats

4.3.a. Caractéristiques de la population d'analyse

Les 66 721 femmes de la population d'analyse ont été suivies pendant un total de 907 632 femmes-années (avec un suivi médian de 10 ans pour les cas de décès et de 15 pour les non-cas). Parmi ces femmes, 2 638 femmes sont décédées au cours du suivi ; la distribution des causes initiales de ces décès est présentée dans le tableau 33.

RÉSULTATS - PARTIE 4

Tableau 33 : Causes initiales des décès survenus dans la population d'analyse, cohorte E3N (N = 66 721).

Cause initiale de décès	Effectif	Pourcentage
Cancer, dont :	1 645	62,36
Sein	505	30,70 ^a
Digestifs et côlon-rectum	226	13,74 ^a
Poumon	169	10,27 ^a
Hémopathie	143	8,69 ^a
Pancréas	134	8,15 ^a
Ovaires	122	7,42 ^a
Maladie de l'appareil circulatoire, dont :	273	10,35
Maladie cérébro-vasculaire	87	31,87 ^b
Cardiopathie non ischémique	64	23,44 ^b
Cardiopathie ischémique	55	20,15 ^b
Embolie pulmonaire	17	6,23 ^b
Hypertension essentielle	11	4,03 ^b
Cause externe, dont :	164	6,22
Accident, autre que transport (noyade, chute, intoxication)	44	26,83 ^c
Suicide	36	21,95 ^c
Accident de transport	21	12,80 ^c
Autre cause, dont :	395	14,97
Maladies de l'appareil respiratoire	58	14,68 ^d
Maladies de l'appareil digestif	47	11,90 ^d
Tumeur non invasive	45	11,39 ^d
Maladie du neurone moteur	36	9,11 ^d
Maladie d'Alzheimer	33	8,35 ^d
Maladie de Parkinson	15	3,80 ^d
Cause inconnue	161	6,10

Note : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

a : Pourcentage parmi les décès par cancer.

b : Pourcentage parmi les décès par maladie de l'appareil circulatoire.

c : Pourcentage parmi les décès par cause externe.

d : Pourcentage parmi les décès par une autre cause que cancer, maladie de l'appareil circulatoire et cause externe.

Plus de la moitié des décès est due à un cancer (62,4 %), et 30,7 % d'entre eux sont des cancers du sein. Les autres causes de décès sont les maladies de l'appareil circulatoire, qui concernent 10,4 % des décès, et les décès par une cause externe (6,2 % des décès). Pour 6,1 % de l'ensemble des décès, la cause initiale n'est pas déterminée.

Les caractéristiques des femmes de la population d'analyse selon la catégorie de l'indice de santé sont présentées dans le tableau 45 (en annexe page 229).

Les femmes de la population d'analyse adhèrent bien aux recommandations de santé publique : 28,7 % d'entre elles adhèrent totalement à quatre des recommandations et totalement

ou partiellement à la cinquième (indice de santé entre 4,5 et 5). Seulement 6,5 % de la population d'analyse adhèrent à deux recommandations, ou moins. Les populations composées des femmes ayant un indice compris entre 0 et 2 et des femmes ayant un indice compris entre 4,5 et 5 présentent des différences statistiquement significatives (tableau 45 en annexe page 229). En particulier, on observe qu'une proportion plus élevée de femmes qui adhèrent le mieux aux recommandations de santé publique résident dans le sud de la France (32,1 %), comparées aux femmes qui y adhèrent le moins bien (24,5 %), de même pour la proportion de femmes retraitées ou n'ayant jamais travaillé (36,5 % *vs.* 21,8 %). Enfin, les femmes qui ont un indice compris entre 4,5 et 5 sont en moyenne plus âgées que les femmes qui ont un indice compris entre 0 et 2 (53,6 ans *vs.* 51,6 ans).

4.3.b. Association entre les comportements individuels, leur combinaison et le risque de décès, toutes causes confondues et par cause initiale spécifique

Les résultats concernant l'estimation du risque de décès, toutes causes confondues et par cause spécifique de décès, associé à chaque score individuel sont présentés dans le tableau 34.

Tableau 34 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de décès, toutes causes confondues (N = 66 721) et par cause initiale spécifique (N = 66 560), cohorte E3N.

Cause initiale ^a	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Décès, toutes causes confondues (N = 2 638)										
0	534 (20,24)	1,00	169 (6,40)	1,00	565 (21,42)	1,00	380 (14,40)	1,00	585 (22,18)	1,00
0,5	617 (23,39)	0,75 [0,67 ; 0,85]	690 (26,16)	0,77 [0,65 ; 0,91]	532 (20,17)	0,92 [0,81 ; 1,03]	460 (17,44)	0,82 [0,71 ; 0,94]	666 (25,25)	0,87 [0,78 ; 0,97]
1	1 487 (56,37)	0,80 [0,72 ; 0,89]	1 779 (67,44)	0,63 [0,54 ; 0,75]	1 541 (58,42)	0,92 [0,83 ; 1,01]	1 798 (68,16)	0,79 [0,71 ; 0,89]	1 387 (52,58)	0,83 [0,75 ; 0,91]
p-valeur de tendance		0,0012		<0,0001		0,1182		0,0003		0,0003
Décès par cancer, toutes localisations confondues (N = 1 645)										
0	341 (20,73)	1,00	88 (5,35)	1,00	367 (22,31)	1,00	248 (15,08)	1,00	348 (21,16)	1,00
0,5	732 (23,77)	0,77 [0,67 ; 0,89]	426 (25,90)	0,89 [0,71 ; 1,13]	344 (20,91)	0,92 [0,79 ; 1,07]	281 (17,08)	0,76 [0,64 ; 0,90]	431 (26,20)	0,95 [0,82 ; 1,09]
1	913 (55,50)	0,82 [0,72 ; 0,94]	1 131 (68,75)	0,73 [0,59 ; 0,92]	934 (56,78)	0,87 [0,76 ; 0,98]	1 116 (67,84)	0,76 [0,66 ; 0,87]	866 (52,64)	0,89 [0,78 ; 1,01]
p-valeur de tendance		0,0286		<0,0001		0,0290		0,0011		0,0448
Décès par maladie de l'appareil circulatoire (N = 273)										
0	57 (20,88)	1,00	31 (11,36)	1,00	53 (19,41)	1,00	36 (13,19)	1,00	70 (25,64)	1,00
0,5	58 (21,25)	0,58 [0,40 ; 0,84]	63 (23,08)	0,44 [0,28 ; 0,67]	57 (20,88)	1,04 [0,71 ; 1,52]	41 (15,02)	0,79 [0,50 ; 1,24]	62 (22,71)	0,69 [0,49 ; 0,97]
1	158 (57,88)	0,64 [0,47 ; 0,88]	179 (65,57)	0,47 [0,32 ; 0,70]	163 (59,71)	0,99 [0,72 ; 1,38]	196 (71,79)	0,89 [0,62 ; 1,28]	141 (51,65)	0,66 [0,49 ; 0,88]
p-valeur de tendance		0,0290		0,0208		0,9115		0,7671		0,0096
Décès par cause externe (N = 164)										
0	31 (18,90)	1,00	8 (4,88)	1,00	34 (20,73)	1,00	24 (14,63)	1,00	37 (22,56)	1,00
0,5	43 (26,22)	0,95 [0,59 ; 1,51]	37 (22,56)	0,88 [0,41 ; 1,90]	22 (13,41)	0,62 [0,36 ; 1,06]	37 (22,56)	1,07 [0,64 ; 1,79]	37 (22,56)	0,76 [0,48 ; 1,20]
1	90 (54,88)	0,89 [0,58 ; 1,37]	119 (72,56)	0,82 [0,39 ; 1,70]	108 (65,85)	1,01 [0,67 ; 1,52]	103 (62,80)	0,80 [0,51 ; 1,26]	90 (54,88)	0,86 [0,58 ; 1,27]
p-valeur de tendance		0,5886		0,4968		0,5517		0,1794		0,5815
Décès d'une autre cause (N = 395)										
0	72 (18,23)	1,00	31 (7,85)	1,00	71 (17,97)	1,00	57 (14,43)	1,00	98 (24,81)	1,00
0,5	93 (23,54)	0,77 [0,56 ; 1,05]	120 (30,38)	0,80 [0,54 ; 1,20]	82 (20,76)	1,12 [0,81 ; 1,55]	72 (18,23)	0,84 [0,59 ; 1,19]	94 (23,80)	0,73 [0,55 ; 0,98]
1	230 (58,23)	0,77 [0,58 ; 1,01]	244 (61,77)	0,58 [0,39 ; 0,86]	242 (61,27)	1,14 [0,86 ; 1,51]	266 (67,34)	0,72 [0,54 ; 0,97]	203 (51,39)	0,71 [0,55 ; 0,90]
p-valeur de tendance		0,1129		0,0003		0,3907		0,0220		0,0105

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

b : Ajusté sur les quatre autres comportements, l'âge, le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, le statut marital, l'apport énergétique journalier, le statut ménopausique et l'utilisation récente de THM ; stratifié sur l'année de naissance et la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète.

Nous pouvons observer une diminution du risque de décès, toutes causes confondues, lors de l'adhésion à chacune des cinq recommandations de santé publique, avec des HR entre 0,63 et 0,92, statistiquement significatifs sauf pour la consommation d'alcool dont l'HR est à la limite de la significativité statistique. Cette diminution de risque pour l'adhésion aux recommandations est retrouvée pour le risque de décès par cancer, avec, cette fois-ci, une diminution du risque statistiquement significative pour une consommation de moins d'un verre d'alcool par jour (HR = 0,87 [0,76 ; 0,98], comparée à une consommation de plus de deux verres par jour) et un résultat à la limite de la significativité statistique pour le niveau d'activité physique récréationnelle (HR = 0,89 [0,78 ; 1,01] pour un niveau de plus de 20 METs-h/semaine comparé à un niveau de moins de 10 METs-h/semaine). En ce qui concerne le risque de décès par maladie de l'appareil circulatoire, l'adhésion à trois des recommandations individuelles de santé publique est associée à une diminution du risque : ne jamais avoir fumé (HR = 0,64 [0,47 ; 0,88], comparé aux fumeuses ou anciennes fumeuses régulières ayant arrêté depuis moins de 10 ans), avoir un IMC entre 18,5 et 25 (HR = 0,47 [0,32 ; 0,70], comparé aux femmes obèses ou en dénutrition, *i.e.* avec un IMC supérieur à 30 ou inférieur à 16) et avoir un niveau d'activité physique récréationnelle d'au moins 20 METs-h/semaine (HR = 0,66 [0,49 ; 0,88], comparé aux femmes avec un niveau de moins de 10 METs-h/semaine). Le risque de décès par cause externe ne semble pas être associé à l'adhésion à ces cinq recommandations de santé publique, tandis que le risque de décès par une autre cause (*i.e.* par maladie, autre que cancer et maladie de l'appareil circulatoire) est statistiquement réduit chez les femmes ayant un IMC entre 18,5 et 25 (HR = 0,58 [0,39 ; 0,86], comparé aux femmes obèses ou en dénutrition, *i.e.* avec un IMC supérieur à 30 ou inférieur à 16), consommant au moins cinq portions de fruits et légumes par jour (HR = 0,72 [0,54 ; 0,97], comparé aux femmes consommant moins de 3,5 portions de fruits et légumes par jour) et pratiquant une activité physique récréationnelle à hauteur d'au moins 20 METs-h/semaine (HR = 0,71 [0,55 ; 0,90], comparé aux femmes avec un niveau de moins de 10 METs-h/semaine).

Les résultats concernant l'association entre le risque de décès, toutes causes confondues et par cause spécifique de décès, et l'indice de santé sont présentés dans le tableau 35.

RÉSULTATS - PARTIE 4

Tableau 35 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de décès, toutes causes confondues (N = 66 721) et par cause spécifique (N = 66 560), cohorte E3N.

Cause initiale ^a	Catégories de l'indice de santé			
	[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Décès, toutes causes confondues (N = 2 638)				
N (%)	236 (8,95)	637 (24,15)	1 083 (41,05)	682 (25,85)
HR [IC 95 %] ^b	1,00	0,71 [0,61 ; 0,83]	0,61 [0,53 ; 0,70]	0,53 [0,45 ; 0,61]
HR (augmentation de 1 point)		0,83 [0,80 ; 0,87]		
P-valeur de tendance		<0,0001		
Décès par cancer, toutes localisations confondues (N = 1 645)				
N (%)	147 (8,94)	402 (24,44)	676 (41,09)	420 (25,53)
HR [IC 95 %] ^b	1,00	0,74 [0,61 ; 0,89]	0,63 [0,53 ; 0,76]	0,55 [0,45 ; 0,66]
HR (augmentation de 1 point)		0,83 [0,79 ; 0,88]		
P-valeur de tendance		<0,0001		
Décès par maladie de l'appareil circulatoire (N = 273)				
N (%)	31 (11,36)	64 (23,44)	105 (38,46)	73 (26,74)
HR [IC 95 %] ^b	1,00	0,51 [0,33 ; 0,78]	0,41 [0,27 ; 0,61]	0,37 [0,24 ; 0,57]
HR (augmentation de 1 point)		0,76 [0,67 ; 0,87]		
P-valeur de tendance		<0,0001		
Décès par cause externe (N = 164)				
N (%)	13 (7,93)	40 (24,39)	68 (41,46)	43 (26,22)
HR [IC 95 %] ^b	1,00	0,83 [0,45 ; 1,56]	0,73 [0,40 ; 1,33]	0,64 [0,34 ; 1,20]
HR (augmentation de 1 point)		0,87 [0,73 ; 1,03]		
P-valeur de tendance		0,1067		
Décès d'une autre cause (N = 395)				
N (%)	30 (7,59)	95 (24,05)	173 (43,80)	97 (24,56)
HR [IC 95 %] ^b	1,00	0,79 [0,53 ; 1,20]	0,71 [0,48 ; 1,05]	0,53 [0,35 ; 0,80]
HR (augmentation de 1 point)		0,82 [0,73 ; 0,92]		
P-valeur de tendance		0,0005		

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

b : Ajusté sur l'âge, le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, le statut marital, l'apport énergétique journalier, le statut ménopausique et l'utilisation récente de THM ; stratifié sur l'année de naissance et la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète.

Comparées aux femmes qui adhèrent le moins aux recommandations de santé publique (indice de santé entre 0 et 2), les femmes qui y adhèrent le plus (indice de santé entre 4,5 et 5) sont à risque diminué de mortalité, toutes causes confondues (HR = 0,53 [0,45 ; 0,61], p-valeur de tendance < 0,001). Cette association est observée pour les décès par cancer (HR = 0,55 [0,45 ; 0,66], p < 0,001), par maladie de l'appareil circulatoire

(HR = 0,37 [0,24 ; 0,57], $p < 0,001$) et par une autre cause de maladie (HR = 0,53 [0,35 ; 0,80], $p < 0,001$). La diminution du risque de mortalité globale ou par cancer, maladie de l'appareil circulatoire ou autre cause est observée dès l'adhésion à une recommandation de santé publique supplémentaire (HR entre 0,76 et 0,83, statistiquement significatifs au seuil de 5 %, $p < 0,0005$). L'association entre l'adhésion aux recommandations de santé et le risque de décès par cause externe n'est pas statistiquement significative ($p = 0,1067$).

4.3.c. Association entre les comportements individuels, leur combinaison et le risque de décès par localisations spécifiques de cancer

Les résultats concernant l'estimation de l'association entre le risque de décès par cancer selon sa localisation et chaque score individuel sont présentés dans le tableau 36.

Tableau 36 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de décès par cancer, selon la localisation de la tumeur, cohorte E3N (N = 66 560).

Cause initiale ^a	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Décès par cancer du sein (N = 503)										
0	81 (16,04)	1,00	26 (5,15)	1,00	120 (23,76)	1,00	86 (17,03)	1,00	107 (21,19)	1,00
0,5	139 (27,52)	1,24 [0,94 ; 1,64]	144 (28,51)	1,03 [0,68 ; 1,57]	105 (20,79)	0,85 [0,65 ; 1,10]	89 (17,62)	0,70 [0,52 ; 0,94]	121 (23,96)	0,90 [0,70 ; 1,18]
1	285 (56,44)	1,23 [0,95 ; 1,58]	335 (66,34)	0,75 [0,50 ; 1,13]	280 (55,45)	0,76 [0,61 ; 0,96]	330 (65,35)	0,64 [0,50 ; 0,82]	277 (54,85)	1,03 [0,82 ; 1,30]
P-valeur de tendance	0,1262		0,0003		0,0235		0,0020		0,6940	
Décès par cancer digestif et du côlon-rectum (N = 226)										
0	44 (19,47)	1,00	12 (5,31)	1,00	53 (23,45)	1,00	41 (18,14)	1,00	47 (20,80)	1,00
0,5	58 (25,66)	0,86 [0,58 ; 1,27]	55 (24,34)	0,80 [0,43 ; 1,51]	47 (20,80)	0,89 [0,60 ; 1,32]	35 (15,49)	0,55 [0,35 ; 0,87]	48 (21,24)	0,79 [0,53 ; 1,18]
1	124 (54,87)	0,80 [0,56 ; 1,14]	159 (70,35)	0,71 [0,39 ; 1,29]	126 (55,75)	0,85 [0,61 ; 1,19]	150 (66,37)	0,56 [0,39 ; 0,80]	131 (57,96)	0,97 [0,69 ; 1,36]
P-valeur de tendance	0,2244		0,1752		0,3718		0,0073		0,8962	
Décès par cancer du poumon (N = 169)										
0	74 (43,79)	1,00	7 (4,14)	1,00	37 (21,89)	1,00	26 (15,38)	1,00	42 (24,85)	1,00
0,5	36 (21,30)	0,31 [0,21 ; 0,47]	32 (18,93)	0,89 [0,39 ; 2,02]	37 (21,89)	1,09 [0,69 ; 1,73]	39 (23,08)	1,09 [0,66 ; 1,79]	52 (30,77)	0,89 [0,59 ; 1,35]
1	59 (34,91)	0,23 [0,16 ; 0,32]	130 (76,92)	1,06 [0,49 ; 2,30]	95 (56,21)	1,09 [0,73 ; 1,63]	104 (61,54)	0,78 [0,50 ; 1,22]	75 (44,38)	0,59 [0,40 ; 0,87]
P-valeur de tendance	<0,0001		0,5529		0,6895		0,1281		0,0035	
Décès par hémopathie maligne (N = 143)										
0	22 (15,38)	1,00	8 (5,59)	1,00	35 (11,89)	1,00	17 (11,89)	1,00	28 (19,58)	1,00
0,5	32 (22,38)	0,94 [0,54 ; 1,63]	43 (15,38)	0,95 [0,44 ; 2,06]	22 (13,99)	0,60 [0,35 ; 1,03]	20 (13,99)	0,79 [0,41 ; 1,51]	35 (24,48)	0,94 [0,57 ; 1,56]
1	89 (62,24)	1,18 [0,72 ; 1,91]	92 (60,14)	0,66 [0,31 ; 1,40]	86 (74,13)	0,79 [0,52 ; 1,21]	106 (74,13)	1,00 [0,59 ; 1,69]	80 (55,94)	0,95 [0,61 ; 1,47]
P-valeur de tendance	0,3594		0,0484		0,4691		0,7085		0,8314	

Suite page suivante

Cause initiale ^a	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Décès par cancer du pancréas (N = 122)										
0	25 (18,66)	1,00	12 (8,96)	1,00	25 (18,66)	1,00	16 (11,94)	1,00	19 (14,18)	1,00
0,5	28 (20,90)	0,75 [0,44 ; 1,29]	26 (17,40)	0,39 [0,19 ; 0,78]	40 (29,85)	1,57 [0,95 ; 2,60]	22 (16,42)	1,81 [1,06 ; 3,10]	45 (33,58)	0,89 [0,47 ; 1,70]
1	81 (60,45)	1,00 [0,63 ; 1,59]	96 (71,64)	0,43 [0,23 ; 0,80]	69 (51,49)	0,93 [0,58 ; 1,50]	96 (71,64)	1,29 [0,77 ; 2,15]	70 (52,24)	0,96 [0,56 ; 1,65]
P-valeur de tendance		0,7148		0,1483		0,3638		0,9806		0,7660
Décès par cancer des ovaires (N = 122)										
0	24 (19,67)	1,00	3 (2,46)	1,00	19 (15,57)	1,00	17 (13,93)	1,00	27 (22,13)	1,00
0,5	27 (22,13)	0,71 [0,41 ; 1,23]	29 (23,77)	1,90 [0,57 ; 6,25]	28 (22,95)	1,43 [0,80 ; 2,58]	20 (16,39)	0,78 [0,41 ; 1,49]	36 (29,51)	0,98 [0,59 ; 1,62]
1	71 (58,20)	0,81 [0,50 ; 1,32]	90 (73,77)	1,92 [0,60 ; 6,12]	75 (61,48)	1,34 [0,79 ; 2,27]	85 (69,67)	0,84 [0,49 ; 1,43]	59 (48,36)	0,74 [0,47 ; 1,18]
P-valeur de tendance		0,5846		0,5088		0,3578		0,6652		0,1340
Décès d'un autre type de cancer (N = 346)										
0	71 (20,52)	1,00	20 (5,78)	1,00	78 (22,54)	1,00	45 (13,01)	1,00	78 (22,54)	1,00
0,5	71 (20,52)	0,65 [0,47 ; 0,91]	97 (28,03)	0,90 [0,55 ; 1,46]	65 (18,79)	0,79 [0,57 ; 1,10]	56 (16,18)	0,84 [0,57 ; 1,25]	94 (27,17)	0,91 [0,67 ; 1,23]
1	204 (58,96)	0,86 [0,65 ; 1,14]	229 (66,18)	0,63 [0,40 ; 1,02]	203 (58,67)	0,84 [0,64 ; 1,11]	245 (70,81)	0,94 [0,68 ; 1,30]	174 (50,29)	0,77 [0,58 ; 1,01]
P-valeur de tendance		0,7426		0,0018		0,3076		0,9724		0,0399

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

b : Ajusté sur les quatre autres comportements, l'âge, le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, le statut marital, l'apport énergétique journalier, le statut ménopausique et l'utilisation récente de THM ; stratifié sur l'année de naissance et la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète.

La diminution du risque de décès par cancer associé au fait de n'avoir jamais fumé (tableau 34) n'est retrouvée que pour le décès par cancer du poumon (HR = 0,23 [0,16 ; 0,32], comparé aux femmes fumeuses ou anciennes fumeuses depuis moins de 10 ans), de même que celle associée à un niveau d'activité physique d'au moins 20 METs-h/semaine (HR = 0,59 [0,40 ; 0,87], comparé aux femmes avec un niveau d'activité physique de moins de 10 METs-h/semaine). Seul le fait d'avoir un IMC entre 18,5 et 25 (HR = 0,43 [0,23 ; 0,80], comparé aux femmes obèses ou en dénutrition) est associé à une diminution statistiquement significative du risque de décès par cancer du pancréas. Une association avec la consommation de moins d'un verre d'alcool par jour n'est retrouvée que pour le risque de décès par cancer du sein (HR = 0,76 [0,61 ; 0,96], comparé aux femmes consommant plus de deux verres par jour). La diminution du risque de décès par cancer observée pour une consommation d'au moins cinq portions de fruits et légumes par jour, comparée à une consommation de moins de 3,5 portions de fruits et légumes par jour (tableau 34) se retrouve pour deux localisations de cancer : le sein (HR = 0,76 [0,61 ; 0,96]) et l'appareil digestif, autre que côlon-rectum (HR = 0,56 [0,39 ; 0,80]).

Les résultats concernant l'association entre les comportements individuels et le risque de décès par hémopathie maligne, cancer des ovaires ou d'une autre localisation ne sont pas statistiquement significatifs, hormis une tendance statistiquement significative entre le risque de décès par hémopathie maligne et l'IMC (p -valeur de tendance = 0,0484).

Les résultats concernant l'association entre le risque de décès par cancer selon sa localisation et l'indice de santé sont présentés dans le tableau 37.

Tableau 37 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de décès par cancer, selon la localisation de la tumeur, cohorte E3N (N = 66 560).

Cause initiale ^a	Catégories de l'indice de santé			
	[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Décès par cancer du sein (N = 505)				
N (%)	54 (10,69)	121 (23,96)	190 (37,63)	140 (27,72)
HR [IC 95 %] ^b	1,00	0,64 [0,46 ; 0,88]	0,53 [0,39 ; 0,72]	0,57 [0,41 ; 0,78]
HR (augmentation de 1 point)		0,86 [0,78 ; 0,95]		
P-valeur de tendance		0,0026		
Décès par cancer digestif et du côlon-rectum (N = 226)				
N (%)	24 (10,62)	42 (18,58)	108 (47,79)	52 (23,01)
HR [IC 95 %] ^b	1,00	0,45 [0,27 ; 0,74]	0,56 [0,36 ; 0,88]	0,36 [0,22 ; 0,59]
HR (augmentation de 1 point)		0,80 [0,69 ; 0,93]		
P-valeur de tendance		0,0032		
Décès par cancer du poumon (N = 169)				
N (%)	22 (13,02)	56 (33,14)	65 (38,46)	26 (15,38)
HR [IC 95 %] ^b	1,00	0,67 [0,41 ; 1,10]	0,39 [0,24 ; 0,63]	0,21 [0,12 ; 0,38]
HR (augmentation de 1 point)		0,59 [0,50 ; 0,70]		
P-valeur de tendance		<0,0001		
Décès par hémopathie maligne (N = 143)				
N (%)	<u>5 (3,50)</u>	<u>40 (27,97)</u>	52 (36,36)	46 (32,17)
HR [IC 95 %] ^b		1,00	0,74 [0,49 ; 1,10]	0,89 [0,58 ; 1,34]
HR (augmentation de 1 point)		0,98 [0,81 ; 1,19]		
P-valeur de tendance		0,8551		
Décès par cancer du pancréas (N = 134)				
N (%)	<u>9 (6,72)</u>	<u>31 (23,13)</u>	55 (41,05)	39 (29,10)
HR [IC 95 %] ^b		1,00	0,89 [0,59 ; 1,35]	0,87 [0,56 ; 1,37]
HR (augmentation de 1 point)		0,94 [0,78 ; 1,14]		
P-valeur de tendance		0,5416		
Décès par cancer des ovaires (N = 122)				
N (%)	<u>7 (5,74)</u>	<u>31 (25,41)</u>	47 (38,52)	37 (30,33)
HR [IC 95 %] ^b		1,00	0,81 [0,53 ; 1,24]	0,88 [0,56 ; 1,40]
HR (augmentation de 1 point)		0,95 [0,78 ; 1,17]		
P-valeur de tendance		0,6551		
Décès d'un autre type de cancer (N = 346)				
N (%)	<u>26 (7,52)</u>	<u>81 (23,41)</u>	159 (45,95)	80 (23,12)
HR [IC 95 %] ^b		1,00	0,80 [0,53 ; 1,22]	0,56 [0,36 ; 0,87]
HR (augmentation de 1 point)		0,84 [0,74 ; 0,94]		
P-valeur de tendance		0,0033		

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

b : Ajusté sur l'âge, le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, le statut marital, l'apport énergétique journalier, le statut ménopausique et l'utilisation récente de THM ; stratifié sur l'année de naissance et la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète.

Comparées aux femmes qui adhèrent le moins aux recommandations de santé publique (indice de santé entre 0 et 2 ou entre 0 et 3 pour certaines localisations de cancer), les femmes qui y adhèrent le plus (indice de santé entre 4,5 et 5) sont à risque diminué de décès par cancer du sein (HR = 0,57 [0,41 ; 0,78], p-valeur de tendance = 0,0026), par cancer digestif, y compris côlon-rectum (HR = 0,36 [0,22 ; 0,59], p = 0,0032), par cancer du poumon (HR = 0,21 [0,12 ; 0,38], p < 0,0001) et par un autre type de cancer (HR = 0,56 [0,36 ; 0,87], p = 0,0033). Pour ces localisations de cancer, la diminution du risque de décès est observée dès l'adhésion à une recommandation supplémentaire (HR compris entre 0,59 et 0,86, statistiquement significatifs au seuil de 5 %, p < 0,004).

4.3.d. Proportions de décès attribuables aux comportements individuels de mode de vie et à leur combinaison

Les fractions attribuables individuelles et combinées, estimant le pourcentage de décès (toutes causes confondues ou par cause spécifique) attribuable à la non-adhésion à chacune des cinq recommandations de santé publique prises en considération et à un mode de vie malsain, respectivement, et leur intervalle de confiance à 95 % sont présentées dans le tableau 38.

Tableau 38 : Proportions [IC 95 %] de décès, toutes causes confondues (N = 66 721) et par cause spécifique (N = 66 560), attribuables à la non-adhésion aux recommandations individuelles de santé publique et à leur combinaison, cohorte E3N.

Cause initiale ^a	Proportions de décès attribuables aux comportements de mode de vie ^b					Proportions de décès attribuables à un mode de vie malsain ^c
	Tabagisme actif	IMC	Consommation d'alcool	Consommation de fruits et légumes	Niveau d'activité physique récréationnelle	
Décès, toutes causes confondues (N = 2 638)	3,03 [-1,57 ; 7,62]	5,86 [3,38 ; 8,32]	2,02 [-2,16 ; 6,18]	3,93 [0,41 ; 7,45]	5,09 [0,69 ; 9,47]	16,62 [9,10 ; 23,94]
Décès par cancer, toutes localisations confondues (N = 1 645)	2,40 [-3,41 ; 8,19]	5,50 [2,43 ; 8,56]	4,28 [-1,07 ; 9,60]	4,13 [-0,36 ; 8,60]	4,03 [-1,50 ; 9,54]	16,36 [6,80 ; 25,62]
Décès par maladie de l'appareil circulatoire (N = 273)	7,31 [-7,76 ; 22,05]	2,88 [-5,18 ; 10,90]	1,07 [-11,60 ; 13,66]	-0,41 [-11,50 ; 10,72]	10,26 [-3,79 ; 23,90]	18,57 [-4,61 ; 39,84]
Décès par cause externe (N = 164)	3,48 [-13,30 ; 20,09]	2,52 [-7,83 ; 12,81]	-8,78 [-24,70 ; 7,64]	8,50 [-4,79 ; 21,50]	0,32 [-16,90 ; 17,53]	13,72 [-1,64 ; 41,48]
Décès d'une autre cause (N = 395)	5,18 [-6,73 ; 16,94]	9,81 [3,34 ; 16,21]	-2,89 [-13,10 ; 7,36]	7,62 [-1,66 ; 16,78]	8,70 [-3,00 ; 20,17]	23,69 [4,56 ; 41,13]
Décès par cancer du sein (N = 503)	-3,49 [-13,10 ; 6,19]	7,97 [2,60 ; 13,30]	7,57 [-2,11 ; 17,10]	8,00 [-0,22 ; 16,11]	-4,08 [-14,10 ; 6,01]	4,94 [-1,32 ; 22,76]
Décès par cancers digestifs et du côlon-rectum (N = 226)	6,50 [-9,04 ; 21,73]	4,16 [-4,16 ; 12,42]	4,50 [-10,20 ; 19,01]	8,51 [-3,97 ; 20,74]	-4,23 [-19,30 ; 10,99]	27,90 [3,27 ; 49,34]
Décès par cancer du poumon (N = 169)	42,24 [22,05 ; 58,95]	-3,64 [-13,20 ; 5,94]	-1,93 [-19,10 ; 15,37]	10,54 [-3,76 ; 24,42]	21,13 [3,98 ; 37,08]	51,27 [27,03 ; 69,40]
Décès par hémopathie maligne (N = 143)	-8,75 [-26,10 ; 9,12]	10,11 [-0,70 ; 20,69]	0,14 [-18,10 ; 18,40]	-4,05 [-18,50 ; 10,60]	0,98 [-17,40 ; 19,27]	-3,79 [-38,10 ; 31,40]
Décès par cancer du pancréas (N = 122)	-7,09 [-26,70 ; 13,13]	2,66 [-8,72 ; 13,97]	13,51 [-4,31 ; 30,51]	-0,76 [-15,70 ; 14,24]	5,77 [-12,50 ; 23,65]	5,05 [-30,00 ; 38,85]
Décès par cancer des ovaires (N = 122)	0,73 [-19,80 ; 21,16]	-2,22 [-12,80 ; 8,35]	-3,92 [-21,30 ; 13,74]	1,13 [-14,90 ; 17,13]	13,20 [-6,78 ; 32,17]	-0,07 [-36,60 ; 36,44]
Décès d'un autre type de cancer (N = 346)	-3,43 [-15,90 ; 9,15]	9,71 [2,76 ; 16,56]	2,51 [-9,11 ; 14,06]	-1,17 [-10,70 ; 8,34]	9,88 [-2,23 ; 21,71]	25,36 [5,18 ; 43,56]

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les classifications utilisées pour repérer les différentes causes initiales de décès sont définies dans le tableau 12 page 93.

b : Pourcentage de cas de cancer attribuables à la non-adhésion aux recommandations individuelles de santé publique (i.e. ne jamais fumer, avoir un IMC entre 18,5 et 25, consommer moins d'un verre d'alcool par jour, consommer au moins 5 portions de fruits et légumes par jour, et pratiquer une activité physique équivalente à au moins 20 METs-h/semaine). La fraction attribuable individuelle associée à un comportement de mode de vie est ajustée sur les quatre autres comportements de mode de vie, l'âge, le niveau d'éducation, la zone de résidence, les antécédents familiaux de cancer chez les parents au premier degré, l'activité professionnelle, le statut marital, l'apport énergétique journalier, le statut ménopausique et l'utilisation récente de THM ; elle est également stratifiée sur l'année de naissance et la présence d'antécédents personnels de cancer, maladie de l'appareil circulatoire et diabète.

c : Pourcentage de cas de cancer attribuables à un mode de vie malsain (i.e. un indice de santé strictement inférieur à 4,5), ajustés sur (a), à l'exception des comportements individuels de mode de vie.

Globalement, 16,6 % [9,1 % ; 23,9 %] des décès, toutes causes confondues, sont attribuables à un mode de vie malsain (*i.e.* un indice de santé strictement inférieur à 4,5). Cette proportion est similaire (16,4 % [6,8 % ; 25,6 %]) en ce qui concerne les décès par cancer. Cette proportion de décès par cancer attribuables à un mode de vie malsain est déterminée par les 51,3 % [27,3 % ; 69,4 %] des décès par cancer du poumon, les 27,9 % [3,3 % ; 49,3 %] des décès par cancer digestif ou du côlon-rectum et les 25,4 % [5,2 % ; 43,6 %] des décès par un autre type de cancer attribuables à un indice strictement inférieur à 4,5.

Un total de, respectivement, 5,9 % [3,4 % ; 8,3 %], 3,9 % [0,4 % ; 7,5 %] et 5,1 % [0,7 % ; 9,5 %] des décès, toutes causes confondues, sont attribuables à l'IMC (surpoids, obésité, maigreur et dénutrition), à une consommation de fruits et légumes de moins de cinq portions par jour et à un niveau d'activité physique inférieur à 20 METs-h/semaine. Les proportions de décès attribuables à la non-adhésion à chacune des recommandations individuelles de santé publique varient selon la nature de la cause initiale de décès. Selon notre étude, 42,2 % [22,1 % ; 59,0 %] des décès par cancer du poumon seraient attribuables au tabagisme actif (actuel et passé), 8,0 % [2,6 % ; 13,3 %] des décès par cancer du sein seraient attribuables à l'IMC (surpoids, obésité, maigreur et dénutrition), et 21,1 % [4,0 % ; 37,1 %] des décès par cancer du poumon seraient attribuables à un niveau d'activité physique inadéquat (inférieur à 20 METs-h/semaine). Les autres fractions attribuables n'atteignent pas la significativité statistique.

4.4. Discussion

Dans cette étude portant sur l'association entre l'adhésion à cinq recommandations de santé publique et le risque de décès, toutes causes confondues et par cause spécifique, nous avons pu observer que l'adhésion à une simple recommandation réduit le risque de décès. Plus important encore, l'adhésion à un mode de vie sain, impliquant l'adhésion conjointe aux recommandations de santé publique concernant le tabagisme actif, l'IMC, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique récréationnelle pourrait permettre d'éviter jusqu'à 16,8 % de décès prématurés. Cette proportion est identique pour les décès par cancer (16,7 %) et plus élevée, mais non statistiquement significative, pour les décès par maladie de l'appareil circulatoire (18,3 % [-5,0 % ; 39,6 %]). Parmi les causes de décès par cancer étudiées, les plus impactées par le mode de vie sont le cancer du poumon avec 51,4 % de décès potentiellement évitables et les cancers digestifs et du côlon-rectum avec 27,6 % de décès potentiellement évitables.

4.4.a. Confrontation à la littérature

Les études ayant analysé l'impact du mode de vie sur le risque de mortalité ont toutes reporté une diminution du risque de mortalité, toutes causes confondues ou par causes spécifiques, avec le nombre de comportements sains adoptés. D'après une récente méta-analyse, la diminution du risque associé à l'adhésion à un mode de vie sain a été estimée à 68 % (HR = 0,32 [0,26 ; 0,40] pour l'adhésion à au moins quatre comportements de santé comparée à l'adhésion à aucun) (184). Dans notre étude, la réduction de risque de mortalité par maladie de l'appareil circulatoire était plus prononcée que par cancer. Les résultats de la littérature concernant cette question sont divergents (section Introduction, partie 3.4.f page 67). Avec une proportion de décès attribuables à la non-adhésion au mode de vie de 16,8 %, nous obtenons une estimation plus faible que dans la littérature, qui fournit une étendue de résultats entre 29 % et 60 % pour la mortalité toutes causes confondues (115;117;128;188-190;195-198). Les variations observées peuvent s'expliquer par le choix des comportements inclus, la manière dont ils sont définis, par la catégorisation utilisée et par des prévalences d'exposition différentes selon les pays et les populations d'analyse.

4.4.b. Conclusion

L'adhésion à une seule recommandation de santé publique, parmi le fait de ne pas fumer, de modérer sa consommation d'alcool, de consommer la quantité recommandée de fruits et légumes, d'être physiquement actif et d'avoir un IMC dans la norme de recommandations de l'OMS, pourrait diminuer le risque de mortalité prématurée. L'adhésion conjointe à ces cinq recommandations pourrait diminuer substantiellement le nombre de décès prématurés, particulièrement par cancers du poumon et digestifs. Ces résultats devraient encourager à poursuivre les démarches mises en œuvre pour alerter la population sur les effets bénéfiques sur la santé de ces cinq comportements et à initier des études de recherches sociologiques permettant de s'interroger sur les freins aux changements de comportements.

CONCLUSION GÉNÉRALE

La cohorte française E3N permet l'étude détaillée de nombreuses problématiques de santé publique. Le nombre élevé de participantes, la longue durée de leur suivi, ainsi que la mise à jour fréquente de leurs informations de santé, constituent un atout de taille doté d'une grande puissance statistique et offrant la possibilité d'effectuer des analyses séparées selon la localisation du cancer diagnostiqué, la nature de la cause conduisant au décès, ainsi que de nombreuses analyses en sous-populations. Le recueil prospectif de l'étude réduit le risque de biais de mémoire différentiel, l'information étant collectée en amont de la survenue de l'événement d'intérêt. Le biais de classement est également limité en raison de la confirmation des cas dans la population d'analyse (par l'obtention de comptes rendus histologiques pour les cas de cancer et la confirmation des décès auprès de la mairie de naissance et du CépiDc pour les cas de décès). Enfin, l'influence des comparaisons multiples est limitée en raison du fait que l'ensemble des analyses sont réalisées à partir d'hypothèses sous-jacentes et que les associations observées se sont avérées cohérentes à celles de la littérature. Ces atouts nous ont permis d'évaluer l'influence qu'ont les facteurs comportementaux et non-comportementaux sur la survenue d'un cancer, en particulier d'un cancer du sein, et sur le risque de mortalité prématurée.

Partie 1 Principaux résultats

Dans un premier temps, nous avons mis en évidence que les principaux facteurs comportementaux, à savoir le tabagisme actif, le poids, la consommation d'alcool, la consommation de fruits et légumes et le niveau d'activité physique récréationnelle, sont associés au risque de cancer, toutes localisations confondues. L'influence individuelle de ces facteurs est seulement modeste, cependant, l'adhésion conjointe aux recommandations de santé publique concernant ces cinq facteurs modifiables, permettrait d'éviter une proportion non négligeable de cas de cancer. Des hétérogénéités ont été observées selon la localisation de la tumeur : parmi les localisations étudiées, les cancers du poumon, digestifs (principalement côlon-rectum), de l'endomètre et du sein en postménopause apparaissent être les plus influencés par le mode de vie.

Dans un deuxième temps, nous nous sommes intéressés au poids relatif des facteurs comportementaux et non-comportementaux sur le risque de survenue de cancer du sein en différenciant les analyses selon le statut ménopausique des femmes. Il a été observé que les facteurs non-comportementaux ont un plus fort impact sur le risque de cancer du sein en préménopause que les facteurs comportementaux. L'inverse est observé après la ménopause : la

CONCLUSION GÉNÉRALE

corpulence à la puberté, l'utilisation de traitements hormonaux de la ménopause, les habitudes alimentaires, la consommation d'alcool, le nombre d'enfants et l'âge à la première grossesse influencent davantage le nombre de cas de cancer du sein postménopausiques diagnostiqués dans la population. Les facteurs non-comportementaux associés au risque de cancer du sein sont : les antécédents personnels de maladie bénigne du sein, l'âge à l'apparition des premières règles (particulièrement pour les cancers survenant avant la ménopause), le poids de naissance (essentiellement avant la ménopause) et les antécédents familiaux de cancer du sein (essentiellement après la ménopause). Des différences ont également été observées selon le statut des récepteurs à l'œstrogène des tumeurs et l'expression de la protéine HER2, permettant de souligner les différences d'étiologies de ces différents types de cancer du sein.

Dans un troisième temps, nous avons évalué dans quelles mesures les facteurs comportementaux et non-comportementaux, associés au risque de cancer du sein préménopausique et postménopausique, peuvent prédire le risque à cinq ans de cancer du sein chez les femmes âgées entre 40 et 70 ans. De même que dans la partie précédente, ce sont les facteurs non-comportementaux qui ont le plus fort pouvoir prédictif avant la ménopause, alors qu'en postménopause, la prise en compte de facteurs non-comportementaux dans la construction du score de prédiction du risque permettrait d'obtenir une prédiction plus performante. Ces résultats mettent en valeur les différences d'impact des facteurs de risque de cancer du sein selon le statut ménopausique.

Enfin, nous avons reproduit l'analyse de l'influence des principaux facteurs comportementaux sur le risque de mortalité. L'adhésion conjointe aux cinq recommandations de santé publique analysées pourrait permettre de réduire le nombre de décès prématurés, que la cause soit un cancer, une maladie de l'appareil circulatoire ou une autre maladie. La diminution de risque est observée dès l'adhésion à une recommandation supplémentaire.

Partie 2 Perspectives de recherche

Dans ce travail, nous avons mis en évidence la différence d'étiologie des différents types de cancer du sein et l'influence non-négligeable des facteurs comportementaux, sur lesquels on peut directement agir. L'influence de ces facteurs est particulièrement importante après la ménopause en ce qui concerne le cancer du sein, soulignant que l'amélioration de notre mode de vie est bénéfique tout au long de la vie. Il serait intéressant de poursuivre ce travail en explorant les différences qui pourraient être observées en périménopause, période durant laquelle apparaissent certains troubles à l'approche de la ménopause.

Par l'estimation des proportions de cas de cancer du sein attribuables aux facteurs de risque reconnus dans la littérature ou par la construction de scores prédictifs dont les performances obtenues sont seulement modestes, on peut déduire de ce travail que les facteurs de risque du cancer du sein ne sont pas encore totalement connus. S'agissant du cancer du sein, le recueil de mammographies (actuellement en cours) pour estimer la densité mammaire, et l'extraction de l'ADN pour prendre en compte l'information génétique pourraient conduire à une meilleure prédiction du risque de cancer du sein. Des recherches sont nécessaires afin de mieux connaître les mécanismes de cancérogenèse mammaire dans le but de mieux cibler les actions préventives et les interventions de santé publique. La recherche exhaustive, que nous effectuons, des comptes rendus histologiques de l'ensemble des cancers déclarés par les participantes, permet d'effectuer des analyses selon les sous-types de cancer. L'allongement du suivi des femmes, qui aboutira à un nombre plus élevé de cancers diagnostiqués, permettra d'améliorer la puissance statistique. En particulier, depuis peu, les médecins renseignent quasi-systématiquement l'information sur les récepteurs hormonaux des cancers du sein dans les comptes rendus histologiques, ce qui nous permettra, à terme, de pouvoir identifier les sous-types de cancer du sein luminal A et luminal B*, dont la classification est actuellement utilisée par les professionnels de santé pour identifier les traitements et interventions médicales.

Ce travail nous incite également à évaluer l'impact des facteurs comportementaux et non-comportementaux pour d'autres localisations de cancer, afin de déterminer pour quels types de cancers les actions préventives seraient les plus efficaces. La validation, actuellement en cours de réalisation, des cas de maladie de l'appareil circulatoire rendra possible l'analyse de l'association entre le mode de vie et ces maladies, une analyse qui semble prometteuse au vu des résultats obtenus sur l'association entre le mode de vie et la mortalité par maladie de l'appareil circulatoire.

La mise en place de la nouvelle cohorte E4N, rassemblant les enfants des femmes E3N, ainsi que leurs pères biologiques, et les petits-enfants des femmes E3N, permettra de reproduire ce travail chez les hommes afin d'évaluer si l'adhésion aux recommandations de santé publique entraîne une diminution du risque de cancer, ainsi qu'une amélioration de l'espérance de vie. Cette nouvelle cohorte inclura des participants plus jeunes, conduisant à une meilleure comparabilité entre les jeunes et les séniors. Avec l'inclusion des enfants et petits-enfants des femmes E3N, les comportements de mode de vie à prendre en considération pourront être adaptés à une société moderne, différente de celle des femmes E3N nées entre 1925 et 1950. Ainsi, l'analyse de l'influence de la sédentarité, de l'utilisation de jeux vidéo dynamiques, voire sportifs (de type Kinect[®]), l'identification de la consommation d'aliments provenant de

l'agriculture biologique ou du commerce équitable, le suivi des régimes alimentaires en vogue ces dernières années (les régimes Atkins, Dukan, paléolithique, la chrononutrition) ou encore l'utilisation de cigarettes électroniques, permettra d'évaluer l'impact du mode de vie moderne sur la santé.

Au terme de ce travail, nous pouvons formuler quelques suggestions concernant les études qui seront menées dans l'avenir. Le cancer du sein regroupe plusieurs types de maladies avec des étiologies différentes, ce qui incite à effectuer systématiquement des analyses séparées suivant le statut ménopausique, et d'identifier les hétérogénéités potentielles selon le statut des récepteurs hormonaux. Ce travail a également permis d'établir l'intérêt de confronter les résultats en incidence et en mortalité. Enfin, l'influence de la corpulence à la puberté sur le risque de cancer du sein à l'âge adulte invite à étudier l'impact qu'aurait dans la vie future les expositions subies dans l'enfance et l'adolescence.

Partie 3 Impact en terme de santé publique

Le cancer est un véritable enjeu de santé publique. Les plans cancer se succèdent afin de mieux prendre en charge les individus atteints d'un cancer et d'améliorer la prévention et le dépistage. D'un autre côté, les comportements individuels, adoptés par la population française, sont en constante évolution. Chez les femmes âgées de 15 à 75 ans, une hausse du tabagisme entre 2005 et 2010 a été observée (26,9 % contre 28,7 % de fumeuses) ; hausse particulièrement marquée pour la tranche d'âge 45-64 ans (16,0 % à 22,5 %) (297). Le surpoids et, de façon plus prononcée, l'obésité sont en constante augmentation en France : 23,3 % et 8,3 % de femmes étaient respectivement en surpoids et obèses en 1997 contre 26,3 % et 15,7 % en 2012 (298). Toutefois, malgré une progression de l'obésité depuis 1997, celle-ci semblerait ralentir ces dernières années (298). En ce qui concerne la consommation d'alcool en France, elle régresse régulièrement depuis les années 1950 en raison, notamment, d'une désaffection pour le vin de table au profit de produits de plus grande qualité (299). En 2005, 7,7 % des femmes déclaraient consommer quotidiennement de l'alcool, contre 5,6 % en 2010 (299). La consommation de fruits et légumes, en revanche, est stable entre 1995 (186,4 kg/personne/an) et 2008 (186,7 kg/personne/an), après avoir connu une forte augmentation depuis les années 1950 (100,3 kg/personne/an) (300). Cependant, la consommation de fruits et légumes frais semble être en légère diminution, au profit de la consommation de fruits et légumes transformés (300). Enfin, les évaluations sur les pratiques d'activité physique montrent une baisse de participation au sport et une plus grande sédentarisation de la population, accentuant par là même le phénomène du

surpoids et de l'obésité (301). Ainsi, les comportements des Français et des Françaises concernant ces cinq facteurs de mode de vie peuvent encore être améliorés.

Nos résultats suggèrent l'importance des recommandations de santé publique, car il a été observé que leur adhésion conjointe améliore nettement la prévention du cancer et l'espérance de vie. Il est important, néanmoins, d'améliorer la communication autour de ces recommandations et de cibler les recommandations afin que la population ne soit pas noyée par l'information et qu'elle ait les outils nécessaires pour comprendre d'où proviennent ces recommandations et pourquoi elles sont pertinentes. L'amélioration de la communication est d'autant plus importante que les messages de santé publique se confrontent à ceux du marketing, très puissants. Une évaluation de l'impact des recommandations sur la population serait également nécessaire afin de comprendre comment les individus réagissent face aux messages de santé publique, comment ils les perçoivent et les interprètent. Une étude de ce genre permettrait d'évaluer dans quelles mesures les individus modifient leurs comportements selon les informations que les organismes de santé publique communiquent et la durée nécessaire à la population pour prendre conscience de la nocivité de certains comportements.

Ce travail apporte des preuves supplémentaires quant à la pertinence des recommandations du Programme National Nutrition Santé, lancé en France en 2001, qui encourage la population générale à adopter une alimentation équilibrée et diversifiée, et à pratiquer une activité physique régulière. Les résultats obtenus permettent également d'insister sur le fait qu'adopter un meilleur mode de vie a des répercussions bénéfiques sur la santé à tout âge. Il n'est jamais trop tard pour abandonner ses mauvaises habitudes de vie.

GLOSSAIRE

Acétaldéhyde : Composé chimique dérivant de l'alcool éthylique par oxydation ménagée ou de l'acétylène par hydratation.

Activité mitotique : Correspond au nombre de cellules en division et s'estime en périphérie de la tumeur après avoir parcouru plusieurs champs, en ne tenant compte que du champ le plus riche en mitose.

Aliquoter : Diviser en plusieurs échantillons.

Aménorrhée : Absence de règles.

Amiante [Du vieux français Abeste] : Fibre minérale naturelle massivement utilisée pendant plus d'un siècle dans des produits à destination industrielle ou domestique, en raison de ses performances techniques remarquables associées à un faible coût. Des expositions courtes et répétées à l'amiante peuvent provoquer de graves maladies respiratoires, se déclarant en moyenne 20 à 40 ans après le début de l'exposition.

Analyse en composantes principales : Technique statistique permettant de générer, à partir d'un ensemble d'information provenant d'un grand nombre de variables, de nouvelles variables non mesurées, appelées composantes principales ou axes principaux.

Androgène : Hormones stéroïdes mâles sécrétées par les testicules, les glandes surrénales et, de façon résiduelle, les ovaires.

Androstènedione : Androgène produit par les glandes surrénales et les gonades.

Angiogenèse : Formation de nouveaux vaisseaux sanguins (néo vascularisation) à partir des vaisseaux existants.

Antioxydants : Substances issues de l'alimentation (fruits et légumes principalement) dont les propriétés chimiques permettent de neutraliser les effets délétères des radicaux libres. Elles peuvent se présenter sous forme de vitamines (A, C, E) ou de polyphénols (isoflavones, lycopène).

Apoptose (ou mort cellulaire programmée) : Processus normal d'autodestruction cellulaire en réponse à un signal interne (déclenché par le vieillissement ou l'altération de la cellule) programmé génétiquement par l'organisme.

Aromatase : Enzyme permettant de transformer en œstrogènes les androgènes sécrétés par la surrénale et, de façon résiduelle, les ovaires (voir également enzyme).

Arsenic : Corps simple d'apparence métallique, très répandu dans les minerais métalliques. Il est employé dans la métallurgie (alliages) et en électronique (fabrication des semi-conducteurs). Les dérivés arsenicaux (dont le poison du même nom, l'anhydride arsénieux) sont utilisés dans les tanneries, dans la fabrication de peinture, de fleurs artificielles, de papiers peints, la coloration des verres ou en agriculture (raticide).

Atrophie : Anomalie de développement se caractérisant par une réduction de la taille.

Atypie : Anomalie, absence de conformisme relativement à une référence.

Benzène : Également appelé benzol. Hydrocarbure insaturé de molécule cyclique (hydrocarbure monocyclique) se présentant sous forme d'un liquide incolore, très mobile, volatil, d'odeur caractéristique très pénétrante. Solvant utilisé, par exemple, pour le dégraissage, la préparation des vernis, l'industrie des matières colorantes, des parfums. Il est facilement inflammable et toxique.

Benzo[a]pyrène : Hydrocarbure aromatique polynucléaire cancérigène se rencontrant dans le goudron de houille. Substance omniprésente dans l'environnement se formant au cours de la combustion des matières organiques.

Béryllium : Métal blanc, argenté, brillant, relativement mou, léger que l'on trouve en particulier dans le béryl (silicate d'aluminium et de béryllium). Utilisé en alliage (associé au cuivre et au nickel) il entraîne une excellente conductivité thermique et électrique. Ses composés, utilisés dans de nombreux secteurs d'activité tels que la métallurgie et l'aéronautique, sont très toxiques.

Bêta-carotène : Forme de carotène la plus répandue et précurseur de la vitamine A. Il est fourni par certains légumes et fruits (dont les carottes, les abricots, les mangues, les légumes verts foncés, les patates douces, le persil).

BRCA1, 2 : Gènes impliqués dans les formes héréditaires de cancer du sein et des ovaires.

Cadmium : Métal argenté naturel présent dans certains minerais sous forme d'impuretés et utilisé pour ses caractéristiques physico-chimiques.

Cancer : Maladie provoquée par la transformation de cellules qui deviennent anormales et prolifèrent de façon excessive. Ces cellules dérégées finissent par former une masse qu'on appelle tumeur maligne. Les cellules cancéreuses ont tendance à envahir les tissus voisins et à se détacher

de la tumeur. Elles migrent alors par les vaisseaux sanguins et les vaisseaux lymphatiques pour aller former une autre tumeur (métastase).

Cancérogène : Qui peut provoquer le développement d'un cancer.

Cancérogenèse : Ensemble de phénomènes qui conduisent à la transformation d'un tissu normal en tissu cancéreux.

Carcinome épidermoïde : Cancer qui se développe à partir d'une cellule épithéliale.

Carcinome basocellulaire : Cancer qui se développe au niveau de kératinocytes situés dans la partie profonde de l'épiderme.

Cas-témoins [Étude] : Étude observationnelle rétrospective consistant à recruter d'une part des personnes présentant l'événement de santé d'intérêt (les cas) et d'autre part un groupe de personnes comparables ne présentant pas l'événement de santé (les témoins).

Cellule cancéreuse différenciée : Cellule ressemblant aux cellules normales et comportant de la même manière dans le tissu où elle s'est développée.

Censure : En statistique, une censure survient lorsque l'événement d'intérêt n'est pas complètement observé. On parle de censure à gauche lorsque l'on ne sait pas ce qui s'est passé avant une date précise, et de censure à droite lorsque l'on ne sait pas ce qui s'est passé après une date précise.

Chrome : Oligo-élément existant sous plusieurs formes, ayant chacune des propriétés très différentes. En tant que métal élémentaire, il est stable et non toxique. En revanche, le chrome hexavalent (ou chrome VI) est le membre le plus toxique de la famille du chrome et peut causer des maladies diverses et fait l'objet d'une surveillance étroite.

Cohorte [Enquête de] : Enquête visant à observer la survenue d'événements de santé dans le temps au sein d'une population définie qui est incluse au démarrage de l'étude et suivie sur plusieurs années.

Composés phytochimiques : Appelés aussi phyto-nutriments, ce sont des substances chimiques présentes à l'état naturel dans certaines plantes et dans certains fruits et légumes et pouvant avoir des effets bénéfiques sur la santé, notamment grâce à leurs propriétés antioxydantes.

Corrélation [statistique] : Liaison entre deux caractères dans le cas d'une corrélation simple, ou plus de deux dans le cas d'une corrélation multiple, telle que les variations de leurs valeurs soient toujours de même sens (corrélation positive) ou de sens opposés (corrélation négative).

Corrélation de Pearson (coefficient de) : Formule mathématique permettant d'estimer la corrélation linéaire entre deux variables aléatoires, fonction de la variance de chaque variable et de la covariance entre les deux variables.

Couche leucoplaquettaire : Appelée aussi couche leucocytaire. Couche visualisable après centrifugation du sang, située entre la couche de plasma et les érythrocytes, et contenant les leucocytes (ou globules blancs) et les plaquettes.

Degré d'anisonucléose : Estimé en rapportant le pourcentage de noyaux réguliers au pourcentage de noyaux irréguliers, voire monstrueux.

Diabète : Trouble de l'assimilation, de l'utilisation et du stockage des sucres apportés par l'alimentation, se traduisant par un taux de glucose dans le sang élevé. On distingue principalement deux types de diabètes.

Diabète de type I : Le diabète de type I (anciennement appelé diabète insulino-dépendant) est habituellement découvert chez les jeunes enfants, adolescents ou jeunes adultes et se traite par injections d'insuline.

Diabète de type II : Le diabète de type II (anciennement appelé diabète non-insulino-dépendant ou gras) apparaît généralement chez le sujet de plus de 40 ans, et est favorisé par une augmentation de la masse grasse abdominale, et donc le surpoids, l'obésité et le manque d'activité physique.

Diacétyle : Dicétone, liquide jaune se trouvant dans des huiles essentielles et dans le beurre, auquel il donne son odeur.

Différenciation cellulaire : Processus conduisant à la spécialisation des cellules qui acquièrent leurs propriétés fonctionnelles.

Dydrogestérone : Hormone dérivée de la progestérone et appartenant à la catégorie des progestatifs.

Dysplasie : Stade précancéreux correspondant à une prolifération de cellules épithéliales anormales et d'apparence variable. Les interactions entre les cellules sont diminuées et l'architecture du tissu désorganisée.

Dysurie : Provoquée par l'existence d'un obstacle à l'évacuation des urines.

Endogène : Produit par la structure elle-même, en dehors de tout apport extérieur (voir également hormone endogène).

Endomètre : Muqueuse tapissant la face interne de l'utérus. Le cancer de l'endomètre est de loin le plus fréquent des cancers du corps utérin, du fait de la rareté des cancers développés à partir du muscle utérin (myomètre).

Enzyme : Protéine fabriquée par l'organisme permettant l'activation ou l'accélération de réactions chimiques.

Épithélium : Tissu formé de cellules jointives, juxtaposées, solidaires les unes des autres par des systèmes de jonctions intercellulaires et séparées du tissu conjonctif sous-jacent par une lame basale. Ce tissu peut être aussi bien un revêtement externe, sur la surface de l'épiderme, qu'interne, sur des muqueuses ou organes.

Érythrocytes : Appelés également hématie ou globule rouge. Cellule sanguine, dépourvue de noyau, la plus abondante du tissu sanguin (voir également hématie).

et al. [Du latin *et alii*] : Signifie « et les autres » et utilisé pour tronquer une liste de personnes.

Exogène : Extérieur à un système (voir également hormone exogène).

Fibres alimentaires : Glucides non digestibles provenant de différentes sources alimentaires (céréales complètes, légumes, légumineuses, fruits, graines oléagineuses), indispensables au bon fonctionnement du transit intestinal.

Folates : Appelés aussi acide folique ou vitamine B9, les folates sont contenues dans de nombreux aliments (foie, épinards, etc.), sont nécessaires à la synthèse de l'ADN et sont prescrits dans certaines anémies.

Formaldéhyde : Connu également sous le nom de formol lorsqu'il est dissous dans l'eau, c'est un composé organique très volatil appartenant à la famille des aldéhydes. Il se trouve dans les sources de combustion (dont la fumée de tabac, les bougies, les bâtonnets d'encens), les produits de construction et de décoration (dont la colle urée-formol), les produits d'usage courant (dont les détergents, les désinfectants, les insecticides, le vernis à ongles) et la réactivité chimique entre l'ozone et certains composés organiques volatils présents dans l'air.

Goitre : Hypertrophie de la glande thyroïde.

Hazard Ratio : Rapport du risque instantané de survenue d'une maladie entre deux groupes.

Helicobacter Pylori : Bactérie qui vit exclusivement chez l'être humain, au niveau de son estomac.

Hématie : Appelée également globule rouge ou érythrocyte, il s'agit d'une cellule biconcave en forme de disque écrasé formée dans la moelle osseuse. L'hématie contient de l'hémoglobine, pigment qui lui donne sa couleur rouge, et sa fonction essentielle est le transport de l'oxygène dans le sang.

Hormone endogène : Hormone présente naturellement dans l'organisme.

Hormone exogène : Hormone de synthèse, non fabriquée par l'organisme.

Hormone stéroïdienne : Substance dérivée du cholestérol et sécrétée par certaines glandes endocrines : glandes corticosurrénales (essentiellement le cortisol, l'aldostérone et les androgènes surrénaliens), placenta (œstrogènes et progestérone), ovaires (œstrogène, progestérone, et en quantité moindre androgènes), et testicules (essentiellement la testostérone).

Hyperplasie : Anomalie de développement se caractérisant par une augmentation du nombre de cellules en gardant la même morphologie que les cellules mères.

Hypertrophie : Anomalie de développement se caractérisant par une augmentation de la taille.

Hystérectomie : Ablation de l'utérus.

i.e. [Du latin *id est*] : Synonyme de « c'est-à-dire ».

Immunochistochimie : Désigne la méthode de localisation des protéines situées dans les cellules d'un tissu en utilisant les anticorps pour détecter les antigènes.

In situ [cancer] : Un cancer est dit *in situ* lorsque les cellules cancéreuses n'ont pas encore franchi la membrane basale.

Invasif [cancer] : Un cancer est dit invasif lorsque les cellules cancéreuses ont envahi les tissus sains voisins, en franchissant la membrane basale.

Lésion précancéreuse : Altération d'un tissu ou d'un organe, due à un traumatisme ou à une maladie, qui contient des cellules anormales pouvant un jour donner naissance à une cellule cancéreuse.

Leucémie : Cancer caractérisé par la prolifération incontrôlée dans la moelle osseuse de cellules à l'origine des globules blancs du sang. La fabrication en trop grand nombre de globules blancs contribue à affaiblir le système immunitaire, ceux-ci n'ayant pas le temps d'arriver à maturité.

Luminal : Les cancers du sein du type luminal sont les cancers présentant des récepteurs à l'œstrogène (RO+) et/ou des récepteurs à la progestérone (RP+). Les cancers luminaux peuvent être différenciés en deux catégories : le cancer luminal A est une tumeur sans surexpression de la protéine HER2 et avec un niveau de prolifération faible, le cancer luminal B est une tumeur sans surexpression de la protéine HER2 et avec un niveau de prolifération élevée, ou une tumeur avec surexpression de la protéine HER2, quel que soit son niveau de prolifération. Le niveau de prolifération est évalué par l'indice Ki-67.

Lymph : Liquide transportant les globules blancs permettant l'évacuation des déchets de l'organisme.

Lymphome : Cancer du système lymphatique qui se développe aux dépens des lymphocytes, cellules jouant un rôle essentiel dans les réactions de défense immunitaire. Le lymphome hodgkinien est marqué par la présence, dans l'organe touché, d'une cellule tumorale particulière, appelée « cellule de Sternberg ».

Maladie de Crohn : Maladie inflammatoire chronique du système digestif, qui évolue par crises et phases de rémission.

Non-médullaire [cancer non médullaire de la thyroïde] : Cancer de la thyroïde qui ne se développent pas à partir des cellules parafolliculaires (ou cellules C) de la thyroïde.

Ménarche : Nom donné à l'apparition des premières règles.

Ménopause : Interruption physiologique des cycles menstruels, due à la cessation de la sécrétion des ovaires.

Mésothéliome : Cancer virulent du mésothélium. Chaque organe du corps humain est recouvert de mésothélium qui forme une membrane protectrice (appelée plèvre au niveau des poumons, péritoine au niveau des viscères abdominaux, péricarde au niveau du cœur).

Méta-analyse : Démarche statistique combinant les résultats d'une série d'études indépendantes sur un problème donné.

Métabolisation : Transformation biochimique d'une substance dans un organisme vivant au cours du métabolisme (ensemble des réactions chimiques).

Métaplasie : Transformation d'un tissu cellulaire en un autre tissu cellulaire de structure et fonction strictement normales.

Métastases : Localisation secondaire d'un cancer atteinte par la propagation des cellules de la localisation primaire *via* la circulation sanguine ou lymphatique.

Microenvironnement : Réseau conjonctif de support d'une cellule, d'un tissu ou d'un organe.

Micronisée : Corps solide réduit en particules ayant des dimensions de l'ordre du micromètre (ou micron).

Mitoses : Processus de division cellulaire qui permet d'obtenir deux cellules filles identiques à partir d'une cellule mère.

Modalité : Variante que peut présenter un caractère statistique qualificatif.

Morbidité : Caractère relatif à une maladie.

Mutation : Apparition brusque, dans tout ou partie des cellules d'un être vivant, d'un changement dans la structure de certains gènes, transmis aux générations suivantes si les gamètes (cellules reproductrices) sont affectés.

Nébulosité : Fraction de la voûte céleste occultée en un lieu et à un instant par l'ensemble des nuages présents au moment de l'évaluation.

Néoplasme : Tissu anormalement formé. Synonyme de tumeur ou pour certains de tumeur maligne (cancer).

Nickel : Métal blanc brillant à cassure fibreuse présentant des analogies avec le fer.

N-Nitrosornicotine : Nitrosamine carcinogène dérivée de la nicotine produite lors de la combustion du tabac.

Œstradiol : Œstrogène le plus actif de l'organisme impliqué dans l'apparition des caractères sexuels secondaires féminins (seins, libido, etc.).

Œstriol : Œstrogène produit lors de la dégradation de l'œstradiol et de l'œstrone, sécrété par le placenta et le fœtus.

Œstrogène : Hormone stéroïdienne, sécrétée principalement en préménopause par les ovaires, responsable du développement pubertaire et du maintien ultérieur des caractères physiques féminins. Il existe trois types d'œstrogènes : l'œstradiol, l'œstrone et l'œstriol (voir également œstradiol, œstrone et œstriol).

Œstrone : Œstrogène pouvant se transformer en œstradiol.

Oncogène : Gène localisé dans un virus ou dans une cellule de l'organisme favorisant la transformation d'une cellule normale en cellule cancéreuse.

Ovariectomie bilatérale : Ablation des deux ovaires.

Plasma : Élément liquide du sang dans laquelle les cellules sanguines (globules rouges, globules blancs et plaquettes) sont en suspension.

Postménopause : Période succédant à la ménopause.

Préménopause : Période précédant la ménopause.

Progestatif : Hormone stéroïdienne d'action similaire à la progestérone.

Progestérone : Hormone stéroïdienne sécrétée principalement par le corps jaune des ovaires pendant la seconde phase du cycle menstruel.

Qualitatif : Une variable est dite qualitative lorsque ses valeurs ne représentent pas une quantité, mais des états.

Quantitatif : Une variable est dite quantitative si elle reflète une notion de grandeur.

Radon : Gaz rare dit noble, radioactif, incolore et inodore. Il est d'origine naturelle et représente le tiers de l'exposition moyenne de la population française aux rayonnements ionisants.

Récréationnelle [Activité physique] : Activité physique pratiquée en loisir.

Rectocolite hémorragique : Maladie auto-immune inflammatoire chronique du côlon et du rectum (appelée aussi colite ulcéreuse).

Risque relatif : Mesure statistique mesurant le risque de survenue d'un événement dans un groupe par rapport à un autre.

Sarcome : Cancer qui se développe au niveau du tissu conjonctif, tissu de connexion des organes.

Sarcome de Kaposi : Cancer provoquant l'apparition de tumeurs violacées ou brunâtres ressemblant à des plaies sur la peau, mais pouvant également s'attaquer aux organes internes ou aux muqueuses de la bouche, du nez et de l'anus.

Sérum : Liquide sanguin débarrassé de la fibrine, une protéine impliquée dans la formation de caillots.

Silice : Minéral dur et transparent qui existe sous différentes formes (roches sédimentaires, roches magmatiques, etc.) dont les particules les plus fines pénètrent par les voies respiratoires jusqu'aux alvéoles pulmonaires et s'y déposent.

Stress oxydatif : Déséquilibre de la balance entre la production de radicaux libres (molécules chimiques synthétisées par des réactions avec l'oxygène) dans l'organisme et les défenses antioxydantes par une insuffisance des défenses ou une production trop intense de radicaux libres. L'activité physique entraîne une consommation plus importante d'oxygène, et donc une production plus importante de radicaux libres.

Stroma : Voir microenvironnement.

Sulfate de déhydroépiandrostérone : Androgène essentiellement fabriqué par les glandes surrénales et semblant impliquée dans le processus de vieillissement.

Syndrome de Li-Fraumeni : Syndrome génétique prédisposant à diverses tumeurs. Environ 70 % des cas de Li-Fraumeni présentent des mutations du gène P53. Les deux autres localisations identifiées sont sur le gène *checkpoint kinase 2* (CHECK2) et sur le bras long du chromosome 1.

Syndrome des ovaires polykystiques [appelé aussi syndrome de Stein-Leventhal] :

Syndrome diagnostiqué par la présence de kystes multiples sur les ovaires et accompagné d'un ensemble de symptômes, notamment l'acné, la difficulté à concevoir (infécondité), des menstruations irrégulières, une prise de poids, une coloration anormale de la peau, une perte de cheveux ou une pilosité excessive ou l'absence d'ovulation.

Système nerveux végétatif : Système nerveux qui règle le fonctionnement des viscères.

Tamoxifène : Médicament à action anti-œstrogène utilisé pour bloquer l'action des œstrogènes sur les cellules mammaires qui contiennent des récepteurs à l'œstrogène dans le cas d'un traitement contre le cancer du sein.

Taux d'incidence : En épidémiologie, le taux d'incidence est le nombre de nouveaux cas d'une pathologie observés dans une population donnée rapporté sur la taille de cette population et la durée de la période d'observation. Il s'exprime généralement en « nombre de personnes pour 100 000 personnes par année ».

Testostérone : Androgène présent en faible quantité chez la femme, produite par les ovaires et les surrénales et responsable du désir sexuel.

Trouble climatérique : Symptômes qui accompagnent les modifications hormonales associées à l'arrêt de la fonction ovarienne lors de la ménopause.

Tumeur : Multiplication excessive de cellules anormales d'un tissu ou d'un organe.

Tumeur bénigne : Conséquence d'une multiplication cellulaire excessive, mais régulée par la tumeur et les tissus environnants : tumeur bien limitée, respectant le tissu environnant, faite de cellules ne présentant pas d'anomalies, qui ne donne pas de métastases et qui ne récidive pas après ablation chirurgicale. Une tumeur bénigne est formée de cellules bien différenciées (*i.e.* ressemblant aux cellules normales).

Tumeur maligne : Conséquence d'une multiplication cellulaire excessive, anarchique et non régulée, tumeur de contours mal délimités, envahissant les tissus environnants, dont les cellules peu différenciées présentent des anomalies microscopiques et métaboliques, donnant des métastases et récidivant souvent après ablation chirurgicale.

Variant de gène : Synonyme de polymorphisme génétique. Existence de plusieurs allèles (formes différentes d'un même gène) dans une population, résultant des mutations génétiques. On parle de polymorphisme, ou variant de gène, lorsque ces formes représentent au moins 1 % de la population.

Vasomoteur [trouble] : Trouble circulatoire dû à un relâchement des vaisseaux ou à leur constriction, en rapport avec des troubles fonctionnels du système nerveux végétatif (voir également système nerveux végétatif).

ANNEXES

Annexe A : Groupes alimentaires considérés dans l'analyse en composantes principales pour la création des profils alimentaires

Tableau 39 : Groupes alimentaires considérés dans l'analyse en composantes principales pour la création des profils alimentaires. [Lien vers la première citation du tableau dans le texte]

Groupes alimentaires	Aliments
Viandes transformées	Saucisson, andouille, pâté, rillettes, saucisse, boudin, andouillette
Volaille et lapin	Volaille, lapin
Poisson	Sardine fraîche, colin, merlu, julienne, maquereau frais, limande, aiglefin, sole, lieu, merlan, cabillaud, saumon frais, truite, saumonette
Viande	Cheval, veau, mouton, porc, bœuf
Abats	Foie, langue, cervelle, cœur, rognons, tripes
Œuf	Œuf dur, froid, au plat, en omelette, à la coque
Jambon	Jambon cuit, jambon cru
Légume sec	Lentille en salade, haricot sec, lentille cuisinée
Légume cru	Endive en salade, poireau en salade, artichaut, chou-fleur en salade, chou blanc ou rouge cru, carotte râpée, radis, betterave, céleri-rave, tomate en salade, avocat, concombre, salade verte
Légume cuit	Maïs en salade, chou de Bruxelles, endive cuite, épinards, champignon, chou-fleur cuit, chou cuit, carotte cuite, céleri en branche, bette, cardon, fenouil cuit, tomate cuite, haricot vert, salsifis, aubergine, poivron cuit, courgette, petits pois
Fruit	Orange, pamplemousse, mandarine, pomme, poire, banane, kiwi, ananas, fraise, framboise, cerise, pêche, brugnion, melon, abricot, raisin, prune
Pâtisserie	Croissant, brioche
Gâteau	Quatre-quarts, génoise, pâtisserie à la crème, tarte aux fruits
Entremet	Crème, dessert lacté, glace, entremet à base de riz, de semoule
Chocolat	Barre chocolatée, chocolat en tablette
Pain blanc	Pain blanc, baguette, ficelle, pain de mie
Pain complet	Pain de champagne, pain Poilane, pain de campagne industriel, pain de campagne rond, pain complet
Biscotte	Pain azyne, cracotte, pain grille, pain croustillant, biscotte, petit pain grillé
Céréale du petit-déjeuner	Céréale sucrée, céréale non sucrée
Margarine	Margarine ordinaire, margarine allégée, margarine en pâte à tartiner
Beurre, crème	Beurre (issu du beurre ordinaire, 1/2 sel, salé, beurre allégé, crème fraîche après cuisson, crème fraîche allégée après cuisson)
Huile d'olive	Huile d'olive (pure ou en vinaigrette maison)
Huile de tournesol	Huile de tournesol (pure ou en vinaigrette maison)
Huile végétale	Huile de maïs, d'arachide, de colza, de soja (pure ou en vinaigrette maison)
Autres matières grasses	Graisse de canard, d'oie, matière grasse pour la cuisson
Mayonnaise	Mayonnaise après cuisson
Vinaigrette non allégée	Vinaigrette du commerce ordinaire
Vinaigrette allégée	Vinaigrette du commerce allégée

Suite page suivante

ANNEXES

Groupes alimentaires	Aliments
Fromage	Camembert, Saint Marcelin, Brie, Munster, Vacherin, Pont-l'évêque, Vache qui rit, bleus, Roquefort, Cantal, Gouda, Port-Salut, Babybel, tommes, Morbier, St Nectaire, Gruyère, Comté, fromage de chèvre
Produit laitier allégé	Lait demi-écrémé, écrémé (pur ou ajouté, concentré ou en poudre), yaourt nature maigre, yaourt aromatisé, aux fruits maigre, aux fruits allégé (sucré à l'aspartam), fromage blanc nature 0% MG, aromatisé, aux fruits 0% MG
Produit laitier non allégé	Lait entier (pur ou ajouté), fromage blanc nature 10-20% MG, fromage blanc nature 30-40% MG, fromage blanc aromatisé, aux fruits 10-20% MG, fromage blanc aromatisé, aux fruits 30-40% MG-, yaourt nature ordinaire, yaourt nature déjà sucré, yaourt aromatisé, aux fruits ordinaire
Alcool faible	Bière standard, bière "de luxe", bière sans alcool, cidre
Alcool fort	Muscat, porto, vermouth, whisky, gin, vodka, Ricard, pastis, cocktail, punch, kir, cognac, calvados, rhum
Vin	Vin rouge, vin blanc, vin rosé, champagne
Frite	Frite
Pomme de terre	Pommes de terre en salade, pommes de terre, purée (frites exclues)
Féculent	Riz ordinaire, riz complet, pâtes, semoule
Pizza, quiche, tarte salée, crêpe salée	Contenant huile végétale, tomates, mozzarella, jambon cuit, champignon, pâte à pizza, lardons fumés, œufs, crème fraîche, gruyère, poireaux, oignons, crêpe, sauce béchamel, pâte brisée
Sandwich, hamburger, croque-monsieur	Contenant pain à hamburger, bœuf à hamburger 15%, ketchup, oignons, pain blanc, jambon cuit, beurre, pain de mie blanc, gruyère, sauce béchamel
Biscuits apéritifs	Olive, biscuit salé, fruits oléagineux (cacahuètes, amandes, noisettes, pistaches, noix)
Thé	Thé
Soupe	Potage (fait maison ou commercial)
Jus de fruits	Fruits pressés ou pur jus de fruits du commerce
Café	Café
Eau du robinet	Eau du robinet
Eau en bouteille	Eau minérale, eau de source en bouteille
Sucre ajouté	Confiture, miel, sucre ajouté au café, thé, chicorée, yaourt et fromage blanc
Édulcorant	Édulcorant ajouté au café, thé, chicorée, yaourt et fromage blanc
Poisson en conserve	Sardine, thon, maquereau en conserve
Fruit de mer	Moule, coquillage, crustacé
Boisson non allégée	Boisson fruitée, cola, soda normal
Boisson allégée	Boisson fruitée, cola, soda allégé en sucre
Biscuit sec	Cookies, biscuit recouvert d'une couche de chocolat
Chicorée	Ricoré, Nescoré
Fruit en conserve	Fruit en compote, au sirop
Boisson chocolatée	Boisson chocolatée
Bonbon	Bonbon, caramel, chewing-gum, réglisse

Annexe B : Jeu de données simulées pour l'illustration des performances d'un modèle prédictif.

Tableau 40 : Exemple de calcul de la sensibilité et de la spécificité à partir d'un jeu de données simulées.
[Lien vers la première citation du tableau dans le texte]

Observation	Probabilité de survenue de la maladie	Vrais positifs	Faux positifs	Faux négatifs	Vrais négatifs	1-Spécificité	Sensibilité
Malade	0,465	1	0	8	23	0,000	0,111
Malade	0,463	2	0	7	23	0,000	0,222
Non-Malade	0,447	2	1	7	22	0,043	0,222
Non-Malade	0,446	2	2	7	21	0,087	0,222
Malade	0,443	3	2	6	21	0,087	0,333
Malade	0,397	4	2	5	21	0,087	0,444
Non-Malade	0,386	4	3	5	20	0,130	0,444
Non-Malade	0,366	4	4	5	19	0,174	0,444
Non-Malade	0,359	4	5	5	18	0,217	0,444
Malade	0,350	5	5	4	18	0,217	0,556
Non-Malade	0,328	5	6	4	17	0,261	0,556
Non-Malade	0,327	5	7	4	16	0,304	0,556
Malade	0,304	6	7	3	16	0,304	0,667
Non-Malade	0,287	6	8	3	15	0,348	0,667
Non-Malade	0,256	6	9	3	14	0,391	0,667
Non-Malade	0,236	6	10	3	13	0,435	0,667
Malade	0,222	7	10	2	13	0,435	0,778
Non-Malade	0,191	7	11	2	12	0,478	0,778
Non-Malade	0,169	7	12	2	11	0,522	0,778
Non-Malade	0,167	7	13	2	10	0,565	0,778
Malade	0,154	8	13	1	10	0,565	0,889
Non-Malade	0,140	8	14	1	9	0,609	0,889
Non-Malade	0,123	8	15	1	8	0,652	0,889
Non-Malade	0,117	8	16	1	7	0,696	0,889
Non-Malade	0,113	8	17	1	6	0,739	0,889
Non-Malade	0,110	8	18	1	5	0,783	0,889
Non-Malade	0,092	8	19	1	4	0,826	0,889
Non-Malade	0,086	8	20	1	3	0,870	0,889
Malade	0,075	9	20	0	3	0,870	1,000
Non-Malade	0,073	9	21	0	2	0,913	1,000
Non-Malade	0,025	9	21	0	1	0,913	1,000
Non-Malade	0,016	9	23	0	0	1,000	1,000

Annexe C : Caractéristiques des populations d'analyse

Tableau 41 : Caractéristiques, au début de l'étude sur l'association entre l'indice de santé et le risque de cancer, de la population d'analyse, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]

	Population entière (N = 64 732, 100,00 %)	Catégories de l'indice de santé			
		[0 ; 2] (N = 3 859, 5,96 %)	[2,5 ; 3] (N = 14 478, 22,37 %)	[3,5 ; 4] (N = 27 488, 42,46 %)	[4,5 ; 5] (N = 18 907, 29,21 %)
Variabiles d'ajustement					
Âge [moyenne (écart-type)]	52,70 (6,62)	51,43 (6,15)	52,06 (6,44)	52,78 (6,64)	53,32 (6,73)
Niveau d'éducation [n (%)]					
Inférieur à Bac	7 091 (10,95)	408 (10,57)	1 560 (10,78)	2 972 (10,81)	2 151 (11,38)
Bac-Bac+2	34 436 (53,20)	1 920 (49,76)	7 439 (51,38)	14 622 (53,19)	10 455 (55,30)
Supérieur à Bac+2	23 205 (35,85)	1 531 (39,67)	5 479 (37,84)	9 894 (36,00)	6 301 (33,32)
Zone de résidence [n (%)]					
Nord	29 876 (46,15)	2 116 (54,83)	7 195 (49,70)	12 602 (45,85)	7 963 (42,12)
Centre	15 794 (24,40)	801 (20,76)	3 321 (22,93)	6 742 (24,53)	4 930 (26,07)
Sud	19 062 (29,45)	942 (24,41)	3 962 (27,37)	8 144 (29,62)	6 014 (31,81)
Antécédents familiaux de cancer chez les parents au premier degré ^a [n (%)]					
Au moins un	30 779 (47,55)	1 888 (48,92)	6 964 (48,10)	12 965 (47,17)	8 962 (47,40)
Aucun	33 953 (52,45)	1 971 (51,08)	7 514 (51,90)	14 523 (52,83)	9 945 (52,60)
Activité professionnelle [n (%)]					
En activité	45 140 (69,73)	3 051 (79,06)	10 788 (74,51)	19 052 (69,31)	12 249 (64,79)
Retraitee/N'a jamais travaillé	19 592 (30,27)	808 (20,94)	3 690 (25,49)	8 436 (30,69)	6 658 (35,21)
Utilisation de contraceptif oral [n (%)]					
Déjà	39 646 (61,25)	2 673 (69,27)	9 415 (65,03)	16 836 (61,25)	10 722 (56,71)
Jamais	25 086 (38,75)	1 186 (30,73)	5 063 (34,97)	10 652 (38,75)	8 185 (43,29)
Statut ménopausique et utilisation de THM [n (%)]					
Préménopause	29 162 (45,05)	1 945 (50,40)	7 065 (48,80)	12 285 (44,69)	7 867 (41,61)
Postménopause avec THM récent	13 353 (20,63)	664 (17,21)	2 723 (18,81)	5 758 (20,95)	4 208 (22,26)
Postménopause sans THM récent	22 217 (34,32)	1 250 (32,39)	4 690 (32,39)	9 445 (34,36)	6 832 (36,13)
Âge à la puberté [n (%)]					
Premier quartile < 12	13 087 (20,22)	918 (23,79)	3 169 (21,89)	5 514 (20,06)	3 486 (18,44)
Second quartile [12 ; 13[16 086 (24,84)	938 (24,31)	3 579 (24,72)	6 881 (25,03)	4 688 (24,80)
Troisième quartile [13 ; 14[17 580 (27,16)	998 (25,86)	3 868 (26,72)	7 450 (27,10)	5 264 (27,84)
Quatrième quartile ≥ 14	17 979 (27,78)	1 005 (26,04)	3 862 (26,67)	7 643 (27,81)	5 469 (28,92)
Nombre d'enfants et âge à la première grossesse à terme [n (%)]					
Nullipare	7 559 (11,68)	537 (13,92)	1 832 (12,65)	3 201 (11,65)	1 989 (10,52)
Un enfant avant 30 ans	7 660 (11,84)	499 (12,93)	1 784 (12,32)	3 139 (11,42)	2 238 (11,84)
Un enfant après 30 ans	2 384 (3,68)	187 (4,85)	572 (3,95)	998 (3,63)	627 (3,32)
Plus d'un enfant dont le premier avant 30 ans	44 441 (68,65)	2 489 (64,50)	9 639 (66,58)	18 982 (69,06)	13 331 (70,50)

Suite page suivante

	Population entière	Catégories de l'indice de santé			
		[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Plus d'un enfant dont le premier après 30 ans	2 688 (4,15)	147 (3,80)	651 (4,50)	1168 (4,24)	722 (3,82)
Comportements de mode de vie					
Tabagisme [n (%)]					
Fumeuse	8 824 (13,63)	2 112 (54,73)	4 050 (27,97)	2 662 (9,68)	0 (0,00)
Ex-fumeuse	21 355 (32,99)	1 268 (32,86)	6 064 (41,88)	10 454 (38,03)	3 569 (18,88)
N'a jamais fumé	34 553 (53,38)	479 (12,41)	4 364 (30,15)	14 372 (52,29)	15 338 (81,12)
IMC, kg.m ⁻² [n (%)]					
Moyenne (écart-type)	22,89 (3,30)	24,69 (5,01)	23,63 (3,99)	22,82 (3,06)	22,08 (2,15)
< 16	85 (0,13)	19 (0,49)	34 (0,23)	32 (0,12)	0 (0,00)
[16 ; 18,5[2 280 (3,52)	275 (7,13)	687 (4,74)	979 (3,56)	339 (1,79)
[18,5 ; 25[49 210 (76,03)	1 882 (48,77)	9 193 (63,50)	20 994 (76,38)	17 141 (90,66)
[25 ; 30[10 760 (16,62)	1 059 (27,44)	3 483 (24,06)	4 791 (17,42)	1 427 (7,55)
≥ 30	2 397 (3,70)	624 (16,17)	1081 (7,47)	692 (2,52)	0 (0,00)
Consommation d'alcool [n (%)]					
Moyenne (écart-type)	11,65 (14,05)	28,43 (19,89)	18,55 (16,59)	10,70 (12,16)	4,33 (4,51)
≤ 1 verre ^b /jour	38 664 (59,73)	520 (13,47)	4 976 (34,37)	16 274 (59,20)	16 894 (89,35)
]1 ; 2] verres/jour	13 237 (20,45)	737 (19,10)	3 684 (25,44)	6 803 (24,75)	2 013 (10,65)
> 2 verres/jour	12 831 (19,82)	2 602 (67,43)	5 818 (40,19)	4 411 (16,05)	0 (0,00)
Consommation de fruits et légumes [n (%)]					
Moyenne (écart-type)	528,65 (243,38)	315,32 (187,26)	440,29 (236,34)	538,12 (234,56)	626,09 (219,07)
Au moins 5 portions ^c par jour (≥ 400 g/jour)	43 951 (67,90)	760 (19,69)	6 625 (45,76)	19 213 (69,90)	17 353 (91,78)
[3,5 ; 5] portions par jour	12 249 (18,92)	926 (24,00)	3 986 (27,53)	5 783 (21,03)	1 554 (8,22)
Moins de 3,5 portions par jour	8 532 (13,18)	2 173 (56,31)	3 867 (26,71)	2 492 (9,07)	0 (0,00)
Niveau d'activité physique récréationnelle, METs-h/semaine [n (%)]					
Moyenne (écart-type)	29,42 (34,87)	13,64 (20,85)	20,90 (27,89)	28,92 (34,22)	39,90 (39,50)
< 10	12 959 (20,02)	2 329 (60,36)	5 506 (38,03)	5 124 (18,64)	0 (0,00)
[10 ; 20[16 740 (25,86)	968 (25,08)	4 425 (30,56)	8 228 (29,93)	3 119 (16,50)
≥ 20	35 033 (54,12)	562 (14,56)	4 547 (31,41)	14 136 (51,43)	15 788 (83,50)

Note : Les proportions observées pour chaque modalité des variables qualitatives qui ne diffèrent pas statistiquement au seuil de 5 % entre les femmes ayant un indice compris entre 0 et 2 et les femmes ayant un indice compris entre 4,5 et 5 sont : les proportions des femmes ayant eu un enfant avant 30 ans (p-valeur = 0,057), ayant eu plusieurs enfants dont le premier avant 30 ans (p-valeur = 0,978), ayant un niveau d'étude inférieur au Bac (p-valeur = 0,150), ayant au moins un antécédent familial de cancer du sein chez les parents au premier degré (p-valeur = 0,084) et ayant un âge à la puberté compris entre 12 et 13 (p-valeur = 0,522). Les moyennes observées pour chaque variable quantitative entre les femmes ayant un indice compris entre 0 et 2 et les femmes ayant un indice compris entre 4,5 et 5 sont toutes significativement différentes au seuil de 5 % (p-valeur < 0,0001).

a : Les parents au premier degré incluent la mère, les sœurs et les filles.

b : Un verre d'alcool correspond à 10 grammes d'éthanol.

c : Une portion de fruits et légumes correspond à 80 grammes de fruits et légumes, qui incluent les légumes crus et cuits, la salade et les fruits frais.

Tableau 42 : Caractéristiques, au début de l'étude sur le poids relatif des facteurs de risque de cancer du sein en préménopause et en postménopause, de la population d'analyse et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]

	Non-cas N = 63 999	Cas de cancer du sein en préménopause N = 497	HR [IC 95 %]	Cas de cancer du sein en postménopause N = 3 138	HR [IC 95 %]
Facteurs de risque non-comportementaux					
Âge (années) ^a					
Moyenne (écart-type)	52,75 (6,66)	46,77 (2,51)	-	54,01 (6,16)	-
Antécédents familiaux de cancer du sein chez les parents au premier degré ^b					
Au moins un	4 587 (7,17)	46 (9,25)	1,37 [1,01 ; 1,86]	345 (10,99)	1,50 [1,34 ; 1,69]
Aucun	49 746 (77,73)	386 (77,67)	1,00	2 321 (73,97)	1,00
Valeur manquante	9 666 (15,10)	65 (13,08)	1,11 [0,85 ; 1,45]	472 (15,04)	0,99 [0,90 ; 1,10]
Niveau d'éducation					
Inférieur au Bac	7 192 (11,24)	33 (6,64)	1,00	340 (10,83)	1,00
Bac-Bac+2	34 019 (53,15)	251 (50,50)	1,15 [0,80 ; 1,66]	1 642 (52,33)	1,00 [0,89 ; 1,13]
Supérieur à Bac+2	22 788 (35,61)	213 (42,86)	1,05 [0,72 ; 1,53]	1 156 (36,84)	1,07 [0,94 ; 1,21]
Taille à l'âge adulte (cm)					
< 160	21 922 (34,25)	138 (27,77)	1,00	1 036 (33,02)	1,00
[160 ; 165[22 394 (34,99)	177 (35,61)	1,06 [0,85 ; 1,33]	1 089 (34,70)	1,04 [0,95 ; 1,13]
[165 ; 170[13 955 (21,81)	119 (23,94)	1,04 [0,81 ; 1,33]	702 (22,37)	1,09 [0,99 ; 1,21]
≥ 170	5 728 (8,95)	63 (12,68)	1,18 [0,87 ; 1,61]	311 (9,91)	1,20 [1,05 ; 1,37]
Antécédents personnels de maladie bénigne du sein ^c					
Au moins un, avec au moins une biopsie pratiquée	1 556 (2,43)	20 (4,02)	2,02 [1,37 ; 2,99]	131 (4,17)	1,75 [1,52 ; 2,02]
Au moins un, aucune biopsie pratiquée	16 860 (26,34)	177 (35,61)	1,42 [1,18 ; 1,71]	1 054 (33,59)	1,42 [1,18 ; 1,71]
Aucun	45 583 (71,23)	300 (60,37)	1,00	1 953 (62,24)	1,00
Âge à la puberté (années)					
< 10	205 (0,32)	2 (0,40)	1,43 [0,35 ; 5,81]	13 (0,42)	1,58 [0,91 ; 2,74]
[10 ; 12[12 758 (19,93)	98 (19,72)	1,26 [0,95 ; 1,66]	649 (20,68)	1,19 [1,07 ; 1,32]
[12 ; 14[33 177 (51,84)	291 (58,55)	1,36 [1,09 ; 1,70]	1 651 (52,61)	1,13 [1,04 ; 1,23]
≥ 14	17 859 (27,91)	106 (21,33)	1,00	825 (26,29)	1,00
Poids à la naissance					
Petit	3 466 (5,41)	14 (2,82)	1,00	154 (4,91)	1,00
Normal	46 189 (72,17)	360 (72,43)	1,72 [1,01 ; 2,95]	2 321 (73,96)	1,13 [0,96 ; 1,33]
Gros	3 531 (5,52)	31 (6,24)	1,99 [1,05 ; 3,76]	163 (5,19)	1,03 [0,82 ; 1,29]
Valeur manquante	10 813 (16,90)	92 (18,51)	2,32 [1,32 ; 4,08]	500 (15,94)	1,34 [1,11 ; 1,60]
Âge à la ménopause (années, pour les femmes ménopausées uniquement) ^d					
< 48	10 426 (16,29)		-	489 (15,58)	1,00
[48 ; 50[10 275 (16,06)		-	548 (17,46)	1,09 [0,97 ; 1,23]
[50 ; 52[18 587 (29,04)		-	993 (31,65)	1,11 [1,00 ; 1,24]
[52 ; 54[13 251 (20,71)		-	623 (19,85)	1,03 [0,92 ; 1,16]
≥ 54	9 673 (15,11)		-	485 (15,46)	1,17 [1,03 ; 1,34]

Suite page suivante

	Non-cas	Cas de cancer du sein en préménopause		Cas de cancer du sein en postménopause	
Facteurs de risque comportementaux					
Tabagisme actif					
Fumeuse	8747 (13,67)	74 (14,89)	0,96 [0,71 ; 1,28]	395 (12,59)	0,99 [0,88 ; 1,13]
Ex-fumeuse	34 257 (53,53)	229 (46,08)	1,15 [0,95 ; 1,40]	1 684 (53,66)	1,01 [0,94 ; 1,09]
N'a jamais fumé	20 995 (32,80)	194 (39,03)	1,00	1 059 (33,75)	1,00
Nombre d'enfants et âge à la première grossesse à terme					
Un enfant avant 30 ans	7 115 (11,12)	51 (10,26)	0,99 [0,73 ; 1,34]	325 (10,36)	0,99 [0,88 ; 1,12]
Un enfant après 30 ans	2 828 (4,42)	38 (7,65)	1,64 [1,16 ; 2,31]	164 (5,22)	1,29 [1,09 ; 1,51]
Plus d'un enfant dont le premier avant 30 ans	42 829 (66,92)	314 (63,18)	1,00	1 998 (63,67)	1,00
Plus d'un enfant dont le premier après 30 ans	3 847 (6,01)	44 (8,85)	1,44 [1,04 ; 1,98]	213 (6,79)	1,22 [1,06 ; 1,40]
Nullipare	7 380 (11,53)	50 (10,06)	0,97 [0,69 ; 1,35]	438 (13,96)	1,28 [1,13 ; 1,45]
Niveau d'activité physique total (METs-h/semaine)					
< 20	12 951 (20,24)	120 (24,15)	1,06 [0,82 ; 1,38]	618 (19,69)	1,02 [0,92 ; 1,13]
[20 ; 40[21 072 (32,92)	176 (35,41)	1,00 [0,79 ; 1,28]	1 033 (32,92)	1,03 [0,94 ; 1,13]
[40 ; 60[13 652 (21,33)	94 (18,91)	0,91 [0,69 ; 1,20]	688 (21,93)	1,03 [0,93 ; 1,14]
≥ 60	16 324 (25,51)	107 (21,53)	1,00	799 (25,46)	1,00
Silhouette à la puberté ^e					
Silhouette 1	13 465 (21,04)	86 (17,30)	1,08 [0,80 ; 1,45]	751 (23,93)	1,33 [1,19 ; 1,49]
Silhouette 2	22 093 (34,52)	185 (37,22)	1,28 [1,00 ; 1,65]	1 140 (36,33)	1,27 [1,15 ; 1,41]
Silhouette 3	14 922 (23,32)	130 (26,16)	1,22 [0,93 ; 1,59]	704 (22,44)	1,18 [1,06 ; 1,32]
Silhouette ≥ 4	13 519 (21,12)	96 (19,32)	1,00	543 (17,30)	1,00
Durée d'allaitement cumulé (mois)					
Aucun	24 032 (37,55)	187 (37,63)	1,09 [0,82 ; 1,45]	1 231 (39,23)	1,02 [0,91 ; 1,14]
[0 ; 3[17 901 (27,97)	134 (26,96)	0,94 [0,71 ; 1,25]	858 (27,34)	1,07 [0,95 ; 1,19]
[3 ; 6[10 279 (16,06)	95 (19,11)	1,03 [0,76 ; 1,38]	513 (16,35)	1,11 [0,99 ; 1,26]
≥ 6	11 787 (18,42)	81 (16,30)	1,00	536 (17,08)	1,00
Profil alimentaire ^f					
Faibles scores pour les deux profils	15 831 (24,73)	100 (20,12)	1,05 [0,78 ; 1,40]	761 (24,25)	1,10 [0,99 ; 1,22]
Faible score pour le profil sain et score élevé pour le profil occidental	16 158 (25,25)	154 (30,99)	1,05 [0,81 ; 1,37]	813 (25,91)	1,20 [1,08 ; 1,33]
Score élevé pour le profil sain et faible score pour le profil occidental	16 292 (25,46)	90 (18,11)	1,00	743 (23,68)	1,00
Scores élevés pour les deux profils	15 718 (24,56)	153 (30,78)	1,36 [1,04 ; 1,79]	821 (26,16)	1,15 [1,04 ; 1,28]
Consommation d'alcool (g/jour)					
[0 ; 10[38 441 (60,06)	302 (60,76)	1,00	1 750 (55,77)	1,00
[10 ; 20[13 060 (20,41)	106 (21,33)	1,01 [0,81 ; 1,27]	659 (21,00)	1,05 [0,96 ; 1,15]
≥ 20	12 498 (19,53)	89 (17,91)	0,89 [0,69 ; 1,14]	729 (23,23)	1,18 [1,07 ; 1,29]
Vitamine D alimentaire (µg) et exposition aux UV du soleil (kJ/m ² /jour) ^g					
Apport de vitamine D alimentaire et exposition aux UV du soleil élevés	15 329 (23,95)	123 (24,75)	1,00	720 (22,94)	1,00
Apport de vitamine D alimentaire élevé et exposition aux UV du soleil faible	16 631 (25,99)	133 (26,76)	0,93 [0,73 ; 1,20]	881 (28,08)	1,14 [1,03 ; 1,26]
Apport de vitamine D alimentaire faible et exposition aux UV du soleil élevé	17 457 (27,28)	119 (23,94)	0,97 [0,74 ; 1,27]	834 (26,58)	1,05 [0,94 ; 1,16]

Suite page suivante

	Non-cas	Cas de cancer du sein en préménopause		Cas de cancer du sein en postménopause	
Apport de vitamine D alimentaire et exposition aux UV du soleil faibles	14 582 (22,78)	122 (24,55)	1,11 [0,85 ; 1,46]	703 (22,40)	1,06 [0,95 ; 1,18]
Utilisation de contraceptif oral ou de progestatif seul					
Récente ^b	10 763 (16,82)	136 (27,36)	1,07 [0,87 ; 1,33]	549 (17,50)	1,38 [1,18 ; 1,61]
Passée, dont la dernière utilisation remonte à moins de 10 ans	16 147 (25,23)	138 (27,77)	1,09 [0,87 ; 1,37]	870 (27,72)	1,06 [0,97 ; 1,15]
Passée, dont la dernière utilisation remonte à plus de 10 ans	14 489 (22,64)	145 (29,17)	1,00	598 (19,06)	1,00
Jamais	17 132 (26,77)	57 (11,47)		858 (27,34)	
Récence inconnue	5 468 (8,54)	21 (4,23)	0,78 [0,46 ; 1,29]	263 (8,38)	0,95 [0,82 ; 1,09]
IMC (kg.m ⁻²)					
< 18,5	2 234 (3,49)	18 (3,62)	1,03 [0,59 ; 1,80]	84 (2,68)	1,00
[18,5 ; 25[48 383 (75,60)	404 (81,29)	1,15 [0,91 ; 1,45]	2 310 (73,61)	
[25 ; 30[10 994 (17,18)	62 (12,47)	1,00	610 (19,44)	1,19 [1,10 ; 1,30]
≥ 30	2 388 (3,73)	13 (2,62)		134 (4,27)	1,25 [1,07 ; 1,46]
Utilisation récente ^b de THM (femmes ménopausées uniquement) ^d					
Œstrogène seul	1 747 (4,95)		-	108 (3,44)	1,07 [0,92 ; 1,23]
Œstrogène combiné à de la progestérone ou de la dydrogestérone	4 621 (13,08)		-	296 (9,43)	1,20 [1,09 ; 1,32]
Œstrogène combiné à un autre type de progestatif	5 253 (14,87)		-	412 (13,13)	1,72 [1,57 ; 1,88]
Autre THM ⁱ	900 (2,55)		-	57 (1,82)	1,17 [0,96 ; 1,44]
Pas d'utilisation récente de THM ^k	23 728 (67,17)		-	2 322 (74,00)	1,00

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : L'âge a été utilisé comme échelle de temps des modèles de régression de Cox, ce qui ne permet pas d'estimer d'HR associés à cette variable.

b : Les parents au premier degré incluent la mère, les sœurs et les filles.

c : Les maladies bénignes du sein incluent les adénomes et les maladies fibro-kystiques.

d : Les femmes non-cas de cancer incluent 28 675 (44,81 %) femmes préménopausées au début du suivi. Les pourcentages sont calculées parmi les femmes ménopausées en début de suivi (N = 35 324, 55,19 %).

e : Les silhouettes sont établies à partir de l'échelle des silhouettes de Sørensen (216).

f : Les valeurs des médianes utilisées pour créer la variable croisée à quatre catégories, en dichotomisant chaque score de profil alimentaire issu d'une analyse en composante principale, sont de -0,119594 pour le profil sain et de -0,122291 pour le profil occidental.

g : Les valeurs des médianes utilisées pour créer la variable croisée à quatre catégories, en dichotomisant la variable d'apport de vitamine D alimentaire et la dose journalière moyennée sur l'année du département de résidence, sont de 2,38 g/jour et de 2,50 kJ/m²/jour, respectivement.

h : L'utilisation récente de pilule contraceptive, de progestatif seul ou de THM a été définie comme une utilisation actuelle ou passée remontant à moins de 12 mois.

i : Les autres progestatifs incluent l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, le diénogest, la drospirénone, le diacétate d'éthinodiol, le gestodène, le lévonorgestrel, le lynestrénol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de mégestrol, l'acétate de nomégestrol, l'acétate de noréthistérone ou la promégestone.

j : Les autres THM incluent la tibolone, les THM de type non spécifié, les traitements injectables et les THM associant un œstrogène et un androgène.

k : 937 femmes (853 (1,3%) non-cas et 54 (1,7%) cas en postménopause) ont utilisé plusieurs types de THM différents.

Tableau 43 : Caractéristiques, au début de l'étude sur la prédiction du risque de cancer du sein en préménopause et en postménopause, de l'échantillon d'apprentissage et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 51 821). [Lien vers la première citation du tableau dans le texte]

	Non-cas	Cas de cancer du sein invasif en préménopause		Cas de cancer du sein invasif en postménopause	
	N = 50 891 N (%)	N = 393 N (%)	HR univarié [IC 95 %]	N = 537 N (%)	HR univarié [IC 95 %]
Statut ménopausique					
Non-ménopausées	22 969 (45,1 %)	393 (100,0 %)	-	-	-
Ménopausées	27 922 (54,9 %)	-	-	537 (100,0 %)	-
Âge (années)					
< 45	5 300 (10,4 %)	50 (12,7 %)	1,00	0 (0,0 %)	} 1,00
[45 ; 50[16 307 (32,0 %)	249 (63,4 %)	1,84 [1,36 ; 2,50]	25 (4,6 %)	
[50 ; 55[11 908 (23,4 %)	88 (22,4 %)	2,36 [1,67 ; 3,35]	166 (30,9 %)	2,14 [1,41 ; 3,26]
[55 ; 60[8 725 (17,2 %)	6 (1,5 %)	2,99 [1,28 ; 6,98]	170 (31,7 %)	2,08 [1,37 ; 3,17]
≥ 60	8 651 (17,0 %)	0 (0,0 %)		176 (32,8 %)	2,12 [1,40 ; 3,23]
Antécédents familiaux de cancer du sein chez les parents au premier degré ^a					
0	45 017 (88,5 %)	330 (84,0 %)	1,00	439 (81,8 %)	1,00
1	5 414 (10,6 %)	58 (14,7 %)	1,52 [1,15 ; 2,01]	87 (16,2 %)	1,59 [1,26 ; 2,00]
≥ 2	460 (0,9 %)	5 (1,3 %)	1,83 [0,76 ; 4,43]	11 (2,0 %)	2,08 [1,14 ; 3,77]
Taille à l'âge adulte (cm)					
< 160	17 399 (34,2 %)	116 (29,5 %)	1,00	191 (35,6 %)	1,00
[160 ; 165[17 798 (35,0 %)	129 (32,8 %)	0,93 [0,72 ; 1,19]	182 (33,9 %)	1,05 [0,85 ; 1,28]
[165 ; 170[11 129 (21,9 %)	103 (26,2 %)	1,09 [0,84 ; 1,42]	126 (23,4 %)	1,24 [0,99 ; 1,56]
≥ 170	4 565 (8,9 %)	45 (11,5 %)	1,04 [0,74 ; 1,47]	38 (7,1 %)	1,03 [0,73 ; 1,46]
Antécédents personnels de maladie bénigne du sein ^b					
0	36 075 (70,9 %)	225 (57,3 %)	1,00	338 (62,9 %)	1,00
≥ 1	14 816 (29,1 %)	168 (42,7 %)	1,59 [1,30 ; 1,94]	199 (37,1 %)	1,60 [1,34 ; 1,91]
Âge à la puberté (années)					
< 10	155 (0,3 %)	1 (0,3 %)	1,01 [0,14 ; 7,27]	3 (0,6 %)	1,94 [0,62 ; 6,08]
[10 ; 12[10 127 (19,9 %)	72 (18,3 %)	1,06 [0,78 ; 1,45]	102 (19,0 %)	1,04 [0,81 ; 1,34]
[12 ; 14[26 449 (52,0 %)	235 (59,8 %)	1,35 [1,05 ; 1,73]	282 (52,5 %)	1,08 [0,88 ; 1,31]
≥ 14	14 160 (27,8 %)	85 (21,6 %)	1,00	150 (28,0 %)	1,00
Nombre d'enfants					
0	5 837 (11,4 %)	48 (12,2 %)	1,26 [0,89 ; 1,78]	74 (13,8 %)	1,27 [0,97 ; 1,68]
1	7 818 (15,4 %)	70 (17,8 %)	1,21 [0,88 ; 1,65]	90 (16,8 %)	1,27 [0,98 ; 1,65]
2	22 238 (43,7 %)	183 (46,6 %)	1,03 [0,80 ; 1,33]	217 (40,4 %)	1,15 [0,94 ; 1,41]
≥ 3	14 998 (29,5 %)	92 (23,4 %)	1,00	156 (29,0 %)	1,00
Âge à la première grossesse à terme (pour les femmes ayant déjà eu des enfants uniquement) ^c					
≤ 20	4 370 (8,6 %)	35 (8,9 %)	1,00	35 (6,5 %)	1,00
[20 ; 25[24 214 (47,6 %)	162 (41,2 %)	0,89 [0,62 ; 1,28]	252 (46,9 %)	1,23 [0,87 ; 1,76]
[25 ; 30[12 478 (24,5 %)	100 (25,5 %)	1,03 [0,70 ; 1,51]	137 (25,5 %)	1,34 [0,92 ; 1,94]
[30 ; 35[3 148 (6,2 %)	33 (8,4 %)	1,28 [0,80 ; 2,06]	29 (5,4 %)	1,17 [0,71 ; 1,91]
> 35	844 (1,6 %)	15 (3,8 %)	1,97 [1,08 ; 3,60]	10 (1,9 %)	1,63 [0,81 ; 3,30]

Suite page suivante

	Non-cas	Cas de cancer du sein invasif en préménopause		Cas de cancer du sein invasif en postménopause	
Consommation d'alcool (g/jour)					
0	6 177 (12,1 %)	46 (11,7 %)	1,01 [0,73 ; 1,39]	58 (10,8 %)	0,94 [0,71 ; 1,25]
[0 ; 10[24 370 (47,9 %)	196 (49,9 %)	1,00	228 (42,5 %)	1,00
[10 ; 20[10 363 (20,4 %)	79 (20,1 %)	0,96 [0,74 ; 1,25]	128 (23,8 %)	1,30 [1,05 ; 1,61]
≥ 20	9 981 (19,6 %)	72 (18,3 %)	0,94 [0,71 ; 1,23]	123 (22,9 %)	1,26 [1,01 ; 1,57]
IMC (kg.m ⁻²)					
< 18,5	1 786 (3,5 %)	16 (4,1 %)	1,02 [0,62 ; 1,69]	13 (2,4 %)	0,78 [0,45 ; 1,35]
[18,5 ; 25[38 610 (75,9 %)	317 (80,7 %)	1,00	390 (72,6 %)	1,00
[25 ; 30[8 613 (16,9 %)	50 (12,7 %)	0,97 [0,72 ; 1,30]	107 (19,9 %)	0,98 [0,79 ; 1,22]
≥ 30	1 882 (3,7 %)	10 (2,5 %)	0,87 [0,47 ; 1,64]	27 (5,1 %)	1,13 [0,77 ; 1,67]
Âge à la ménopause (années, pour les femmes ménopausées uniquement) ^d					
< 48	2 713 (5,4 %)	-	-	28 (5,2 %)	1,00
[48 ; 50[9 096 (17,9 %)	-	-	174 (32,4 %)	1,83 [1,23 ; 2,73]
[50 ; 52[8 109 (15,9 %)	-	-	163 (30,4 %)	1,92 [1,29 ; 2,87]
[52 ; 54[4 950 (9,7 %)	-	-	101 (18,8 %)	1,95 [1,29 ; 2,97]
≥ 54	3 054 (6,0 %)	-	-	71 (13,2 %)	2,22 [1,43 ; 3,44]
Utilisation récente de THM (pour les femmes ménopausées uniquement) ^d					
Œstrogène seul par voie orale	1 493 (2,9 %)	-	-	27 (5,0 %)	0,94 [0,64 ; 1,38]
Œstrogène seul combiné à de la progestérone ou de la dydrogestérone	5 159 (10,1 %)	-	-	117 (21,8 %)	1,15 [0,91 ; 1,45]
Œstrogène seul combiné à un autre type de progestatif ^e	4 225 (8,3 %)	-	-	88 (16,4 %)	1,26 [1,02 ; 1,55]
Autre THM ^f	773 (1,5 %)	-	-	18 (3,3 %)	1,25 [0,78 ; 2,00]
Pas d'utilisation récente de THM ^g	17 033 (61,0 %)	-	-	301 (56,1 %)	1,00

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les parents au premier degré incluent la mère, les sœurs et les filles.

b : Les maladies bénignes du sein incluent les adénomes et les maladies fibro-kystiques.

c : Les non-cas de cancer incluent 5 836 (11,5 %) femmes nullipares (45 (12,2 %) en préménopause et 74 (13,8 %) en postménopause).

d : Les non-cas de cancer incluent 22 969 (45,1 %) femmes préménopausées.

e : Les autres types de progestatif incluent l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, le diénogest, la drospirénone, le diacétate d'éthinodiol, le gestodène, le lévonorgestrel, le lynestrérol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de mégéstrol, l'acétate de nomégéstrol, l'acétate de noréthistérone ou la promégestone.

f : Les autres THM incluent la tibolone, les THM de type non spécifié, les traitements injectables et les THM associant un œstrogène et un androgène.

g : 720 femmes (706 (1,4%) non-cas et 14 (2,6%) cas en postménopause) ont utilisé plusieurs types de THM différents sur une période de 12 mois.

Tableau 44 : Caractéristiques, au début de l'étude sur la prédiction du risque de cancer du sein en préménopause et en postménopause, de l'échantillon de validation et HR [IC 95 %] associés à chaque variable, cohorte E3N (N = 13 174). [Lien vers la première citation du tableau dans le texte]

	Non-cas	Cas de cancer du sein invasif en préménopause		Cas de cancer du sein invasif en postménopause	
	N = 12 998 N (%)	N = 78 N (%)	HR univarié [IC 95 %]	N = 98 N (%)	HR univarié [IC 95 %]
Statut ménopausique					
Non-ménopausées	5 765 (44,4 %)	78 (100 %)	-	-	-
Ménopausées	7 233 (55,6 %)	-	-	98 (100 %)	-
Âge (années)					
< 45	1 355 (10,4 %)	12 (15,4 %)	1,00	1 (1,0 %)	} 1,00
[45 ; 50[4 028 (31,0 %)	47 (60,2 %)	1,49 [0,79 ; 2,81]	2 (2,0 %)	
[50 ; 55[3 058 (23,5 %)	17 (21,8 %)	1,84 [0,88 ; 3,86]	37 (37,8 %)	3,91 [1,21 ; 12,69]
[55 ; 60[2 291 (17,6 %)	2 (2,6 %)	4,82 [1,08 ; 21,52]	40 (40,8 %)	3,85 [1,19 ; 12,43]
≥ 60	2 266 (17,5 %)	0 (0,0 %)		18 (18,4 %)	1,72 [0,51 ; 5,85]
Antécédents familiaux de cancer du sein chez les parents au premier degré ^a					
0	11 478 (88,3 %)	65 (83,3 %)	1,00	81 (82,7 %)	1,00
1	1 375 (10,6 %)	12 (15,4 %)	1,47 [0,79 ; 2,72]	13 (13,2 %)	1,39 [0,78 ; 2,50]
≥ 2	145 (1,1 %)	1 (1,3 %)	1,71 [0,24 ; 12,34]	4 (4,1 %)	3,13 [1,15 ; 8,54]
Taille à l'âge adulte (cm)					
< 160	4 438 (34,2 %)	21 (26,9 %)	1,00	40 (40,8 %)	1,00
[160 ; 165[4 539 (34,9 %)	31 (39,8 %)	1,25 [0,72 ; 2,17]	31 (31,6 %)	0,82 [0,51 ; 1,31]
[165 ; 170[2 833 (21,8 %)	16 (26,5 %)	0,95 [0,49 ; 1,81]	22 (22,5 %)	1,01 [0,60 ; 1,71]
≥ 170	1 188 (9,1 %)	10 (12,8 %)	1,31 [0,62 ; 2,78]	5 (5,1 %)	0,61 [0,24 ; 1,54]
Antécédents personnels de maladie bénigne du sein ^b					
0	9 224 (71,0 %)	49 (62,8 %)	1,00	59 (60,2 %)	1,00
≥ 1	3 774 (29,0 %)	29 (37,2 %)	1,21 [0,77 ; 1,92]	39 (39,8 %)	1,91 [1,28 ; 2,87]
Âge à la puberté (années)					
< 10	53 (0,4 %)	1 (1,3 %)	2,96 [0,39 ; 22,28]	0 (0,0 %)	} 1,34 [0,77 ; 2,47]
[10 ; 12[2 628 (20,2 %)	16 (20,5 %)	1,20 [0,60 ; 2,37]	22 (22,4 %)	
[12 ; 14[6 694 (51,5 %)	44 (56,4 %)	1,23 [0,70 ; 2,45]	53 (54,1 %)	1,36 [0,83 ; 2,22]
≥ 14	3 623 (27,9 %)	17 (21,8 %)	1,00	23 (23,5 %)	1,00
Nombre d'enfants					
0	1 536 (11,8 %)	6 (7,7 %)	0,83 [0,33 ; 2,10]	13 (13,3 %)	1,53 [0,77 ; 3,06]
1	2 053 (15,8 %)	14 (18,0 %)	1,17 [0,57 ; 2,36]	18 (18,4 %)	1,81 [0,96 ; 3,39]
2	5 714 (44,0 %)	41 (52,5 %)	1,18 [0,67 ; 2,08]	46 (46,9 %)	1,73 [1,03 ; 2,90]
≥ 3	3 695 (28,4 %)	17 (21,8 %)	1,00	21 (21,4 %)	1,00
Âge à la première grossesse à terme (pour les femmes pares uniquement) ^c					
< 20	1 101 (8,5 %)	6 (7,7 %)	1,00	4 (4,1 %)	1,00
[20 ; 25[6 141 (47,2 %)	30 (38,4 %)	0,90 [0,37 ; 2,16]	49 (50,0 %)	2,19 [0,79 ; 6,06]
[25 ; 30[3 220 (24,8 %)	24 (30,8 %)	1,34 [0,55 ; 3,28]	20 (20,4 %)	1,74 [0,60 ; 5,09]
[30 ; 35[779 (6,0 %)	10 (12,8 %)	2,18 [0,79 ; 5,99]	10 (10,2 %)	3,67 [1,15 ; 11,71]
≥ 35	221 (1,7 %)	2 (2,6 %)	1,43 [0,29 ; 7,10]	2 (2,0 %)	2,67 [0,49 ; 14,58]

Suite page suivante

	Non-cas	Cas de cancer du sein invasif en préménopause		Cas de cancer du sein invasif en postménopause	
Consommation d'alcool (g/jour)					
0	1 584 (12,2 %)	6 (7,7 %)	0,77 [0,32 ; 1,82]	16 (16,3 %)	1,34 [0,75 ; 2,38]
]0 ; 10[6 182 (47,6 %)	35 (44,9 %)	1,00	42 (42,9 %)	1,00
[10 ; 20[2 690 (20,7 %)	13 (16,7 %)	0,88 [0,46 ; 1,66]	23 (23,5 %)	1,24 [0,75 ; 2,07]
≥ 20	2 542 (19,5 %)	24 (30,7 %)	1,69 [1,00 ; 1,28]	17 (17,3 %)	0,96 [0,55 ; 1,69]
IMC (kg.m ⁻²)					
< 18,5	410 (3,2 %)	2 (2,6 %)	0,69 [0,17 ; 2,83]	4 (4,1 %)	1,54 [0,56 ; 4,21]
[18,5 ; 25[9 866 (75,9 %)	65 (83,3 %)	1,00	68 (69,4 %)	1,00
[25 ; 30[2 260 (17,4 %)	10 (12,8 %)	0,89 [0,46 ; 1,74]	21 (21,4 %)	1,09 [0,67 ; 1,78]
≥ 30	462 (3,5 %)	1 (1,3 %)	0,45 [0,06 ; 3,27]	5 (5,1 %)	1,26 [0,51 ; 3,13]
Âge à la ménopause (années, pour les femmes ménopausées uniquement) ^d					
< 48	724 (5,6 %)	-	-	4 (4,1 %)	1,00
[48 ; 50[2 287 (17,6 %)	-	-	26 (26,5 %)	2,02 [0,71 ; 5,80]
[50 ; 52[2 107 (16,2 %)	-	-	33 (33,7 %)	2,78 [0,99 ; 7,86]
[52 ; 54[1 298 (10,0 %)	-	-	16 (16,3 %)	2,18 [0,73 ; 6,53]
≥ 54	817 (6,3 %)	-	-	19 (19,4 %)	4,30 [1,46 ; 12,63]
Utilisation récente de THM (pour les femmes ménopausées uniquement) ^d					
Œstrogène seul par voie orale	1 327 (10,2 %)	-	-	25 (25,5 %)	0,50 [0,16 ; 1,59]
Œstrogène seul combiné à de la progestérone ou de la hydrogestérone	421 (3,2 %)	-	-	3 (3,1 %)	1,75 [1,06 ; 2,88]
Œstrogène seul combiné à un autre type de progestatif ^e	1 083 (8,3 %)	-	-	21 (21,4 %)	1,71 [1,07 ; 2,74]
Autre THM ^f	189 (1,5 %)	-	-	2 (2,0 %)	0,87 [0,21 ; 3,53]
Pas d'utilisation récente de THM ^g	4 412 (61,0 %)	-	-	48 (49,0 %)	1,00

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

a : Les parents au premier degré incluent la mère, les sœurs et les filles.

b : Les maladies bénignes du sein incluent les adénomes et les maladies fibro-kystiques.

c : Les non-cas de cancer incluent 1 536 (11,8 %) femmes nullipares (6 (7,7 %) en préménopause et 13 (13,3 %) en postménopause).

d : Les non-cas de cancer incluent 5 764 (44,3 %) femmes préménopausées.

e : Les autres types de progestatif incluent l'acétate de chlormadinone, l'acétate de cyprotérone, la démégestone, le diénogest, la drospirénone, le diacétate d'éthinodiol, le gestodène, le lévonorgestrel, le lynestrérol, la médrogestone, l'acétate de médroxyprogestérone, l'acétate de mégéstrol, l'acétate de nomégéstrol, l'acétate de noréthistérone ou la promégestone.

f : Les autres THM incluent la tibolone, les THM de type non spécifié, les traitements injectables et les THM associant un œstrogène et un androgène.

g : 182 femmes (181 (1,4%) non-cas et 1 (1,0%) cas en postménopause) ont utilisé plusieurs types de THM différents sur une période de 12 mois.

Tableau 45 : Caractéristiques, au début de l'étude sur l'association entre l'indice de santé et le risque de mortalité, de la population d'analyse, cohorte E3N (N = 66 721). [Lien vers la première citation dans le texte]

	Population entière (N = 66 721, 100,00 %)	Catégories de l'indice de santé			
		[0 ; 2] (N = 4 339, 6,50 %)	[2,5 ; 3] (N = 15 325, 22,97 %)	[3,5 ; 4] (N = 27 904, 41,82 %)	[4,5 ; 5] (N = 19 153, 28,71 %)
Variabiles d'ajustement					
Âge [moyenne (écart-type)]	52,87 (6,66)	51,61 (6,27)	52,21 (6,49)	52,96 (6,68)	53,55 (6,76)
Apport énergétique journalier (kcal/jour) [moyenne (écart-type)]	2 212 (558)	2 262 (601)	2 243 (586)	2 207 (552)	2 184 (531)
Niveau d'éducation [n (%)]					
Inférieur à Bac	7 194 (10,78)	444 (10,23)	1 603 (10,46)	2 979 (10,68)	2 168 (11,32)
Bac-Bac+2	35 665 (53,45)	2 176 (50,15)	7 890 (51,48)	14 963 (53,62)	10 636 (55,53)
Supérieur à Bac+2	23 862 (35,76)	1 719 (39,62)	5 832 (38,06)	9 962 (35,70)	6 349 (33,15)
Zone de résidence [n (%)]					
Nord	30 789 (46,15)	2 371 (54,64)	7 613 (49,68)	12 759 (45,72)	8 046 (42,01)
Centre	16 273 (24,39)	903 (20,81)	3 488 (22,76)	6 915 (24,78)	4 967 (25,93)
Sud	19 659 (29,46)	1 065 (24,54)	4 224 (27,56)	8 230 (29,49)	6 140 (32,06)
Antécédents familiaux de cancer chez les parents au premier degré ^a [n (%)]					
Au moins un	31 965 (47,91)	2 144 (49,41)	7 424 (48,44)	13 268 (47,55)	9 129 (47,66)
Aucun	34 756 (52,09)	2 195 (50,59)	7 901 (51,56)	14 636 (52,45)	10 024 (52,34)
Activité professionnelle [n (%)]					
En activité	45 900 (68,79)	3 394 (78,22)	11 285 (73,64)	19 056 (68,29)	12 165 (63,51)
Retraîtée/N'a jamais travaillé	20 821 (31,21)	945 (21,78)	4 040 (26,36)	8 848 (31,71)	9 988 (36,49)
Statut ménopausique et utilisation récente de THM [n (%)]					
Préménopause	29 030 (43,51)	2 136 (49,23)	7 251 (47,31)	12 004 (43,02)	7 639 (39,88)
Postménopause avec THM récent	14 601 (21,88)	817 (18,83)	3 130 (20,42)	6 176 (22,13)	4 478 (23,38)
Postménopause sans THM récent	23 090 (34,61)	1 386 (31,94)	4 911 (32,26)	9 724 (34,85)	7 036 (36,74)
Comportements de mode de vie					
Tabagisme [n (%)]					
Fumeuse ou ex-fumeuse régulière depuis moins de 10 ans	12 756 (19,12)	2 853 (65,75)	5 866 (38,28)	4 037 (14,47)	0 (0,00)
Ex-fumeuse régulière depuis plus de 10 ans ou occasionnelle	17 744 (26,59)	982 (22,63)	4 916 (32,08)	8 782 (31,47)	3 064 (16,00)
N'a jamais fumé	36 221 (54,29)	504 (11,62)	4 543 (29,64)	15 085 (54,06)	16 089 (84,00)
IMC, kg.m ⁻² [n (%)]					
Moyenne (écart-type)	22,95 (3,30)	24,69 (4,94)	23,63 (3,95)	22,87 (3,04)	22,13 (2,17)
< 16	76 (0,11)	15 (0,35)	33 (0,22)	28 (0,10)	0 (0,00)
[16 ; 18,5[2 208 (3,31)	280 (6,45)	669 (4,37)	923 (3,31)	336 (1,75)
[18,5 ; 25[50 516 (75,71)	2 171 (50,03)	9 828 (64,13)	21 262 (76,20)	17 255 (90,09)
[25 ; 30[11 448 (17,16)	1 194 (27,52)	3 683 (24,03)	5 009 (17,95)	1 562 (8,16)
≥ 30	2 473 (3,71)	679 (15,65)	1 112 (7,26)	682 (2,44)	0 (0,00)

Suite page suivante

	Population entière	Catégories de l'indice de santé			
		[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Consommation d'alcool [n (%)]					
Moyenne (écart-type)	11,57 (13,84)	28,23 (19,50)	18,22 (16,14)	10,27 (11,69)	4,37 (4,57)
< 1 verre ^b /jour	39858 (59,74)	582 (13,41)	5 310 (34,65)	16 933 (60,68)	17 033 (88,93)
[1 ; 2[verres/jour	13 724 (20,57)	837 (19,29)	3 981 (25,98)	6 786 (24,32)	2 120 (11,07)
≥ 2 verres/jour	13 139 (19,69)	2 920 (67,30)	6 034 (39,37)	4 185 (15,00)	0 (0,00)
Consommation de fruits et légumes [n (%)]					
Moyenne (écart-type)	529 (241)	327 (195)	449 (234)	540 (232)	625 (217)
5 portions ^c par jour ou plus (≥ 400 g/jour)	45 564 (68,29)	954 (21,99)	7 390 (48,22)	19 675 (70,51)	17 545 (91,60)
[3,5 ; 5[portions par jour	12 596 (18,88)	1 065 (24,54)	4 135 (26,98)	5 788 (20,74)	1 608 (8,40)
Moins de 3,5 portions par jour	8 561 (12,83)	2 320 (53,47)	3 800 (24,80)	2 441 (8,75)	0 (0,00)
Niveau d'activité physique récréationnelle, METs-h/semaine [n (%)]					
Moyenne (écart-type)	29,37 (34,80)	14,00 (19,39)	22,06 (31,29)	28,72 (33,44)	39,62 (38,96)
< 10	13 292 (19,92)	2 536 (58,45)	5 552 (36,23)	5 204 (18,65)	0 (0,00)
[10 ; 20[17 353 (26,01)	1 116 (25,72)	4 635 (30,24)	8 329 (29,85)	3 273 (17,09)
≥ 20	39 076 (54,07)	687 (15,83)	5 138 (33,53)	14 371 (51,50)	15 880 (82,91)

Note : Les proportions observées pour chaque modalité des variables qualitatives entre les femmes ayant un indice compris entre 0 et 2 et les femmes ayant un indice compris entre 4,5 et 5 sont toutes significativement différentes au seuil de 5 % (p-valeur < 0,040), sauf pour les proportions de femmes ayant un apport énergétique journalier inférieur à 1 816 kcal/jour (p-valeur = 0,081) et ayant un apport énergétique entre 2 159 et 2 551 kcal/jour (p-valeur = 0,2697). Les moyennes observées pour chaque variable quantitative entre les femmes ayant un indice compris entre 0 et 2 et les femmes ayant un indice compris entre 4,5 et 5 sont toutes significativement différentes au seuil de 5 % (p-valeur < 0,0001).

a : Les parents au premier degré incluent la mère, les sœurs et les filles.

b : Un verre d'alcool correspond à 10 grammes d'éthanol.

c : Une portion de fruits et légumes correspond à 80 grammes de fruits et légumes, qui incluent les légumes crus et cuits, la salade et les fruits frais.

Annexe D : Résultats complémentaires pour l'analyse de l'association entre le mode de vie et le risque de cancer

Tableau 46 : HR [IC 95 %] de l'association entre les cinq comportements de mode de vie et le risque de mélanome et de cancer cutané non basal non mélanome, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]

Localisation ^a Score	Tabagisme actif		IMC		Consommation d'alcool		Consommation de fruits et légumes		Niveau d'activité physique récréationnelle	
	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b	N (%)	HR [IC 95 %] ^b
Mélanome (N = 304)										
0	40 (13,16)	1,00	9 (2,96)	1,00	55 (18,09)	1,00	32 (10,53)	1,00	39 (12,83)	1,00
0,5	105 (34,54)	1,02 [0,71 ; 1,48]	61 (20,07)	1,19 [0,59 ; 2,39]	60 (19,74)	1,04 [0,72 ; 1,50]	62 (20,39)	1,29 [0,84 ; 1,97]	77 (25,33)	1,50 [1,02 ; 2,20]
1	159 (52,30)	0,96 [0,67 ; 1,38]	234 (76,97)	1,15 [0,59 ; 2,25]	189 (62,17)	1,16 [0,85 ; 1,58]	210 (69,08)	1,19 [0,82 ; 1,73]	188 (61,84)	1,78 [1,26 ; 2,52]
P-valeur de tendance		0,7262		0,9250		0,7608		0,7506		0,4469
Cancer cutané non basal non mélanome (N = 382)										
0	42 (10,99)	1,00	10 (2,62)	1,00	66 (17,28)	1,00	39 (10,21)	1,00	56 (14,66)	1,00
0,5	104 (27,23)	0,89 [0,62 ; 1,27]	72 (18,85)	1,25 [0,65 ; 2,43]	81 (21,20)	1,15 [0,83 ; 1,59]	72 (18,59)	1,17 [0,79 ; 1,73]	97 (25,39)	1,27 [0,91 ; 1,76]
1	236 (61,78)	1,14 [0,81 ; 1,60]	300 (78,53)	1,43 [0,76 ; 2,70]	235 (61,52)	1,14 [0,87 ; 1,51]	272 (71,20)	1,23 [0,80 ; 1,58]	229 (59,95)	1,36 [1,01 ; 1,82]
P-valeur de tendance		0,1211		0,0515		0,6189		0,1781		0,1173

Note : Les mélanomes sont identifiés par les codes CIM 9 172 et CIM 10 C43 ; les cancers cutané non basal non mélanome par les codes CIM 9 173 et CIM 10 C44. Dans les deux catégories, les codes CIM-O-3 M809-M811 ont été exclus.

Tableau 47 : Proportions [IC 95 %] de cas de mélanome et de cancer cutané non basal et de cancer cutané non basal et de cancer cutané non basal non mélanome attribuables à la non-adhésion aux recommandations individuelles de santé publique et à leur combinaison, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]

	Proportions de cancers attribuables aux comportements de mode de vie ^a				Proportions de cancers attribuables à un mode de vie non sain ^b
	Tabagisme actif	IMC	Consommation d'alcool	Consommation de fruits et légumes	
Mélanome (N = 304)	2,5 [-12,0 ; 16,9]	0,1 [-9,2 ; 9,4]	-5,2 [-16,9 ; 6,8]	-0,5 [-11,1 ; 10,1]	-14,9 [-27,8 ; -1,5]
Cancer cutané non basal non mélanome (N = 382)	-9,9 [-22,7 ; 3,2]	-3,5 [-11,4 ; 4,4]	-2,4 [-12,9 ; 8,2]	-0,7 [-10,4 ; 9,0]	-22,2 [-39,3 ; -3,5]

Note : Les mélanomes sont identifiés par les codes CIM 9 172 et CIM 10 C43 ; les cancers cutané non basal non mélanome par les codes CIM 9 173 et CIM 10 C44. Dans les deux catégories, les codes CIM-O-3 M809-M811 ont été exclus.

Tableau 48 : HR [IC 95 %] de l'association entre l'indice de santé et le risque de mélanome et de cancer cutané non basal non mélanome, cohorte E3N (N = 64 732). [Lien vers la première citation du tableau dans le texte]

	Catégories de l'indice de santé			
	[0 ; 2]	[2,5 ; 3]	[3,5 ; 4]	[4,5 ; 5]
Mélanome (N = 304)				
N (%)	450 (6,48)	1554 (22,40)	2974 (42,87)	1960 (28,25)
HR [IC 95 %] ^c	1,00	0,95 [0,54 ; 1,69]	1,19 [0,69 ; 2,03]	1,44 [0,84 ; 2,48]
HR (augmentation de 1 point)		1,18 [1,03 ; 1,36]		
P-valeur de tendance		0,0172		
Cancer cutané non basal non mélanome (N = 382)				
N (%)	32 (15,02)	61 (28,64)	85 (39,90)	35 (16,44)
HR [IC 95 %] ^c	1,00	1,59 [0,84 ; 3,00]	1,70 [0,92 ; 3,13]	2,16 [1,17 ; 4,00]
HR (augmentation de 1 point)		1,21 [1,06 ; 1,37]		
P-valeur de tendance		0,0034		

Note : Les mélanomes sont identifiés par les codes CIM 9 172 et CIM 10 C43 ; les cancers cutané non basal non mélanome par les codes CIM 9 173 et CIM 10 C44. Dans les deux catégories, les codes CIM-O-3 M809-M811 ont été exclues.

Annexe E : Résultats complémentaires pour l'analyse des poids relatifs des facteurs de risque de cancer du sein en préménopause et en postménopause

Tableau 49 : Proportions [IC 95 %] de cas de cancer du sein en préménopause RO+ et RO- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]

	Cancer du sein en préménopause	
	RO+ N = 314	RO- N = 97
Facteurs de risque non-comportementaux	PAF [IC 95 %]	PAF [IC 95 %]
Antécédents familiaux de cancer du sein chez les parents au premier degré (réf. = aucun)	-0,26 [-3,19 ; 2,67]	5,82 [-2,24 ; 13,80]
Niveau d'éducation (réf. = inférieur au Bac)	5,33 [-34,50 ; 43,54]	-19,00 [-76,60 ; 55,40]
Taille à l'âge adulte (réf. = inférieure à 1,60 m)	7,91 [-14,60 ; 29,60]	-26,20 [-65,10 ; 23,61]
Antécédents personnels de maladie bénigne du sein (réf. = aucun)	16,21 [5,58 ; 26,48]	11,79 [-5,62 ; 28,51]
Âge à la puberté (réf. = supérieur à 14 ans)	14,14 [-4,52 ; 31,84]	38,62 [6,11 ; 63,72]
Poids à la naissance (réf. = petit)	21,37 [-17,70 ; 54,58]	61,01 [17,56 ; 84,56]
Âge à la ménopause, pour les femmes ménopausées uniquement (réf. = inférieur à 48 ans)	-	-
Combinaison des facteurs de risque non-comportementaux	52,18 [-74,90 ; 97,21]	70,32 [-82,60 ; 99,42]
Facteurs de risque comportementaux		
Tabagisme actif (réf. = n'a jamais fumé)	8,87 [-4,43 ; 21,86]	-1,14 [-23,70 ; 21,50]
Niveau d'activité physique total (réf. = élevé)	-4,31 [-29,50 ; 21,46]	12,69 [-30,50 ; 51,55]
Nombre d'enfants et âge à la première grossesse à terme (réf. = plus d'un enfant dont le premier avant 30 ans)	4,84 [-10,30 ; 19,79]	11,16 [-15,90 ; 36,65]
Durée d'allaitement cumulée (réf. = plus de 6 mois)	2,26 [-27,70 ; 31,82]	-1,43 [-51,10 ; 48,93]
Silhouette à la puberté (réf. = ronde)	9,66 [-14,00 ; 32,28]	13,03 [-27,00 ; 49,22]
Profil alimentaire (réf. = scores élevés pour les deux profils alimentaires)	12,64 [-12,40 ; 36,12]	14,72 [-27,80 ; 52,46]
Consommation d'alcool (réf. = moins d'un verre par jour)	-0,92 [-13,00 ; 11,17]	-6,45 [-26,70 ; 14,32]
Vitamine D alimentaire et exposition aux UV solaires (réf. = niveaux élevés)	3,57 [-21,40 ; 28,06]	-10,90 [-52,50 ; 34,99]
Utilisation de contraceptif oral ou de progestatif seul (réf. = utilisation passée remontant à plus de 10 ans, ou jamais utilisé)	5,91 [-13,50 ; 24,90]	-1,08 [-33,00 ; 31,11]
IMC (réf. = supérieur 25 (en préménopause) ou inférieur à 25 (en postménopause))	25,01 [3,81 ; 44,07]	-12,40 [-49,80 ; 28,82]
Utilisation récente de THM pour les femmes ménopausées uniquement (réf. = pas d'utilisation récente de THM)	-	-
Combinaison des facteurs de risque comportementaux	50,56 [-50,90 ; 93,21]	21,80 [-99,60 ; 99,83]
Combinaison des facteurs de risque comportementaux et non-comportementaux	76,30 [-86,90 ; 99,75]	76,83 [-99,80 ; 100,00]

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

Tableau 50 : Proportions [IC 95 %] de cas de cancer du sein en postménopause HER2+, HER2-, RO+ et RO- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]

	HER2+ N = 178	HER2- N = 747	RO+ N = 2 050	RO- N = 466
Facteurs de risque non-comportementaux				
Antécédents familiaux de cancer du sein chez les parents au premier degré (ref = aucun)	0,09 [-4,39 ; 4,58]	4,26 [1,75 ; 6,76]	4,25 [2,76 ; 5,73]	2,20 [-0,69 ; 5,09]
Niveau d'éducation (ref = inférieur au Bac)	-26,00 [-68,70 ; 30,13]	25,54 [4,73 ; 44,23]	6,45 [-6,86 ; 19,53]	4,73 [-22,00 ; 30,84]
Taille à l'âge adulte (ref = inférieure à 1,60 m)	19,35 [-7,83 ; 43,86]	12,71 [-0,99 ; 25,95]	8,70 [0,34 ; 16,93]	-0,35 [-18,10 ; 17,41]
Antécédents personnels de maladie bénigne du sein (ref = aucun)	16,55 [3,00 ; 29,50]	11,70 [5,31 ; 17,99]	14,51 [10,57 ; 18,41]	15,31 [6,98 ; 23,44]
Âge à la puberté (ref = supérieur à 14 ans)	5,47 [-18,80 ; 29,14]	9,72 [-2,49 ; 21,65]	14,47 [7,37 ; 21,42]	3,51 [-11,90 ; 18,72]
Poids à la naissance (ref = petit)	3,69 [-48,00 ; 53,45]	14,40 [-10,90 ; 37,93]	5,81 [-10,10 ; 21,42]	24,63 [-4,53 ; 49,92]
Âge à la ménopause, pour les femmes ménopausées uniquement (ref = inférieur à 48 ans)	39,10 [6,24 ; 64,30]	-1,52 [-20,80 ; 17,89]	12,83 [2,04 ; 23,31]	-7,53 [-33,00 ; 19,01]
Combinaison des facteurs de risque non-comportementaux	53,04 [-77,60 ; 97,66]	56,74 [-11,40 ; 88,58]	50,82 [9,25 ; 77,31]	38,07 [-58,50 ; 89,98]
Facteurs de risque comportementaux				
Tabagisme actif (ref. = n'a jamais fumé)	8,06 [-7,87 ; 23,59]	-0,68 [-8,54 ; 7,20]	-0,20 [-4,88 ; 4,47]	1,80 [-7,61 ; 11,17]
Niveau d'activité physique total (ref. = élevé)	-4,03 [-34,50 ; 27,22]	4,08 [-10,90 ; 18,89]	3,08 [-5,93 ; 12,03]	-0,34 [-19,10 ; 18,46]
Nombre d'enfants et âge à la première grossesse à terme (ref. = plus d'un enfant dont le premier avant 30 ans)	-1,08 [-19,90 ; 17,86]	4,00 [-6,00 ; 13,92]	7,05 [1,00 ; 13,05]	-6,95 [-18,60 ; 4,85]
Durée d'allaitement cumulé (ref. = plus de 6 mois)	4,48 [-31,40 ; 39,23]	15,15 [-2,98 ; 32,31]	7,82 [-3,44 ; 18,89]	-0,65 [-24,00 ; 22,76]
Silhouette à la puberté (ref. = ronde)	8,72 [-24,10 ; 39,77]	15,99 [0,92 ; 30,35]	15,25 [6,03 ; 24,20]	6,82 [-13,80 ; 26,91]
Profil alimentaire (ref. = scores élevés pour les deux profils alimentaires)	3,59 [-27,50 ; 34,01]	16,06 [1,07 ; 30,34]	12,85 [3,75 ; 21,73]	7,82 [-11,60 ; 26,67]
Consommation d'alcool (ref. = moins d'un verre par jour)	4,78 [-11,70 ; 21,01]	9,54 [1,18 ; 17,77]	3,60 [-1,33 ; 8,52]	8,48 [-1,83 ; 18,61]
Vitamine D alimentaire et exposition aux UV solaires (ref. = niveaux élevés)	-10,10 [-43,50 ; 25,73]	8,35 [-7,89 ; 24,16]	8,50 [-1,14 ; 17,98]	4,64 [-15,70 ; 24,59]
Utilisation de contraceptif oral ou de progestatif seul (ref. = utilisation passée remontant à plus de 10 ans, ou jamais utilisé)	-4,69 [-24,50 ; 15,51]	2,11 [-8,31 ; 12,49]	3,81 [-2,31 ; 9,90]	2,17 [-9,80 ; 14,07]
IMC (ref. = supérieur 25 (en prémenopause) ou inférieur à 25 (en postménopause))	12,21 [-1,06 ; 25,06]	8,12 [2,03 ; 14,15]	6,47 [2,69 ; 10,23]	-1,05 [-9,70 ; 7,61]
Utilisation récente de THM, pour les femmes ménopausées uniquement (ref. = pas d'utilisation récente de THM)	0,63 [-19,70 ; 20,93]	12,99 [2,06 ; 23,60]	17,51 [10,99 ; 23,88]	5,00 [-8,34 ; 18,17]
Combinaison des facteurs de risque comportementaux	22,39 [-97,30 ; 98,90]	63,91 [-15,90 ; 93,21]	59,81 [12,83 ; 84,87]	25,28 [-82,70 ; 93,48]
Combinaison des facteurs de risque comportementaux et non-comportementaux	63,57 [-98,90 ; 99,94]	84,33 [-24,40 ; 99,13]	80,21 [20,27 ; 96,43]	53,61 [-91,60 ; 99,21]

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

Tableau 51 : Proportions [IC 95 %] de cas de cancer du sein en postménopause RO+/HER2+, RO+/HER2-, RO-/HER2+ et RO-/HER2- attribuables à un facteur ou à la combinaison de plusieurs facteurs de risque, cohorte E3N (N = 67 634). [Lien vers la première citation du tableau dans le texte]

	RO+/HER2+ N = 113	RO+/HER2- N = 629	RO-/HER2+ N = 63	RO-/HER2- N = 107
Facteurs de risque non-comportementaux				
Antécédents familiaux de cancer du sein chez les parents au premier degré (réf. = aucun)	4,23 [-2,29 ; 10,72]	4,14 [1,41 ; 6,87]	-7,19 [-10,00 ; -4,35]	6,24 [-0,72 ; 13,14]
Niveau d'éducation (réf. = inférieur au Bac)	-20,90 [-72,40 ; 45,61]	27,10 [4,60 ; 46,99]	-24,70 [-83,30 ; 59,97]	9,36 [-46,20 ; 59,61]
Taille à l'âge adulte (réf. = inférieure à 1,60 m)	27,51 [-6,13 ; 55,54]	13,11 [-1,84 ; 27,48]	7,94 [-37,00 ; 49,88]	14,72 [-20,80 ; 46,79]
Antécédents personnels de maladie bénigne du sein (réf. = aucun)	15,62 [-1,29 ; 31,66]	10,91 [3,96 ; 17,76]	16,06 [-6,82 ; 37,33]	15,75 [-1,64 ; 32,21]
Âge à la puberté (réf. = supérieur à 14 ans)	2,21 [-27,70 ; 31,75]	11,91 [-1,27 ; 24,68]	14,58 [-26,00 ; 50,74]	-7,19 [-39,40 ; 26,59]
Poids à la naissance (réf. = petit)	-12,20 [-72,40 ; 58,59]	11,73 [-16,20 ; 37,94]	28,38 [-48,90 ; 80,72]	23,12 [-38,00 ; 70,21]
Âge à la ménopause, pour les femmes ménopausées uniquement (réf. = inférieur à 48 ans)	40,69 [-0,21 ; 69,93]	4,60 [-15,80 ; 24,66]	34,23 [-24,10 ; 74,38]	-31,00 [-76,10 ; 34,15]
Combinaison des facteurs de risque non-comportementaux	53,87 [-88,50 ; 98,91]	59,73 [-12,30 ; 90,54]	58,51 [-97,00 ; 99,79]	33,93 [-96,20 ; 99,05]
Facteurs de risque comportementaux				
Tabagisme actif (réf. = n'a jamais fumé)	11,16 [-9,88 ; 31,25]	-1,89 [-10,50 ; 6,72]	5,08 [-19,40 ; 28,97]	11,28 [-10,20 ; 31,79]
Niveau d'activité physique total (réf. = élevé)	1,16 [-36,30 ; 38,28]	4,69 [-11,70 ; 20,78]	-17,10 [-63,60 ; 38,49]	-5,21 [-43,00 ; 34,12]
Nombre d'enfants et âge à la première grossesse à terme (réf. = plus d'un enfant dont le premier avant 30 ans)	7,31 [-17,80 ; 31,48]	7,01 [-3,93 ; 17,77]	-18,30 [-42,00 ; 7,76]	-16,20 [-40,00 ; 9,63]
Durée d'allaitement cumulé (réf. = plus de 6 mois)	3,77 [-41,50 ; 47,59]	14,02 [-5,94 ; 32,91]	5,33 [-48,90 ; 56,54]	17,36 [-28,20 ; 56,56]
Silhouette à la puberté (réf. = ronde)	12,38 [-28,20 ; 49,18]	19,85 [3,82 ; 34,88]	0,11 [-52,20 ; 52,37]	-0,62 [-42,70 ; 41,72]
Profil alimentaire (réf. = scores élevés pour les deux profils alimentaires)	0,53 [-38,20 ; 39,08]	16,56 [0,23 ; 32,03]	12,09 [-37,50 ; 56,28]	13,81 [-25,90 ; 49,57]
Consommation d'alcool (réf. = moins d'un verre par jour)	8,22 [-13,20 ; 28,85]	8,94 [-0,18 ; 17,91]	0,62 [-25,10 ; 26,21]	14,23 [-8,20 ; 35,30]
Vitamine D alimentaire et exposition aux UV solaires (réf. = niveaux élevés)	-15,00 [-54,40 ; 29,85]	8,28 [-9,57 ; 25,61]	2,59 [-51,30 ; 55,02]	14,15 [-25,80 ; 50,00]
Utilisation de contraceptif oral ou de progestatif seul (réf. = utilisation passée remontant à plus de 10 ans, ou jamais utilisé)	-13,40 [-35,70 ; 10,40]	3,20 [-8,34 ; 14,66]	13,12 [-25,50 ; 48,10]	-5,60 [-29,50 ; 18,96]
IMC (réf. = supérieur 25 (en préménopause) ou inférieur à 25 (en postménopause))	11,33 [-5,15 ; 27,21]	10,45 [3,94 ; 16,87]	13,33 [-9,46 ; 34,81]	-6,05 [-23,80 ; 12,12]
Utilisation récente de THM, pour les femmes ménopausées uniquement (réf. = pas d'utilisation récente de THM)	8,04 [-18,30 ; 33,31]	15,00 [3,06 ; 26,53]	-13,50 [-43,30 ; 19,02]	-0,55 [-26,80 ; 25,73]
Combinaison des facteurs de risque comportementaux	33,75 [-98,40 ; 99,60]	68,00 [-15,30 ; 94,80]	11,42 [-100,00 ; 99,97]	36,07 [-98,50 ; 99,66]
Combinaison des facteurs de risque comportementaux et non-comportementaux	69,46 [-99,70 ; 99,99]	87,08 [-25,70 ; 99,44]	63,00 [-100,00 ; 100,00]	57,78 [-99,90 ; 99,99]

Note : Les résultats en gras correspondent à des résultats statistiquement significatifs au niveau 5 % (p-valeur < 0,05).

Annexe F : « Association between five lifestyle habits and cancer risk : Results from the E3N cohort » - Article publié dans le journal « Cancer Prevention Research »

Published OnlineFirst February 26, 2014; DOI: 10.1158/1940-6207.CAPR-13-0325

Cancer
Prevention
Research

Research Article

Association between Five Lifestyle Habits and Cancer Risk: Results from the E3N Cohort

Laureen Dartois^{1,2,3}, Guy Fagherazzi^{1,2,3}, Marie-Christine Boutron-Ruault^{1,2,3}, Sylvie Mesrine^{1,2,3}, and Françoise Clavel-Chapelon^{1,2,3}

Abstract

Although some modifiable lifestyle characteristics have been associated with decreased cancer risk, little is known about their combined effect or about the proportion of cancer cases that could be prevented by improving lifestyle behaviors. We aimed to quantify the association between lifestyle habits and all-site and site-specific cancer risk in middle-aged women. The study included 64,732 women from the French E3N prospective cohort, ages 43 to 68 years at baseline. During a 15-year follow-up period, 6,938 cases of invasive cancer were diagnosed. We defined an index that aggregated five lifestyle characteristics: smoking, body mass index, alcohol consumption, fruit and vegetable consumption, and physical activity. Proportional hazard Cox regressions were performed to evaluate the association between lifestyle and cancer risk and to estimate multivariate HRs and their 95% confidence intervals (CI). In addition, population-attributable fractions were used to estimate the proportion of cancer cases that could be prevented by healthier behaviors. A significant decrease in all-site cancer risk was observed and was associated with a healthy lifestyle (HR, 0.81; 95% CI, 0.73–0.89 when comparing the highest with the lowest health index category; P_{trend} across categories < 0.01). Combining all five characteristics would have prevented 6.3% (2.2%–10.3%) of any-site, 6.3% (0.5%–12.1%) of postmenopausal breast, and 47.5% (26.8%–64.1%) of lung cancers. In conclusion, compliance with only five modifiable lifestyle behaviors could prevent a significant number of cancers, notably postmenopausal breast and lung cancers. *Cancer Prev Res*; 7(5): 516–25. ©2014 AACR.

Introduction

Cancer is the primary cause of death in developed countries and is the second-most common cause of death in developing countries (1). Among women, the global cancer incidence was 164.4 for 100,000 women-years in 2008 (137.2 in less-developed regions and 226.3 in more-developed regions), and in France, cancer incidence reached 254.9 for 100,000 women-years in 2008 (1). Lifestyle characteristics, such as not smoking, maintaining a normal weight, limiting alcohol intake, adopting specific dietary habits such as a high fruit and vegetable consumption, and exercising regularly have been associated with decreased mortality (2) and a decreased risk of cancer (2), cardiovascular disease (3, 4), and diabetes (5), and have then been recognized as healthy behaviors. National public health recommendations have been established to make the gen-

eral public aware of improved health expectancy and decreased cancer risk (6–8) when adopting a healthy lifestyle combining these behavioral habits.

The individual effects of these healthy lifestyle characteristics have been widely explored. Smoking has been associated with lung, laryngeal and pharyngeal, upper digestive tract and oral, bladder, kidney, pancreatic, liver, and cervical cancer risk (9). High body mass has been associated with endometrial, postmenopausal breast, renal, thyroid, pancreatic, colon, and esophageal cancer risk (10). Alcohol consumption has been associated with head and neck, esophageal, liver, colorectal, and breast cancer risk (11), whereas consumption of fruit and/or vegetables has been inversely associated with head and neck, esophageal, stomach, and colorectal cancer risk (2). Physical activity has been inversely associated with colorectal, breast, and endometrial cancer risk (2). However, as compliance with healthy behaviors is clustered (12, 13), these five healthy behaviors should be considered simultaneously to take into account their associations. However, the combined effect of these behaviors has not been well documented in the literature (14). The few studies that have investigated the combination of five (15–17) or four (18–28) such behaviors have reported a reduced all-site cancer risk (17–21) as well as a reduced risk of colorectal (23, 24), pancreatic (15), and postmenopausal breast (25) cancer and a reduced risk of cancer mortality (16, 26–28) when comparing the healthiest subjects with the less healthy ones. In addition, very few

Authors' Affiliations: ¹Inserm (Institut National de la Santé et de la Recherche Médicale), Centre for Research in Epidemiology and Population Health (CESP); ²Univ Paris Sud; and ³Gustave Roussy, Villejuif, France

Note: Supplementary data for this article are available at *Cancer Prevention Research Online* (<http://cancerprevres.aacrjournals.org/>).

Corresponding Author: Françoise Clavel-Chapelon, Inserm, Centre for research in Epidemiology and Population Health (CESP), 114 rue Edouard Vaillant, 94805 Villejuif, France. Phone: +331 42 11 41 48; Fax: +331 42 11 40 00; E-mail: francoise.clavel@gustaveroussy.fr

doi: 10.1158/1940-6207.CAPR-13-0325

©2014 American Association for Cancer Research.

studies have estimated the proportion of cases that could be prevented if exposure to one or several risk factors was removed from the population (15–19, 23, 28). Taking simultaneously into account the prevalence of the studied exposures and their association with the disease permits the computation of the population-attributable fraction (PAF), which quantifies the burden of a disease and helps focus prevention strategies. The aim of the present study was to assess the proportion of all-site and site-specific cancer cases that would have been prevented if all women had followed a healthy lifestyle based on five modifiable characteristics.

Materials and Methods

Study population

The E3N [Étude Épidémiologique auprès des femmes de la Mutuelle Générale de l'Éducation Nationale (MGEN)] prospective cohort was launched in 1990 to investigate cancer risk factors (29). The cohort includes 98,995 French women who were regularly asked about their health status and lifestyle. Self-administered questionnaires were sent every 2 to 3 years to update the information. All women signed informed consent, in compliance with the rules of the French National Commission for Data Protection and Privacy, from which approval was obtained.

For the present study, follow-up started at the date of return of the 1993 questionnaire, which first recorded dietary habits, thereafter considered as baseline. Responders ($N = 74,522$) contributed person-years of follow-up until the date of any cancer diagnosis other than basal cell skin carcinoma, the date of the last completed questionnaire or the date the last available questionnaire was mailed (June 2008), whichever occurred first. We excluded women diagnosed with a cancer before inclusion ($N = 4,705$), those with no follow-up questionnaire ($N = 769$), cancer cases with no specified date of diagnosis ($N = 54$), and women who did not report health behaviors at baseline ($N = 4,262$). Our final population for analysis consisted of 64,732 women ages 43 to 68 years at baseline.

Data collection

At baseline, women were asked to fill in a dietary questionnaire using quantitative and qualitative estimates of consumed items, including alcohol consumption and fruit and vegetable consumption. Body mass index (BMI) was computed as weight/height² in kg/m², using self-reported height and weight. Validation studies, conducted to determine the accuracy of the reported anthropometric measurements and dietary data, demonstrated the reliability of the reported data (30, 31). Walking, cycling, and sports activities were combined to derive overall recreational physical activity. Durations were averaged over the summer and winter. The assigned metabolic equivalent task (MET) values were 3.0 for walking, 6.0 for cycling and other sports (32).

Construction of the health index

The five lifestyle behaviors considered in the present analysis included tobacco smoking, BMI, alcohol consumption, fruit and vegetable consumption, and recreational physical activity, all assessed at baseline. An *a priori* score was developed according to how well participants adhered to health recommendations expressed by the PNNS (the French national program for health and nutrition; ref. 6), the ANSES (the French food safety agency; ref. 7), and the World Health Organization (8). Each score ranked from 0 (not meeting the recommendation) to 1 (full compliance), with partial compliance scored as 0.5. As presented in Table 1, fully compliant women were never smokers, had a BMI within the 18.5 to 25 normal range, drank less than one alcoholic drink per day, consumed at least 5 servings of fruit and vegetables per day, and had a recreational physical activity of at least 20 MET-hour per week (i.e., 300 minutes of moderate or 150 minutes of vigorous recreational physical activity per week; ref. 8). Partial compliance was defined as former smoking, being overweight (BMI between 25 and 30), or underweight (BMI between 16 and 18.5), consuming one to two alcoholic drinks per day, 3.5 to 5 daily servings of fruit and vegetables, and having 10

Table 1. Health index construction, based on five lifestyle characteristics

Health recommendations	Compliance with public health recommendations		
	Full (1 point)	Partial (0.5 point)	Poor (0 point)
– Smoking status	Never smoker	Former smoker	Current smoker
– BMI (kg/m ²)	[18.5–25) (normal range)	[25–30) (overweight) [16–18.5) (underweight)	≥30 (obesity) <16 (severely underweight)
– Alcohol consumption (drinks ^a /day)	<1	[1; 2)	≥2
– Fruit and vegetable consumption (servings ^b /day)	≥5	[3.5; 5)	<3.5
– Recreational physical activity level (MET-hour/week)	≥20	[10; 20)	<10

^aOne drink of alcohol corresponds to 10 g of ethanol.
^bOne serving of fruits and vegetables corresponds to 80 g of fruits and vegetables. Fruits and vegetables included raw and cooked vegetables, lettuce, and fresh fruits.

Dartois et al.

to 20 MET-hour weekly recreational activity (i.e., between 150 and 300 minutes of moderate or between 75 and 150 minutes of vigorous recreational physical activity per week; ref. 8). If women did not meet any of the above criteria, they were considered noncompliant.

We assigned a health index to participants by summing up the individual scores of each lifestyle behavior. Consequently, the health index ranged from 0 (the least healthy) to 5 (the most healthy), with intervals of 0.5.

Ascertainment of cancer cases

All questionnaires enquired about occurrence of any cancer, type of cancer, addresses of physicians, and permission to contact them. Cancer incidence data were coded according to the 10th revision of the International Classification of diseases (ICD-10) and the 3rd edition of the International Classification of Diseases for Oncology (ICD-O-3). We considered only invasive cancers (ICD-O-3 behavior code of 3) and censored *in situ* tumors with no concomitant invasive tumor at the date of diagnosis. In addition to reported cancer cases, some cases (5.0% of all included invasive cancer cases) were identified from information on the causes of death obtained from the French National Service on Causes of Death; when the precise date of diagnosis could not be retrieved, we recorded the date of diagnosis as 1 year before the date of death, according to the median survival time after diagnosis of late-stage cancers in our cohort. Invasive cancer cases were confirmed by pathology reports or death certificates, obtained for 86.1% of our population cases. For analyses, we considered breast (ICD-10 C50), nonbasal skin (ICD-10 C43-C44, excluding ICD-O-3 M809-M811), colorectal (ICD-10 C18-C20), digestive other than colorectal (ICD-10 C15-C17 and C21-C26), endometrial (ICD-10 C54), thyroid (ICD-10 C73), ovarian (ICD-10 C56), and lung (ICD-10 C34) cancers as well as hematopoietic malignancies (ICD-10 C81-C96). Bladder, kidney, head and neck, brain, cervical cancers, other sites (such as eye and anus), and unknown sites of cancer were not included, because of the low number of cases.

Statistical methods

Cox proportional hazards regression models with age as the timescale (33) were used to estimate HRs and 95% confidence intervals (CI) of all-site and site-specific cancer risks associated with the healthiest behaviors compared with the less healthy behaviors. Multivariate models were adjusted for potential effect modifiers, i.e., level of education (undergraduate, graduate, and postgraduate), residence (North, Center, South of France), first-degree family history of any cancer (at least one, none), professional activity (active, retired/never worker), use of oral contraceptives (ever, never), menopausal status and use of menopausal hormone therapy (premenopausal, postmenopausal with current hormone therapy, and postmenopausal with no current hormone therapy), age at menarche (quartiles) and number of children and age at first full-term pregnancy (no child, one child before age 30, one child after age 30, more than one child with the first before age 30,

more than one child with the first after age 30). Models for the individual effect of each of the five lifestyle characteristics scores were further adjusted for the four other characteristics. Information on menopausal status and use of menopausal hormone therapy was updated at each questionnaire, whereas all other potential effect modifiers were assessed at baseline. Analyses were further stratified by generation (1925–1930, 1931–1935, 1936–1945, and 1946–1950) using the "strata" option in the SAS "proc phreg" command, to consider a possible cohort effect (34). We used competitive risk Cox models to assess the risk associated with specific cancer types to deal with multiple censoring types. Trends for the health index were assessed, incorporating the health index as a continuous variable in the models.

For each potential effect modifier, missing values represented less than 5% of observations and were then all imputed to the modal category for categorical variables and the median value for quantitative variables.

Estimation of the population-attributable fractions

To quantify the proportion of cancer cases that could have been avoided if the population distribution of some risk factors was changed (all other modifiable and nonmodifiable risk factors remaining unchanged), we estimated the PAF associated with a specific modification in health behaviors with the assumption of a causal relationship. Point estimates and 95% CIs were evaluated using a method described by Spiegelman and colleagues (35). The estimation of PAFs took into account exposure prevalence and HRs of cancer risk associated with exposure and potential effect modifiers. CIs were estimated using the multivariate delta method (35). The PAF of each healthy behavior was estimated separately and expressed as the percentage of cancer cases that would have been prevented if all women had been in the healthiest category for each behavior (i.e., lifelong smoking abstinence, BMI between 18.5 and 25, drinking less than one alcohol unit per day, eating at least 5 daily servings of fruits and vegetables, or weekly exercise of at least 20 MET-hour), the four other characteristics remaining unchanged. In addition, we estimated the PAF corresponding to the combination of the five behaviors, expressing the percentage of cancer cases that would have been prevented if all women had scored 4.5 or 5 for the health index (i.e., entirely adhering to four health recommendations and partly or entirely adhering to the fifth one). A positive PAF quantified the percentage of cancer cases that would have been prevented; a negative PAF represented the percentage of cancer cases that would have additionally occurred.

Sensitivity analyses

We tested a reverse-causation hypothesis by censoring cancer cases occurring within the first 5 years of follow-up. To analyze the effect of the threshold choices, analyses were performed after modifying the categorization of each individual score, i.e., considering underweight (BMI between 16 and 18.5) as full compliance and severe underweight (BMI < 16) as full or partial compliance; alcohol abstinence as full

compliance and consumption of less than one alcoholic drink daily as partial compliance; modifying the threshold of 3.5 servings of fruits and vegetables daily to 2.5 or 3 servings a day; and modifying the thresholds of 10 to 20 MET-hour/week to 15 to 20, 15 to 25, 15 to 30, 10 to 25, or 10 to 30 MET-hour/week. We tested the potentially confounding effect of sun exposure for nonbasal skin cancer analyses by adjusting the effects for the yearly mean sun UV dose of the residential town at baseline. In addition, we performed an analysis separating women according to their age (less than 50, more than 50) to assess effects of lifestyle according to genetic predispositions. Because the etiology of melanoma and other types of nonbasal skin cancer differs, we analyzed separately the different types of nonbasal skin cancer (melanoma, spinocellular, or other/unspecified type). Finally, we performed a sensitivity analysis excluding nonvalidated cancer cases to avoid misclassification of cases. All analyses were conducted using SAS software, version 9.2 (SAS Institute Inc.).

Results

Population characteristics

During 840,097 person-years (median follow-up of 8 years for invasive cancer cases and 15 years for noncases), 6,938 women were diagnosed with any type of invasive cancer. Specific types (including concomitant tumors) included 3,483 breast cancers (609 premenopausal and 2,874 postmenopausal at diagnosis), 686 nonbasal skin cancers, 481 colorectal cancers, 411 hematopoietic malignancies (including mostly non-Hodgkin and Hodgkin lymphoma, leukemia, and multiple myeloma), 274 noncolorectal digestive cancers (including pancreatic, stomach, liver, biliary tract, esophageal, and small bowel cancers), 270 endometrial cancers, 267 thyroid cancers, 248 ovarian cancers, and 213 lung cancers.

The main characteristics of the population according to health index categories are presented in Supplementary Table S1. Women in the highest category of the health index were more likely to live in the south of France, to be retired/never have worked, to be parous, to be postmenopausal, and to have ever used menopause hormone therapy. Table 2 presents correlations between the five lifestyle characteris-

tics. Pearson correlation coefficients suggested a significant correlation for all pairs ($P < 0.05$), except between smoking and BMI ($P = 0.16$).

Risk of cancer

Because age-adjusted and fully adjusted estimates provided very close figures, only the latter will be addressed below.

All-site and site-specific cancer risks are presented in Table 3. Compared with women with the lowest compliance (health index 0–2), women with the best compliance to health recommendations (health index between 4.5 and 5) exhibited a reduced all-site cancer risk: HR (95% CI) = 0.81 (0.73–0.89), $P_{\text{trend}} < 0.001$ across score categories.

The association was also observed for several specific tumor sites: lung cancer [HR = 0.19 (0.11–0.30), $P_{\text{trend}} < 0.001$], endometrial cancer [HR = 0.45 (0.29–0.71), $P_{\text{trend}} < 0.001$], digestive cancers other than colorectal [HR = 0.59 (0.36–0.95), $P_{\text{trend}} = 0.036$], and colorectal cancer [HR = 0.66 (0.45–0.97), $P_{\text{trend}} = 0.013$]. A significant trend was observed for postmenopausal breast cancer ($P_{\text{trend}} = 0.016$); however, the reduced risk for women with the best compliance compared with those with the lowest compliance was of borderline significance: HR = 0.87 (0.74–1.03). On the opposite, we observed an increased cancer risk when comparing women with the best compliance with women with lowest compliance for nonbasal skin cancer [HR = 1.75 (1.17–2.62), $P_{\text{trend}} < 0.001$]. Associations between the risk of other cancer sites (premenopausal breast, hematopoietic, thyroid, and ovarian cancers) and compliance to a healthy lifestyle did not reach significance.

Population-attributable fractions

PAFs of various tumor sites according to lifestyle characteristics, considered separately and in combination, are presented in Table 4.

Overall, if all women had followed a healthy lifestyle, i.e., had an index between 4.5 and 5, a total of 6.3% (2.2%–10.3%) of any-site cancer cases would have been prevented. The proportion of preventable cases was 6.3% (0.5%–12.1%) for postmenopausal breast cancer and 47.5% (26.8%–64.1%) for lung cancer. However, if all

Table 2. Pearson correlation coefficients between the five lifestyle characteristics scores

	Smoking status	BMI	Alcohol consumption	Fruit and vegetable consumption	Physical activity level
Smoking status	1	–0.00557	0.17957 ^a	0.05515 ^a	–0.00870 ^b
BMI		1	–0.01827 ^a	–0.01829 ^a	0.08793 ^a
Alcohol consumption			1	0.05885 ^a	–0.04013 ^a
Fruit and vegetable consumption				1	0.08809 ^a
Recreational physical activity level					1

^a P value < 0.001.

^b P value < 0.01.

Dartois et al.

Table 3. Association between the adherence to public health recommendations and the risk of any-site and site-specific cancer, E3N cohort ($N = 64,732$)

	Health index ^a categories				P_{trend}
	[0; 2]	[2.5; 3]	[3.5; 4]	[4.5; 5]	
All-site cancer ($N = 6,938$)					
All, number of cases (%)	450 (6.48)	1,554 (22.40)	2,974 (42.87)	1,960 (28.25)	
HR (95% CI) ^b	1.00 (reference)	0.88 (0.79–0.98)	0.86 (0.78–0.95)	0.81 (0.73–0.89)	<0.001
Breast cancer ($N = 3,483$)					
Premenopausal, number of cases (%)	47 (7.72)	141 (23.15)	268 (44.01)	153 (25.12)	
HR (95% CI) ^b	1.00 (reference)	0.81 (0.59–1.13)	0.90 (0.66–1.22)	0.80 (0.58–1.12)	0.288
Postmenopausal, number of cases (%)	169 (5.88)	623 (21.68)	1,284 (44.68)	798 (27.76)	
HR (95% CI) ^d	1.00 (reference)	0.95 (0.80–1.12)	0.99 (0.85–1.17)	0.87 (0.74–1.03)	0.016
Nonbasal skin cancer ($N = 686$)					
All, number of cases (%)	26 (3.79)	125 (18.22)	283 (41.25)	252 (36.74)	
HR (95% CI) ^b	1.00 (reference)	1.22 (0.80–1.87)	1.40 (0.94–2.09)	1.75 (1.17–2.62)	<0.001
Colorectal cancer ($N = 481$)					
All, number of cases (%)	33 (6.86)	110 (22.87)	207 (43.04)	131 (27.23)	
HR (95% CI) ^b	1.00 (reference)	0.83 (0.56–1.22)	0.76 (0.53–1.10)	0.66 (0.45–0.97)	0.013
Hematopoietic cancer ($N = 411$)					
All, number of cases (%)	18 (4.38)	97 (23.60)	160 (38.93)	136 (33.09)	
HR (95% CI) ^b	1.00 (reference)	1.33 (0.80–2.20)	1.08 (0.66–1.76)	1.27 (0.77–2.08)	0.894
Digestive cancer ^e ($N = 274$)					
All, number of cases (%)	22 (8.03)	63 (22.99)	115 (41.97)	74 (27.01)	
HR (95% CI) ^b	1.00 (reference)	0.72 (0.44–1.16)	0.65 (0.41–1.04)	0.59 (0.36–0.95)	0.036
Endometrial cancer ($N = 270$)					
All, number of cases (%)	26 (9.63)	66 (24.44)	105 (38.89)	73 (27.04)	
HR (95% CI) ^b	1.00 (reference)	0.61 (0.39–0.97)	0.48 (0.31–0.73)	0.45 (0.29–0.71)	<0.001
Thyroid cancer ($N = 267$)					
All, number of cases (%)	16 (5.99)	65 (24.34)	103 (38.58)	83 (31.09)	
HR (95% CI) ^b	1.00 (reference)	1.08 (0.62–1.87)	0.92 (0.54–1.56)	1.08 (0.63–1.86)	0.573
Ovarian cancer ($N = 248$)					
All, number of cases (%)	15 (6.05)	59 (23.79)	101 (40.73)	73 (29.43)	
HR (95% CI) ^b	1.00 (reference)	0.99 (0.56–1.75)	0.87 (0.51–1.51)	0.89 (0.51–1.55)	0.297
Lung cancer ($N = 213$)					
All, number of cases (%)	32 (15.02)	61 (28.64)	85 (39.91)	35 (16.44)	
HR (95% CI) ^b	1.00 (reference)	0.47 (0.31–0.72)	0.33 (0.22–0.49)	0.19 (0.11–0.30)	<0.001

NOTE: Bold values signify P -values < 0.05.^aThe health index ranged from 0 (the less healthy) to 5 (the healthiest), with an interval of 0.5 and was calculated by summing the five individual scores (0, 0.5, and 1) based on recommendations on smoking, BMI, alcohol consumption, fruit and vegetable consumption, and physical activity.^bAdjusted for level of education, residence, first-degree family history of any cancer, professional activity, use of oral contraceptives, age at menarche and number of children, age at first full-term pregnancy, menopausal status, and use of menopausal hormone therapy.^cAdjusted for level of education, residence, first-degree family history of any cancer, professional activity, use of oral contraceptives, age at menarche and number of children, and age at first full-term pregnancy.^dAdjusted for level of education, residence, first-degree family history of any cancer, professional activity, use of oral contraceptives, age at menarche and number of children, age at first full-term pregnancy, and use of menopausal hormone.^eNoncolorectal digestive cancers included cancers of the pancreas, stomach, liver, biliary tract, esophagus, and small bowel.

women had followed a healthy lifestyle, an excess of 21.5% (7.5%–34.6%) of nonbasal skin cancers cases would have occurred. This proportion was not significant for the other sites of cancer.

The proportions of any-site and site-specific cancer cases that would be prevented by following one of the four other

healthy behaviors varied according to cancer site. If all women had been lifelong smoking abstainers, 41.8% (24.8%–56.3%) of lung cancer cases would have been prevented; if all women had been within a healthy range of BMI (i.e., between 18.5 and 25 kg/m²), 13.3% (3.7%–22.5%) of endometrial cancer cases and 8.6% (0.2%–16.9%) of

Table 4. PAFs of any-site and site-specific cancer when adhering to public health recommendations, E3N cohort (N = 64,732)

	PAF for each single characteristic ^a					PAF for the five combined characteristics ^b
	Smoking	BMI	Alcohol consumption	Fruit and vegetable consumption	Physical activity	
All-type cancer (N = 6,938)	1.3 (-1.8 to 4.4)	1.4 (-0.5 to 3.4)	3.2 (0.7-5.7)	1.5 (-0.8 to 3.8)	0.0 (-2.8 to 2.8)	6.3 (2.2-10.3)
Breast cancer (N = 3,483)						
Premenopausal (N = 609)	2.0 (-5.2 to 9.3)	-6.3 (-9.9 to -2.6)	2.1 (-3.4 to 7.6)	2.5 (-3.0 to 8.0)	-1.0 (-7.4 to 5.3)	-2.3 (-12.0 to 7.5)
Postmenopausal (N = 2,874)	-2.6 (-7.0 to 1.8)	-0.3 (-3.2 to 2.6)	5.2 (1.6-8.9)	2.4 (-0.5 to 5.6)	-0.8 (-4.8 to 3.3)	6.3 (0.5-12.1)
Nonbasal skin cancer (N = 696)	-4.5 (-14.1 to 5.3)	-2.0 (-8.0 to 4.1)	-3.6 (-11.5 to 4.3)	-0.6 (-7.8 to 6.5)	-10.6 (-19.4 to -1.7)	-21.5 (-34.6 to -7.9)
Colorectal cancer (N = 481)	7.0 (-5.0 to 18.8)	1.0 (-6.3 to 8.2)	6.6 (-2.9 to 16.0)	5.1 (-4.1 to 14.1)	-5.0 (-15.5 to 5.6)	12.8 (-2.4 to 27.4)
Hematopoietic cancer (N = 411)	-0.7 (-13.2 to 11.8)	8.6 (0.2-16.9)	-1.0 (-10.9 to 9.0)	-4.6 (-13.3 to 4.0)	-2.9 (-14.3 to 8.6)	-6.7 (-23.6 to 10.6)
Digestive ^c cancer (N = 274)	3.9 (-11.7 to 19.3)	6.1 (-3.8 to 15.9)	2.5 (-10.6 to 15.4)	1.7 (-9.5 to 12.8)	2.0 (-12.3 to 16.3)	12.1 (-8.4 to 31.6)
Endometrial cancer (N = 270)	0.7 (-15.6 to 17.0)	13.3 (3.7-22.9)	9.8 (-3.1 to 22.3)	1.6 (-10.3 to 13.5)	2.9 (-11.9 to 17.6)	15.0 (-5.8 to 34.6)
Thyroid cancer (N = 267)	-3.6 (-18.1 to 11.0)	4.6 (-6.0 to 15.1)	-7.8 (-19.7 to 4.2)	1.2 (-10.0 to 12.4)	3.0 (-11.3 to 17.1)	-7.5 (-28.7 to 14.4)
Ovarian cancer (N = 248)	4.5 (-11.6 to 20.3)	7.2 (-3.5 to 17.8)	-0.6 (-13.9 to 12.8)	0.5 (-11.4 to 12.3)	4.7 (-10.4 to 19.6)	2.9 (-18.8 to 24.3)
Lung cancer (N = 213)	41.8 (24.8-56.3)	-5.7 (-16.6 to 5.4)	1.2 (-14.2 to 16.5)	5.4 (-8.5 to 19.1)	22.5 (6.3-37.9)	47.5 (26.8-64.1)

NOTE: Bold values signify P-values < 0.05.

^aIndividual PAFs estimated the percentage of cancer cases that would be prevented if all women had entirely adhered to one single health recommendation (individual score of 1 for the health behavior), adjusted for level of education, residence, first-degree family history of any cancer, professional activity, use of oral contraceptives, menopausal status and use of menopausal hormone therapy, age at menarche, number of children and age at first full-term pregnancy, and the four other characteristics. Analyses on premenopausal breast cancer were not adjusted for menopausal status or use of menopausal hormone therapy. Analyses of postmenopausal breast cancers were not adjusted for menopausal status.^bPercentage of cancer cases that would be prevented if all women had followed a healthy lifestyle (health index between 4.5 and 5), adjusted for all above variables, except the five lifestyle characteristics.^cNoncolorectal digestive cancers included cancers of the pancreas, stomach, liver, biliary tract, esophagus, and small bowel.

Dartois et al.

hematopoietic malignancies cases would have been prevented, whereas 6.3% (2.6%–9.9%) premenopausal breast cancer cases in excess would have occurred. If all women had consumed less than one alcoholic drink daily, 3.2% (0.7%–5.7%) of any-site cancer and 5.2% (1.6%–8.9%) of postmenopausal breast cancer cases would have been prevented. If all women had achieved a high level of physical activity, 22.5% (6.3%–37.5%) of lung cancer cases would have been prevented, whereas 10.6% (1.7%–19.4%) skin cancer cases in excess would have occurred; no other association with a statistically significant PAF was observed.

Sensitivity analyses

To test a reverse causation hypothesis, we censored cases diagnosed within the first 5 years of follow-up ($N = 1,897$). A decreased risk of any-site cancer was nonetheless observed when comparing women with the healthiest behaviors with those with the lowest index: HR = 0.85 (0.75–0.96). The percentage of prevented cases was 5.4% (0.6%–10.2%). In addition, when censoring non validated cancer cases ($N = 963$), associations persisted, with HR = 0.79 (0.71–0.88) and 6.2% (2.0%–10.8%) of any-site cancers prevented. Analyses of nonbasal skin cancer adjusted for sun UV dose of the baseline residential town provided similar results to the main analyses, with HR = 1.65 (1.03–2.64) and 19.50% (6.99%–31.30%) additional diagnosed nonbasal skin cancer cases when adopting healthy behaviors. Associations between recreational physical activity and nonbasal skin cancer remained with a PAF estimated at –9.16% (–17.00% to –1.21%). Analyses considering the subtypes of nonbasal skin cancer (875 cancers in 686 women) resulted in similar results for melanomas ($N = 391$) and other/unspecified types of nonbasal skin cancers ($N = 257$), with HR = 1.44 (0.88–2.37) and PAF = –20.8% (–38.2% to –2.0%) and HR = 2.41 (1.12–5.20) and PAF = –28.3% (–49.0% to –4.5%), respectively, but no association was observed for spinocellular cancers ($N = 227$), with HR = 1.25 (0.62–2.51) and PAF = –8.1% (–30.0% to 14.6%). When performing separate analyses according to age, results remained similar in women over 50 years old (6,304 cancers diagnosed during 705,771 women-years) and those 43 to 50 years old (634 cancers diagnosed during 56,326 women-years) with HR = 0.82 (0.73–0.92) and 0.72 (0.54–0.97), respectively. The PAF in the oldest women was 6.8% (2.6%–11.0%), although in the youngest women, due to the shorter follow-up duration, the PAF barely reached –1.0% (–9.9% to 7.9%). Modifying the thresholds of the lifestyle factor scores resulted in HRs of between 0.76 and 0.84 (all statistically significant at the 5% level) for any-site cancer risk associated with the highest compared with the lowest category of the modified health index, and resulted in a percentage of preventable cancer cases of between 4.9% and 7.5%.

Discussion

In this large prospective cohort of French women, we observed that adhering to one single health recommenda-

tion had a moderate impact on the proportion of prevented cancer cases. However, adhering more closely to a healthy lifestyle, by respecting recommendations for smoking, BMI, alcohol consumption, fruit and vegetable consumption, and physical activity (6–8), could reduce the number of any-site cancer cases by up to 6.3%, a percentage that varied across cancer sites, reaching 47.5% for lung cancer. Behaviors associated with the highest PAF included smoking for lung cancers, BMI for endometrial cancers and hematopoietic malignancies, alcohol consumption for postmenopausal breast cancers, and physical activity for lung and skin cancers. Cancer is a multifactorial disease that involves multiple endogenous (genetics, hormones, immunity, and metabolism) and exogenous (lifestyle and environment) causes. Therefore, the analyzed healthy behaviors, chosen to be easily modifiable lifestyle factors, are five of multiple factors related to cancer.

Literature confrontation

The few studies that quantified the impact of lifestyle on cancer risk reported a decreased risk of any-site (17–22), colorectal (19, 22–24), breast (19, 22, 25), endometrial (19, 22), digestive (15, 19), and lung (19) cancer in women complying with public health recommendations. Compliance with recommendations was estimated to be able to prevent 11% to 31% of cancer (17–19) and, specifically, 6% of colorectal cancer (23). The variation of exposure throughout the world and the use of different estimation methods could contribute to the variability of the findings. The widely used Levin formula (36) was demonstrated to provide biased estimates when calculated with multivariate relative risks (37), whereas the Spiegelman formula (35) seems to be better adapted to estimate-adjusted PAF.

The healthy lifestyle characteristics analyzed in the present study seemed to play a minor role in the etiology of thyroid cancer, for which radiation exposure is the only well-established risk factor (38). The strong association observed between tobacco smoking and lung cancer is in line with a recent metaanalysis (39) and emphasizes the causal nature of the relationship. Surprisingly, we did not observe that physical activity can prevent colorectal or postmenopausal breast cancer cases, despite convincing evidence of the association between physical activity and these cancer sites (2). The increased number of skin cancer cases that would have occurred if all women had been highly physically active suggests residual confounding by individual sun exposure habits through outdoors activities, which has been associated with an increased risk of skin cancer (40), and underlines the importance of recommending that patients avoid sunburn and excessive sun exposure without protection (41). We observed that approximately 22% of lung cancer cases would have been prevented if all women had achieved a high physical activity, in agreement with the inverse relationship between risk of lung cancer and physical activity found in metaanalyses (42, 43). However, the association may be a reflection of reverse causation, as people with chronic

lung disease who are at an increased risk of lung cancer may be less prone to practicing physical activity (2). In agreement with the individual effects of the analyzed healthy behaviors on cancer risk observed in this study, previously published metaanalyses concluded that women should limit alcohol consumption to prevent an increased risk of breast cancer (44, 45), that BMI was inversely associated with the risk of premenopausal breast cancer (46), and that lower BMI was associated with a decreased risk of endometrial cancer (47). Low BMI was also associated with lower risk of non-Hodgkin lymphoma (48), leukemia (49), and multiple myeloma (50), three hematopoietic malignancies with very few established risk factors (51). Finally, consistent with our results, other studies observed that weight management, physical activity, alcohol consumption, and smoking likely explained most of the observed association between healthy lifestyle scores and cancer risk (17, 19, 21, 22). The results remained unchanged when excluding cases diagnosed within the first 5 years of follow-up and when excluding women younger than 50 years. These results suggest that reverse causation was unlikely and that lifestyle may play a role in the etiology of cancer in older women, which is less related to genetic factors.

Strengths and limits

The study has some limitations. The choice of a threshold for creating a health index may be disputable, although our approach was chosen to be as close as possible to public health recommendations (6–8). However, our results are strengthened by the sensitivity analyses, in which the health dimension categories were modified but produced similar results. In addition, the E3N cohort, like most cohorts of volunteers, is prone to a “healthy cohort effect.” As the impact of lifestyle varies according to the prevalence of exposure, higher PAFs can be expected in the general population. The nonsignificant associations observed for several site-specific cancer sites might be a reflection of a lack of statistical power due to a small number of cases. As the values of the five characteristics taken into account in the construction of the health index were not updated at each questionnaire, this approach cannot capture the effect of adopting healthy behaviors over time. The estimation of PAFs was based on the assumption of a causal relationship between exposure and cancer and should therefore be interpreted with caution.

The study has also several strengths. A major strength is that it was based on a large prospective cohort with an extended follow-up. The design permitted adjustment for many potential confounders and provided high statistical power and the ability to examine the impact of lifestyle on various cancer sites. Information on exposure collected before cancer diagnosis resulted in little risk of memory bias. Self-report bias was limited through the histologic confirmation of the vast majority of cancer cases. Moreover, excluding nonvalidated cases of cancer produced similar results. In addition, dietary and anthropometric data were validated (30, 31), limiting declaration bias.

An even larger cohort with longer follow-up and information on lifestyle changes would nicely complete our study and enable the identification of a key period during which changes in lifestyle may modify cancer incidence.

Conclusion

Several studies have been published using similar evaluation of lifestyle exposure and cancer outcomes, though few have quantified the impact of lifestyle on overall cancer risk using adequate methodology. While adhering to one specific recommendation has a moderate impact on the proportion of prevented cancer cases, combining healthy behaviors—being a nonsmoker, drinking moderately, consuming the recommended amounts of fruits and vegetables, being physically active, and having a BMI within the recommended range—may substantially decrease the incidence of some specific cancers, especially lung and postmenopausal breast cancers. These results should encourage research into ways of enforcing these five simple health behaviors in the general population to improve cancer prevention.

Ethical Approval

The E3N study received approval from the national commission overseeing ethical data collection in France (“Commission Nationale de l’Informatique et des Libertés”), classification AG662, 12489 (September 25, 1988). All women signed an informed consent.

Data Sharing

The dataset can be requested from the principal investigator of the E3N study at francoise.clavel@gustaveroussy.fr.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: L. Dartois, G. Fagherazzi, S. Mesrine
Development of methodology: L. Dartois, G. Fagherazzi
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): F. Clavel-Chapelon
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Dartois, G. Fagherazzi, S. Mesrine
Writing, review, and/or revision of the manuscript: L. Dartois, G. Fagherazzi, M.-C. Boutron-Ruault, S. Mesrine, F. Clavel-Chapelon
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L. Dartois
Study supervision: M.-C. Boutron-Ruault, F. Clavel-Chapelon

Acknowledgments

The authors thank all participants for providing data and practitioners for providing pathology reports. They also thank all members of the E3N-EPIC study group, particularly Rafika Chait, for the management of cancer databases.

Grant Support

The E3N study was funded by the MGEN, the Ligue contre le Cancer, Gustave Roussy Institute, and the Institut National de la Santé et de la Recherche Médicale (Inserm), and received grants from the “Cohortes Santé TIGR Program,” Agence Nationale de la Recherche (ANR), Institut National du Cancer (INCa), Fondation ARC pour la recherche sur le cancer, World Cancer Research Fund (WCRF), Ligue contre le Cancer, Fondation de France, ANSES, the French Ministry of Health, and the Université Paris Sud.

Dartois et al.

M.-C. Boutron-Ruault and F. Clavel-Chapelon were supported by Inserm, L. Dartois by the Fondation de France (2011-00023087), S. Mesrine by the Université Paris Sud, and G. Fagherazzi by the Fondation ARC.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked

advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 10, 2013; revised January 27, 2014; accepted February 15, 2014; published OnlineFirst February 26, 2014.

References

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN2008. *Int J Cancer* 2010;127:2893-2917.
2. World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: AICR; 2007.
3. Goldstein LB, Adams R, Alberts MJ, Appel LJ, Brass LM, Bushnell CD, et al. Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Stroke* 2006;113:e873-e923.
4. World Health Organization. Prevention of cardiovascular disease. Guidelines for assessment and management of cardiovascular risk. Geneva, Switzerland: WHO; 2007.
5. National Institute for Health and Clinical Excellence. Preventing type 2 diabetes—risk identification and interventions for individuals at high risk. NICE Public health guidance 38; 2012.
6. Ministère du Travail de l'Emploi et de la Santé. Programme national nutrition santé 2011-2015. PNNS 2011.
7. Agence Nationale de Sécurité Sanitaire Alimentation Environnement Travail. Nutrition et cancer. ANSES 2011.
8. World Health Organization. National cancer control programmes. Policies and managerial guidelines. 2nd ed. Geneva, Switzerland: WHO; 2002.
9. Gandini S, Botteri E, Iodice S, Boniol M, Lowerre AB, Mauboussin P, et al. Tobacco smoking and cancer: a meta-analysis. *Int J Cancer* 2008;122:155-64.
10. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371:569-78.
11. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. IARC Monogr Eval Carcinog Risks Hum 2010;963-1383.
12. Poortings W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev Med* 2007;44:124-8.
13. Wallstrom P, Wirfalt E, Janzon L, Mattiasson I, Elmstahl S, Johansson U, et al. Fruit and vegetable consumption in relation to risk factors for cancer: a report from the malmo diet and cancer study. *Public Health Nutr* 2000;3:263-71.
14. Balter K, Moller E, Fondell E. The effect of dietary guidelines on cancer risk and mortality. *Curr Opin Oncol* 2012;24:90-102.
15. Jiao L, Mitrou PN, Reedy J, Graubard BI, Hollenbeck AR, Schatzkin A, et al. A combined healthy lifestyle score and risk of pancreatic cancer in a large cohort study. *Arch Intern Med* 2009;169:764-70.
16. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. *BMJ* 2008;337:a1440.
17. Cerhan JR, Potter JD, Gilmore JM, Janney CA, Kushi LH, Lazovich D, et al. Adherence to the AICR cancer prevention recommendations and subsequent morbidity and mortality in the Iowa Women's Health Study cohort. *Cancer Epidemiol Biomarkers Prev* 2004;13:1114-20.
18. Sasazuki S, Inoue M, Iwasaki M, Sawada N, Shimazu T, Yamaji T, et al. Combined impact of five lifestyle factors and subsequent risk of cancer: the Japan Public Health Center Study. *Prev Med* 2012;54:112-6.
19. Romaguera D, Vergaud AC, Peeters PH, van Gils CH, Chan DS, Ferrati P, et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. *Am J Clin Nutr* 2012;96:150-63.
20. Meng L, Maskarinec G, Lee J, Kolonel LN. Lifestyle factors and chronic diseases: application of a composite risk index. *Prev Med* 1999;29:296-304.
21. Ford ES, Bergmann MM, Kroger J, Schienkiewitz A, Welkert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. *Arch Intern Med* 2009;169:1355-62.
22. Harnack L, Nicodemus K, Jacobs DR Jr, Folsom AR. An evaluation of the Dietary Guidelines for Americans in relation to cancer occurrence. *Am J Clin Nutr* 2002;76:889-96.
23. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjønneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *BMJ* 2010;341:c5504.
24. Odegaard AO, Koh WP, Yuan JM. Combined lifestyle factors and risk of incident colorectal cancer in a Chinese population. *Cancer Prev Res* 2013;6:360-7.
25. Hastert TA, Baresford SA, Patterson RE, Kristal AP, White E. Adherence to WCRF/AICR cancer prevention recommendations and risk of postmenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev* 2013;22:1498-508.
26. McCullough ML, Patel AV, Kushi LH, Patel R, Willett WC, Doyle C, et al. Following cancer prevention guidelines reduces risk of cancer, cardiovascular disease, and all-cause mortality. *Cancer Epidemiol Biomarkers Prev* 2011;20:1089-97.
27. Khaw KT, Wareham N, Bingham S, Welch A, Luben R, Day N. Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. *PLoS Med* 2008;5:e12.
28. Kvasavik E, Batty GD, Ursin G, Huxley R, Gale CR. Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey. *Arch Intern Med* 2010;170:711-8.
29. Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulaud H, Le CC, et al. E3N, a French cohort study on cancer risk factors. E3N Group. Etude Epidémiologique auprès de femmes de l'Education Nationale. *Eur J Cancer Prev* 1997;6:473-8.
30. Tehard B, van Liere MJ, Com NC, Clavel-Chapelon F. Anthropometric measurements and body silhouette of women: validity and perception. *J Am Diet Assoc* 2002;102:1779-84.
31. van Liere MJ, Lucas F, Clavel F, Slimani N, Villemont S. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol* 1997;26 Suppl 1:S128-36.
32. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR Jr, Tudor-Locke C, et al. 2011 Compendium of physical activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;43:1575-81.
33. Thiebaut AC, Benichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Stat Med* 2004;23:3803-20.
34. Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2006;56:159-75.
35. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control* 2007;18:571-9.
36. Levin ML. The occurrence of lung cancer in man. *Acta Univ Int Contra Cancrum* 1953;9:531-41.

37. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health* 1998;88:15-9.
38. Agate L, Lorusso L, Elisei R. New and old knowledge on differentiated thyroid cancer epidemiology and risk factors. *J Endocrinol Invest* 2012; 35:3-9.
39. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 2012;12:385.
40. Lin JS, Eder M, Weinmann S, Zuber SP, Bell TL, Plaut D, et al. behavioral counseling to prevent skin cancer: systematic evidence review to update the 2003 U.S. Preventive Services Task Force recommendation. *Ann of Intern Med* 2011;154:190-201.
41. Bataille V, de VE. Melanoma-part 1: epidemiology, risk factors, and prevention. *BMJ* 2008;337:a2249.
42. Tardon A, Lee WJ, Delgado-Rodriguez M, Dosemeci M, Albanes D, Hoover R, et al. Leisure-time physical activity and lung cancer: a meta-analysis. *Cancer Causes Control* 2005;16:389-97.
43. Buffart LM, Singh AS, van Loon EC, Vermeulen HI, Brug J, Chinapaw MJ. Physical activity and the risk of developing lung cancer among smokers: a meta-analysis. *J Sci Med Sport* 2014;17:67-71.
44. Seitz HK, Pelucchi C, Bagnardi V, La VC. Epidemiology and pathophysiology of alcohol and breast cancer: update 2012. *Alcohol Alcohol* 2012;47:204-12.
45. Bagnardi V, Rota M, Botteri E, Tramacere I, Islami F, Fedirko V, et al. Light alcohol drinking and cancer: a meta-analysis. *Ann Oncol* 2013; 24:301-8.
46. Amadou A, Ferrari P, Muwonge R, Moskal A, Bessy C, Romieu I, et al. Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. *Obes Rev* 2013;14:665-78.
47. Crosbie EJ, Zvahlen M, Kitchener HC, Egger M, Renehan AG. Body mass index, hormone replacement therapy, and endometrial cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2010;19: 3119-30.
48. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: a meta-analysis of prospective studies. *Eur J Cancer* 2011;47:2422-30.
49. Larsson SC, Wolk A. Overweight and obesity and incidence of leukemia: a meta-analysis of cohort studies. *Int J Cancer* 2008;122:1418-21.
50. Wallin A, Larsson SC. Body mass index and risk of multiple myeloma: a meta-analysis of prospective studies. *Eur J Cancer* 2011;47: 1606-15.
51. Merhi M, Raynal H, Cahuzac E, Vinson F, Cravedi JP, Gamet-Payraette L. Occupational exposure to pesticides and risk of hematopoietic cancers: meta-analysis of case-control studies. *Cancer Causes Control* 2007; 18: 1209-26.

Annexe G : « Proportion of premenopausal and postmenopausal breast cancers attributable to established risk factors: Estimates from the E3N-EPIC cohort study » - Article en revue au journal « The BMJ »

Page 1 of 34

BMJ

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Proportion of premenopausal and postmenopausal breast cancers attributable to established risk factors: Estimates from the French E3N cohort

Laureen Dartois, MSc^{1,2,3}, Guy Fagherazzi, PhD^{1,2,3}, Laura Baglietto, PhD^{4,5}, Marie-Christine Boutron-Ruault, MD, PhD^{1,2,3}, Sylvie Mesrine, MD, PhD^{1,2,3}, Françoise Clavel-Chapelon, PhD^{1,2,3}

1 Inserm (Institut National de la Santé et de la Recherche Médicale), Centre for Research in Epidemiology and Population Health (CESP), U1018, Team 9, F-94805, Villejuif, France

2 Univ Paris Sud, UMRS 1018, F-94805, Villejuif, France

3 Gustave Roussy, F-94805, Villejuif, France

4 Cancer Epidemiology Centre, Cancer Council of Victoria, Melbourne, Victoria, Australia,

5 Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Victoria, Australia

Correspondence to:

F. Clavel-Chapelon, francoise.clavel@gustaveroussy.fr

Inserm U1018 Team 9, 114 rue Edouard Vaillant, F-94805 Villejuif Cedex, France

Phone: +33 1 42 11 41 48

Fax: +33 1 42 11 40 00

Abbreviated running title: Population-attributable fraction for invasive breast cancer

Key words: Breast cancer; Women; Risk factors; Cohort study; Estrogen receptor; hormone receptor; postmenopausal women; premenopausal women; Attributable fraction; hormone replacement therapy; birth weight

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Abstract word count: 240

Text word count (excluding references): 3721

Total number of tables: 4

Total number of supplementary tables: 3

Total number of figures: 1

Total number of references: 62

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The authors are indebted to all participants for providing data and to practitioners for providing pathology reports. They are grateful to all members of the E3N-EPIC study group, particularly to Agnès Fournier for her help on hormonal treatments, Rafika Chaït for the management of cancer databases, and Siham Eltaïef, Marie Fangon, Pascale Gerbouin-Rérolle, Lyan Hoang, Celine Kemaleguen, Camille Laplanche, Maryvonne Niravong and Maxime Valdenaire for their technical assistance.

Sponsoring: The E3N study was funded by the MGEN (Mutuelle Générale de l'Éducation Nationale), the Ligue contre le Cancer, Gustave Roussy and the Institut National de la Santé et de la Recherche Médicale (Inserm), and received grants from the "Cohortes Santé TGIR Program", Agence Nationale de la Recherche (ANR), Institut National du Cancer (INCa), Fondation ARC pour la recherche sur le cancer, World Cancer Research Fund (WCRF), Ligue contre le Cancer, Fondation de France, Agence Nationale de Sécurité Sanitaire (ANSES), the French Ministry of Health and the Université Paris Sud. For the submitted work, MCBR and FCC were supported by Inserm, LD by the Fondation de France (2011-00023087), SM by the Université Paris Sud and GF by the Fondation ARC.

Ethical approval: The E3N study received approval from the national commission for computed data and liberties ("Commission Nationale de l'Informatique et des Libertés"), classification AG662, 12489 (September 25th, 1988). All women signed an informed consent.

Transparency declaration: FCC affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Data sharing: The dataset can be requested from the principal investigator of the E3N study at francoise.clavel@gustaveroussy.fr

Conflicts of interest: none declared

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Licence agreement: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article to be published in BMJ editions and any other BMJPGJ products and sublicences such use and exploit all subsidiary rights, as set out in our licence.

Contributors: LD, GF and FCC designed the analysis plan. LD performed the statistical analyses and wrote the manuscript. LD, GF, LB, MCBR, SM and FCC critically revised the manuscript and approved the final version.

ABSTRACT:

Background: Breast cancer is the most frequently diagnosed cancer among women worldwide. Risk factors for breast cancer have been widely explored individually; however, little is known about their combined impact.

Methods: The study included 67 634 women from the French E3N prospective cohort, aged 42 to 72 at baseline. During a 15-year follow-up period, 3635 cases of invasive breast cancer were diagnosed: 497 in premenopause and 3138 in postmenopause. Population-attributable fractions (PAFs) were used to estimate the proportion of breast cancer cases attributable to risk factors under various hypothetical scenarios of exposure. We examined overall premenopausal and postmenopausal invasive breast cancers, as well as tumour subtypes (ER status and HER2 expression).

Results: Premenopausal breast cancer was more frequently attributable to non-behavioural factors (61.2%) than to behavioural factors (39.9%). The contrary was observed in postmenopause (41.9% and 53.5%, respectively). Individually, the highest PAFs for non-behavioural factors were obtained for birth weight (33.6%) and age at menarche (19.8%) in premenopause and for body shape at menarche (17.1%) and personal history of benign breast diseases (14.9%) in postmenopause. In postmenopause, the highest PAFs for behavioural factors were obtained for use of hormone replacement therapy (14.5%), dietary pattern (10.1%), alcohol consumption (5.6%) and body mass index (5.1%). These proportions were higher for ER+, HER2- and ER+/HER2- postmenopausal breast cancers.

Conclusion: Our data support the hypothesis that after age 40, modifying one's behaviour can reduce the number of diagnosed cancers, even more so after menopause.

Acronyms used in text

BBD (Benign breast disease)

BMI (Body Mass Index)

CI (Confidence Intervals)

E3N (*Étude Épidémiologique des femmes de la Mutuelle Générale de l'Éducation Nationale*)

FFTP (First full-term pregnancy)

HR (Hazard Ratio)

ICD-10 (International Classification of Diseases, 10th edition)

MET (Metabolic Equivalent Task)

MGEN (*Mutuelle Générale de l'Éducation Nationale*)

MHT (Menopause Hormone Therapy)

PAF (Population-attributable fraction)

UVRd (Ultraviolet radiation doses)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

What is already known on this subject
<ul style="list-style-type: none">- The etiology of breast cancer differs according to menopausal status at diagnosis.- Many behavioural and non-behavioural risk factors of breast cancer have been established in the literature, however little is known about proportions of cancer cases attributable to these factors when considered individually or together.
What this study adds
<ul style="list-style-type: none">- A higher proportion of premenopausal breast cancer cases was attributable to non-behavioural factors, while the contrary was observed in postmenopause where behavioural factors were predominant.- Behavioural modifications on alcohol consumption diet, weight and use of menopause hormone therapy may have a large impact on postmenopausal breast cancer risk.

INTRODUCTION:

2 Breast cancer is the most frequently diagnosed neoplasm among women, with an incidence
3 reaching 43.3 per 100 000 woman-years worldwide in 2012 (1). Risk factors for breast cancer
4 have been widely explored and include reproductive and genetic factors, anthropometry, and
5 lifestyle factors. They are either of a behavioural (such as daily alcohol consumption) or a
6 non-behavioural (such as a family history of breast cancer) nature (2-4).

7 Taking into account the prevalence of the risk factors and their association with the disease
8 allows the computation of the population-attributable fraction (PAF) that quantifies the
9 burden of a disease and helps focus prevention strategies (5). PAFs enable an estimation of
10 the proportion of cases attributable to one or several exposures: the artificial removal of these
11 exposures could potentially prevent those cases. Previously published combined PAFs for
12 breast cancer risk factors have been estimated from several observational studies of women in
13 the US (6-9), Germany (10), Iran (11), Italy (12;13), Sweden (14) and Great Britain (15).
14 However, only a few of these studies assessed the combined effect of several risk factors
15 simultaneously (8;9;15), few were prospective studies exempt of recall bias (6;14), or
16 accounted for menopausal status (8-13).

17 Although recent studies have provided evidence of the differential association between risk
18 factors and breast cancer incidence according to the expression of the Human Epidermal
19 Growth Factor Receptor 2 (HER2) protein and the oestrogen-receptor (ER) status of the
20 tumour (16-21), differences in PAFs accounting for ER subtypes have been evaluated in only
21 one study (10), and no study evaluated PAFs according to HER2 subtypes.

22 Based on data from the large prospective E3N-EPIC cohort study, we estimated the PAFs of
23 several subtypes of invasive breast cancers (premenopausal, postmenopausal, ER and HER2
24 subtypes) corresponding to individual and combined behavioural and non-behavioural breast
25 cancer risk factors.

26

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

METHODS

28 *Study population*

The E3N (*Étude Épidémiologique auprès des femmes de la Mutuelle Générale de l'Éducation Nationale (MGEN)*) prospective cohort was launched in 1990 to investigate risk factors for cancer in women (22). The cohort included 98 995 French women who were regularly asked about their health status and lifestyle. Self-administered questionnaires were sent every 2 to 3 years to update the information. All women signed an informed consent form, in compliance with the rules of the French National Commission for Data Protection and Privacy from which approval was obtained.

For the present study, follow-up started at the return date of the 1993 questionnaire, which first recorded dietary habits, thereafter considered as baseline. Responders (N=74 522) contributed person-years of follow-up until the date of any cancer diagnosis other than basal cell skin carcinoma, the date of the last completed questionnaire or the date the last available questionnaire was mailed (June 2008), whichever occurred first. We excluded women diagnosed with a cancer before baseline (N=4705), those with no follow-up questionnaire (N=769), and cancer cases with no specified date of diagnosis (N=54). In addition, we excluded women with unlikely values for the ratio between energy intake and required energy (*i.e.*, the lowest or highest percentile for the cohort, N=1360). Our final population for analysis consisted of 67 634 women, aged 42-72 years at baseline.

46 *Data collection*

Women were considered postmenopausal if they reported amenorrhea for more than one year (unless due to hysterectomy) or if they self-reported that they were postmenopausal (23).

Categorizations and references used for each risk factor are specified in Table 1 and Table 2 and were based on the literature (2-4). Non-behavioural variables included risk factors on which we cannot directly act, and behavioural variables included all other risk factors.

52 Non-behavioural risk factors

53 Benign breast disease (BBD) included adenomas and fibrocystic disease. Self-reported age
54 at menarche was recorded in 1992, with an age ranking from 8 to 19 years. Low birth weight
55 was defined as below 2.5 kg, normal birth weight as between 2.5 and 4.0 kg, and high birth
56 weight as over 4.0 kg. Because preterm and low weight full-term babies have been associated
57 with different levels of breast cancer risk (24), premature babies were considered as missing
58 values for birth weight. Age at menopause was defined as age at last menstrual period, age at
59 bilateral oophorectomy, or self-reported age at menopause (23).

60 Behavioural risk factors

61 Overall physical activity included walking, gardening, home improvement, floor climbing,
62 cycling and other sports; durations were averaged over the summer and winter. The assigned
63 metabolic equivalent task (MET) values per hour were 3.0 for walking, 4.0 for gardening, 4.5
64 for home improvement, 0.067 for each floor climbed and 6.0 for cycling and other sports (25).
65 The total number of months women had spent breastfeeding was calculated by summing the
66 months of breastfeeding for each birth. Nulliparous women were considered as never having
67 breastfed. Body shape around menarche was assessed using drawings in a series of Sorensen's
68 body shapes (26).

69 At baseline, women were asked to fill in a dietary questionnaire, using quantitative and
70 qualitative estimates of consumed items (27). Dietary patterns were computed from a factor
71 analysis, based on 57 predefined food groups, using a principal component analysis. We
72 extracted the first two factors, respectively labelled "Healthy pattern" and "Western pattern"
73 (see reference (28) for further explanations), and created a four-modality variable, using the
74 median value of both of the dietary scores as thresholds. Alcohol consumption and dietary
75 vitamin D were extracted from the dietary questionnaire using a food composition table
76 derived from the updated French national database (29). Baseline regions of residence were

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

linked to a database containing mean daily ultraviolet radiation doses (UVRd in $\text{kJ/m}^2/\text{day}$) in French departments (30). Daily dietary vitamin D and UVRd were compiled into a four-modality variable, using the median value of both variables as thresholds (31).

BMI was computed as $\text{weight}/\text{height}^2$ in kg/m^2 . Lifetime menopause hormone therapy (MHT) use included any systemic use of oestrogens (alone or combined with a progestagen) or tibolone. Vaginally administered oestrogens and estriol were not considered as hormone therapy because they have not been found associated with breast cancer risk (33;34). At each questionnaire, a woman was considered as a recent user of a specific type of MHT when she reported any use within the preceding 12 months.

Ascertainment of cancer cases

All questionnaires enquired about any cancer occurrence, type of cancer, addresses of physicians and permission to contact them. Only invasive breast cancers were considered, and *in situ* tumours with no concomitant invasive tumour were censored at the date of diagnosis. Invasive breast cancer cases were confirmed by pathology reports, obtained for 91.5% of cases. Information on oestrogen-receptor (ER) status and on the expression of the human epidermal growth factor receptor 2 (HER2) were extracted from pathology reports. Invasive premenopausal breast cancers were classified accordingly into two categories, ER+ and ER-, and invasive postmenopausal breast cancer into eight non-exclusive categories: HER2+, HER2-, ER+, ER- and their combinations.

Statistical methods

Multivariate Cox proportional hazards regression models with age as the timescale (35) were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) of breast cancer risk associated with the risk factors mentioned above. Models were stratified by menopausal status because of the V-relationship between adiposity and the risk of breast cancer before and after menopause (36;37). Menopausal status, use of MHT, BMI, tobacco smoking, personal

1
2
3 102 history of BBD, and use of oral contraceptives and progestagens alone were updated at each
4
5 questionnaire and were thus considered as time dependent, whereas all other variables were
6
7
8 104 assessed at baseline, or in 2002 for birth weight. All analyses were adjusted according to
9
10 previous mammogram (in the previous follow-up period yes/no, time-dependent variable),
11
12 106 and further stratified by cohort of birth ([1925-1930], [1931-1935], [1936-1940], [1941-
13
14 1945], [1946-1950]) (38).

15
16 108 As the family history of breast cancer and the birth weight variables presented more than
17
18 5% of missing values, missing categories were created for both variables. Causes of missing
19
20 110 values for birth weight were prematurity (18.6%), no response to the questionnaire answering
21
22 about the birth weight (23.5%) and no response to the question (57.9%). For the other
23
24 112 variables, missing values were imputed to the modal category for categorical variables and to
25
26 the median value for quantitative variables.

27 28 29 114 *Estimation of the population-attributable fractions (PAF)*

30
31 To quantify the attributable proportion of cancer cases, we estimated the population-
32
33 116 attributable fraction (PAF) associated with specific modifications in health behaviours with
34
35 the assumption of a causal relationship on a particular set of risk factors, all others remaining
36
37 unchanged. Point estimates and 95% CIs were evaluated using a method described by
38
39 118 Spiegelman et al (39). The estimation of PAFs took into account exposure prevalence
40
41 throughout follow-up, and HRs of cancer risk associated with exposure and potential effect
42
43 120 modifiers.

44
45
46
47 122 PAFs were estimated separately for each factor included in the model and expressed as the
48
49 percentage of cancer cases attributed to a hypothetical scenario in which all women would be
50
51 124 in the reference category. In addition, we estimated the combined PAFs of the non-
52
53 behavioural factors, the behavioural factors and all of them simultaneously, including all
54
55 126 factors considered in the individual PAFs, whatever their significance.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

We performed stratified analyses according to receptor status. Women with unknown receptor status were excluded from the stratified analyses, and cases with other types of cancer than the one of interest were censored at the date of diagnosis. A positive PAF quantifies the percentage of potentially avoidable cancer cases, a negative PAF the percentage of potentially additional cancer cases. Because PAFs were sensitive to threshold choices (40), variables categorizations were chosen so that 20% or more of the population were in the reference category, whenever possible.

134 *Sensitivity analyses*

The influence of the imputation method was tested by excluding individuals with missing values for a specific risk factor, as a reverse-causation hypothesis by censoring cancer cases occurring within the first three years of follow-up. Because of correlations between BMI and adult height, and between dietary patterns, alcohol consumption and tobacco smoking, analyses for one variable not adjusting for the other(s) were performed. The association between birth weight and breast cancer risk was tested by performing additional analyses with further adjustment for length at birth and with premature babies considered as a separate category instead of missing values. Finally, a sensitivity analysis excluding non-validated cancer cases was performed to avoid the misclassification of cases.

144 All analyses were conducted using SAS software, version 9.2 (SAS Institute Inc., Cary, North Carolina, USA).

146 **RESULTS**

147 *Population characteristics*

148 The 67 634 women included in our analyses were aged 42 to 72 at study entry, with a mean age of 52.8 years (SD=6.6 years). During 876 468 person-years (median follow-up of 7 years for cases and 13 years for non-cases), 3635 women were diagnosed with invasive breast cancer; 497 were premenopausal and 3138 postmenopausal. Baseline characteristics of the

1
2
3 152 population and multivariate HRs from Cox regression models associated with the risk factors
4
5 used in the estimation of the PAFs are presented in Table 1 and Table 2.
6

7
8 154 *Population-attributable fractions of only one risk factor*

9
10 As univariate and multivariate estimations of PAFs produced similar results, we only
11
12 156 present results from multivariate analyses. Figure 1 presents PAFs as percentages for one risk
13
14 factor or a combination of risk factors for premenopausal and postmenopausal invasive breast
15
16 158 cancers.

17
18 **Premenopausal breast cancer.** PAFs reaching statistical significance were 14.1% for
19
20 160 personal history of BBD, 19.8% for age at menarche and 33.6% for birth weight, meaning
21
22 that approximately one out of 7, 5 and 3 premenopausal breast cancers, respectively, could be
23
24 162 attributed to these factors.

25
26
27 **Postmenopausal breast cancers.** PAFs reaching statistical significance were 3.6% for
28
29 164 family history of breast cancer, 9.7% for age at menarche and 14.9% for personal history of
30
31 BBD regarding non-behavioural factors and 5.1% for BMI, 5.4% for number of children and
32
33 age at FFTP, 5.6% for alcohol consumption, 10.1% for dietary pattern, 14.5% for use of MHT
34
35 166 and 17.1% for body shape at menarche regarding behavioural factors.

36
37
38 168 *Population-attributable fractions of combinations of several risk factors*

39
40 As many as 76.6% (-69.5% to 99.4%) of premenopausal breast cancers could be attributable
41
42 170 to the combination of all the considered variables. This percentage was 61.2% (-15.5% to
43
44 91.9%) and 39.9% (-71.0% to 93.9%) for the combination of the non-behavioural factors and
45
46 172 the behavioural factors, respectively (Figure 1).

47
48
49 Regarding postmenopausal analyses, 72.9% (17.6% to 93.2%) of postmenopausal breast
50
51 174 cancers could be attributable to the combination of all the considered variables. This
52
53 percentage was 41.9% (5.0% to 68.7%) and 53.5% (12.8% to 78.7%) for the combination of
54
55 176 the non-behavioural factors and the behavioural factors, respectively (Figure 1).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Population-attributable fractions for premenopausal breast cancer according to

178 *hormonal receptors status*

During follow-up, 314 premenopausal women were diagnosed with ER+ and 97 with ER- breast cancer. Due to low numbers, HER2 status could not be considered. PAFs corresponding to ER+ and ER- status are presented in Table 3 (95% CI available in Supplementary Table 1).

As many as 3 out of 4 cancers were attributed to the combination of all risk factors for each stratum. This result was driven by BMI (25.0%) and personal history of breast cancer (16.2%) for ER+ cancers and by birth weight (61.0%) and age at menarche (38.6%) for ER- cancers.

186 *Population-attributable fractions for postmenopausal breast cancer according to*

hormonal receptors status

ER status was available for 2516 postmenopausal breast cancers, HER2 status for 925 cancers and both statuses for 912 cancers; the distribution of cases is provided in Table 3. Point estimates of the individual and combined PAFs are presented in Tables 3 and 4, and their 95% CIs are presented in Supplementary Tables 2 and 3.

As many as 80.2% (20.3% to 96.4%) of postmenopausal ER+ cancers and 53.6% (-91.6% to 99.2%) of ER- cancers could be attributable to the combination of all considered variables. Statistically significant PAFs observed for all postmenopausal women were confirmed for ER+ cancers, with additional statistically significant PAFs for height at adulthood (8.7%) and age at menopause (12.8%). Only one PAF remained statistically significant for ER- cancers, obtained for personal history of BBD (15.3%). The percentage of cancers attributable to all the considered variables for the other subtypes were estimated from 57.8% for ER-/HER2- cancers to 87.1% for ER+/HER2- cancers. The results for HER2- and ER+/HER2- cancers were similar to the overall analysis, except for age at menarche, parity and age at FFTP, and alcohol consumption for which the PAFs were no longer statistically significant. Regarding

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

202 HER2+ cancers, only one PAF remained statistically significant, obtained for personal history
of BBD (11.7%). In addition, we observed statistically significant PAFs associated with level
204 of education for HER2- (25.5%) and ER+/HER2- (27.1%) cancers and age at menopause for
HER2+ (39.1%) cancers.

206 *Sensitivity analyses*

All the sensitivity analyses provided similar results to those observed in the main analyses.

208

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

210 This article adds evidence to the literature that a high proportion of breast cancers can be avoided, both overall and in subgroups defined by tumour hormone receptor status.

212 To our knowledge, this is the first study to estimate combined PAFs associated with premenopausal and postmenopausal invasive breast cancer including an extended number of risk factors. Indeed, as many as 61% of premenopausal breast cancers after 40 years of age could be attributed to non-behavioural factors and 40% to behavioural factors, while in postmenopause, the impact of behavioural factors was higher (42% and 54%, respectively).

Literature confrontation

218 To our knowledge, this is the first study to estimate PAFs associated with anthropometric measures around menarche and birth weight for premenopausal and postmenopausal breast cancer separately. One study (41) estimated the PAFs for overall breast cancer risk associated with BMI around menarche at 15% and with birth weight at 7%. Values of PAFs for body shape around menarche in our study were close for both premenopausal and postmenopausal breast cancer, although not statistically significant in premenopause. In addition, a high body shape around menarche has previously been associated with a decreased risk of breast cancer in terms of HR (42-44). Our results regarding birth weight are consistent with two recent meta-analyses showing a positive association between birth weight and breast cancer risk that was stronger in pre- than in post-menopausal women (45;46).

228 Regarding premenopausal breast cancer, the statistically significant PAF observed in our analyses associated with a late menarche were consistent with recent published estimate of 230 22.2% (9).

Regarding postmenopausal breast cancer, the individual PAF observed for MHT was within 232 the range observed in the literature (2.4% to 19.4%) (8-10;13;47-50). The results suggested that a statistically significant number of postmenopausal breast cancer cases could be

1
2
3 234 attributed to the recent use of MHT, restricted to a combination of systemic oestrogen with a
4
5 progestagen (PAF = 13.8 (10.1 to 17.4), data not shown) (23). The postmenopausal
6
7 236 individual PAF observed for BMI was within the range noted in the literature (0.2% to
8
9 24.8%) (9-12;49-51), as were those for alcohol (-7.6% to 9.1%) (8-10;12;50), age at
10
11 238 menopause (5.9% to 14.5%) (8-10;13) and age at menarche (7.7% to 18.8%) (7-10). The
12
13 postmenopausal PAF associated with family history of breast cancer was lower than
14
15 240 previously published estimations (5.7% to 15.7%) (7-11;13) because of a lower prevalence of
16
17 exposure in our study population (up to 20.4% in the literature vs. 7.3% in our population,
18
19 242 data not shown) and lower relative risk estimates (1.49 to 6.29 in the literature vs. 1.37 and
20
21 1.50 among premenopausal and postmenopausal women in our analyses, see Tables 1 and 2).
22
23
24 244 Few studies quantified the joint influence of behavioural and/or non-behavioural factors on
25
26 premenopausal (12;13), postmenopausal (8;10-13) and overall (7;9;14;15) breast cancer.
27
28 246 Among them, very few jointly analysed the impact of only behavioural factors (8;10;12), only
29
30 non-behavioural factors (8;10;13), or their combination (11;13). Previously published
31
32 248 combined PAFs for premenopausal breast cancer were 42.8% for the combination of a low
33
34 consumption of alcohol and a high level of physical activity (12), and 52.0% for the
35
36 250 combination of high level of education, having a child before age 20 years and having no
37
38 family history of breast cancer (13).
39
40
41 252 For postmenopausal breast cancer, different combinations of behavioural risk factors were
42
43 associated with PAFs of 26.3% (including physical activity, use of MHT, BMI and alcohol
44
45 254 consumption) (10), 40.7% (including physical activity, use of MHT, alcohol consumption and
46
47 weight gain since age 18 years) (8) and from 12.4% to 40.7% when combining two factors
48
49 256 among alcohol consumption, physical activity and body mass index (12). In our study, the
50
51 observed PAF for behavioural risk factors (42.3%) was higher than these published estimates.
52
53
54 258 Regarding non-behavioural risk factors, previously published PAFs were 37.2% (10)
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

combining age at menopause, parity, BBD, age at menarche and family history of breast and
260 57.3% when adding age at FFTP and height at adulthood to the estimation (8). In our study,
we observed a PAF associated with non-behavioural factors of 51.7%. Differences found for
262 both the combination of behavioural and non-behavioural factors may result from a higher
number of factors included in our study.

264 Our results are in line with the only previously published study that evaluated individual and
combined postmenopausal PAFs stratifying on hormonal receptor status (10), in which the
266 use of MHT, age at menopause and age at menarche were associated with higher PAF
estimates for ER+ than ER- cancers (10). In addition, previous pooled- and meta-analyses
268 have found that a high BMI was associated with a decreased risk of ER+/PR+ premenopausal
breast cancer in terms of HR (17;20) and that the decreased risk of postmenopausal breast
270 cancer for a 5-unit increase in BMI was restricted to ER+/PR+ postmenopausal breast cancer
(20). Though not statistically significant, one study found that a late menarche was associated
272 with a decreased risk of ER- premenopausal breast cancer and increased risk of ER+
premenopausal breast cancer (52). A late menarche has previously been associated with a
274 decreased risk of ER+ (18;53-55) breast cancer. High BMI around menarche has been
associated with a decreased risk of ER- (56), ER+/PR- (57), and ER+/PR+ (42)
276 postmenopausal breast cancer. Use of MHT has been associated with an increased risk of
ER+/PR+ (53;58), ER+ (54;59) and HER2- (58;60) postmenopausal breast cancer.

278 *Strengths and limitations*

The cut-points used to define exposure levels can have an effect on PAFs estimates (40).
280 However to avoid bias, the categorization used had been chosen *a priori* and based on the
literature and public health recommendations. Age and body shape at menarche were
282 retrospectively collected from women aged 40-65 years at baseline and thus require long-term
memory, which may generate measurement errors. However, an independent validation study

1
2
3 284 showed that this information was reliable (correlations between original and recalled was 0.8
4
5 for age and 0.6 for body shape) (61). In addition, like most cohorts, the E3N population is not
6
7 286 representative of the general population and is prone to a healthy cohort effect. As PAF
8
9 estimates depend on the prevalence of exposures among the population of interest, PAF of
10
11 healthy behaviours (tobacco smoking, physical activity level, alcohol consumption or BMI)
12 288 would be higher in a less healthy population. However, assessing PAFs in the general
13
14 population is not feasible because of the lack of data in their prevalence. If PAFs estimates are
15 290 useful to rank risk factors, it is obvious that public health interventions are not possible for all
16
17 of (age at menarche for instance). Finally, results for premenopausal breast cancer should be
18 292 interpreted with caution because of the limited number of premenopausal breast cancer cases
19
20 resulting from the selection of the cohort population restricted to women over 40 years, where
21 294 genetic factors are less predominant than in younger women (62).
22
23

24
25 296 The main strength of these analyses is the size of the cohort and an extended follow-up
26
27 enabling a high statistical power and a wide set of available risk factors. The study design also
28
29 enabled analysis of the heterogeneity of subgroups of breast cancer defined by its menopausal
30 298 and receptor status. Histological confirmation was obtained for the vast majority of breast
31
32 cancer cases, and exclusion of non-validated cases produced similar results. Because of the
33
34 prospective design, information on exposure collected prior to cancer diagnosis resulted in
35 300 low risk of recall bias. In addition, dietary and anthropometric data were validated (27;32),
36
37 and demonstrated limited declaration bias and good reliability of the reported data. Another
38
39 strength is the formula used to estimate the PAF, allowing a non-biased estimation of adjusted
40 304 PAF (5).
41
42

43 306 *Conclusion*

44
45 Behavioural modifications could have a large impact on breast cancer risk after age 40, even
46 308 more so after menopause. Health agencies should favour more effective information on the
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

most important behavioural factors for breast cancer: alcohol consumption, diet,
310 postmenopausal obesity and menopausal hormone therapy and encourage research on how to
modify people's behaviours.

Confidential: For Review Only

Table 1 Non-behavioural baseline characteristics and multivariate-adjusted hazard ratios with 95% confidence intervals regarding invasive breast cancer among premenopausal and postmenopausal women, E3N cohort (N=67 634)

	Non cases	Invasive premenopausal breast cancer cases		Invasive postmenopausal breast cancer cases	
	N=63 999	N=497	HR [95%CI]	N=3138	HR [95%CI]
Age (years)^a					
Mean (SD)	52.75 (6.66)	46.77 (2.51)	-	54.01 (6.16)	-
First-degree family history of breast cancer^b					
At least one	4587 (7.17)	46 (9.25)	1.37 [1.01; 1.86]	345 (10.99)	1.50 [1.34; 1.69]
None	49 746 (77.73)	386 (77.67)	Reference	2321 (73.97)	Reference
Missing	9666 (15.10)	65 (13.08)	1.11 [0.85; 1.45]	472 (15.04)	0.99 [0.90; 1.10]
Level of education					
Undergraduate	7192 (11.24)	33 (6.64)	Reference	340 (10.83)	Reference
Graduate	34 019 (53.15)	251 (50.50)	1.15 [0.80; 1.66]	1642 (52.33)	1.00 [0.89; 1.13]
Post-graduate	22 788 (35.61)	213 (42.86)	1.05 [0.72; 1.53]	1156 (36.84)	1.07 [0.94; 1.21]
Height at adulthood (cm)					
<160	21 922 (34.25)	138 (27.77)	Reference	1036 (33.02)	Reference
[160; 165)	22 394 (34.99)	177 (35.61)	1.06 [0.85; 1.33]	1089 (34.70)	1.04 [0.95; 1.13]
[165; 170)	13 955 (21.81)	119 (23.94)	1.04 [0.81; 1.33]	702 (22.37)	1.09 [0.99; 1.21]
≥ 170	5728 (8.95)	63 (12.68)	1.18 [0.87; 1.61]	311 (9.91)	1.20 [1.05; 1.37]
History of benign breast diseases^c					
At least one, with at least one biopsy performed	1556 (2.43)	20 (4.02)	2.02 [1.37; 2.99]	131 (4.17)	1.75 [1.52; 2.02]
At least one, no biopsy performed	16 860 (26.34)	177 (35.61)	1.42 [1.18; 1.71]	1054 (33.59)	1.48 [1.37; 1.60]
None	45 583 (71.23)	300 (60.37)	Reference	1953 (62.24)	Reference
Age at menarche (years)					
<10	205 (0.32)	2 (0.40)	1.43 [0.35; 5.81]	13 (0.42)	1.58 [0.91; 2.74]
[10; 12)	12 758 (19.93)	98 (19.72)	1.26 [0.95; 1.66]	649 (20.68)	1.19 [1.07; 1.32]
[12; 14)	33 177 (51.84)	291 (58.55)	1.36 [1.09; 1.70]	1651 (52.61)	1.13 [1.04; 1.23]
≥ 14	17 859 (27.91)	106 (21.33)	Reference	825 (26.29)	Reference
Birth weight					
Low	3466 (5.41)	14 (2.82)	Reference	154 (4.91)	Reference
Normal	46 189 (72.17)	360 (72.43)	1.72 [1.01; 2.95]	2321 (73.96)	1.13 [0.96; 1.33]
High	3531 (5.52)	31 (6.24)	1.99 [1.05; 3.76]	163 (5.19)	1.03 [0.82; 1.29]
Missing	10 813 (16.90)	92 (18.51)	2.32 [1.32; 4.08]	500 (15.94)	1.34 [1.11; 1.60]
Age at menopause (years, for postmenopausal women)^d					
< 48	10 426 (16.29)	-	-	489 (15.58)	Reference
[48; 50)	10 275 (16.06)	-	-	548 (17.46)	1.09 [0.97; 1.23]
[50; 52)	18 587 (29.04)	-	-	993 (31.65)	1.11 [1.00; 1.24]
[52; 54)	13 251 (20.71)	-	-	623 (19.85)	1.03 [0.92; 1.16]
≥ 54	9673 (15.11)	-	-	485 (15.46)	1.17 [1.03; 1.34]

NOTE: Bold values means p-value<0.05.

a: Age was considered as the timescale in Cox regression models, which did not allow the estimation of HR associated with age.

b: Family history of breast cancer included history among first-degree relatives (mother, sisters and daughters).

c: Benign breast diseases included adenomas and fibrocystic diseases.

d: Non cases included 1787 (2.79%) of women who were still premenopausal at the end of follow-up.

Table 2 Behavioural baseline characteristics and multivariate-adjusted hazard ratios with 95% confidence intervals regarding invasive breast cancer among premenopausal and postmenopausal women, E3N cohort (N=67 634)

	Non cases	Invasive premenopausal breast cancer cases		Invasive postmenopausal breast cancer cases	
	N=63 999	N=497	HR [95% CI]	N=3138	HR [95% CI]
Tobacco smoking					
Current smoker	8747 (13.67)	74 (14.89)	0.96 [0.71; 1.28]	395 (12.59)	0.99 [0.88; 1.13]
Past smoker	34 257 (53.53)	229 (46.08)	1.15 [0.95; 1.40]	1684 (53.66)	1.01 [0.94; 1.09]
Never smoker	20 995 (32.80)	194 (39.03)	Reference	1059 (33.75)	Reference
Number of children and age at FFTP					
One child before 30	7115 (11.12)	51 (10.26)	0.99 [0.73; 1.34]	325 (10.36)	0.99 [0.88; 1.12]
One child after 30	2828 (4.42)	38 (7.65)	1.64 [1.16; 2.31]	164 (5.22)	1.29 [1.09; 1.51]
More than one child, the first before 30	42 829 (66.92)	314 (63.18)	Reference	1998 (63.67)	Reference
More than one child, the first after 30	3847 (6.01)	44 (8.85)	1.44 [1.04; 1.98]	213 (6.79)	1.22 [1.06; 1.40]
Nulliparous	7380 (11.53)	50 (10.06)	0.97 [0.69; 1.35]	438 (13.96)	1.28 [1.13; 1.45]
Physical activity level (METs/hour/week)					
< 20	12 951 (20.24)	120 (24.15)	1.06 [0.82; 1.38]	618 (19.69)	1.02 [0.92; 1.13]
[20; 40)	21 072 (32.92)	176 (35.41)	1.00 [0.79; 1.28]	1033 (32.92)	1.03 [0.94; 1.13]
[40; 60)	13 652 (21.33)	94 (18.91)	0.91 [0.69; 1.20]	688 (21.93)	1.03 [0.93; 1.14]
≥ 60	16 324 (25.51)	107 (21.53)	Reference	799 (25.46)	Reference
Body shape at menarche^a					
Body shape 1	13 465 (21.04)	86 (17.30)	1.08 [0.80; 1.45]	751 (23.93)	1.33 [1.19; 1.49]
Body shape 2	22 093 (34.52)	185 (37.22)	1.28 [1.00; 1.65]	1140 (36.33)	1.27 [1.15; 1.41]
Body shape 3	14 922 (23.32)	130 (26.16)	1.22 [0.93; 1.59]	704 (22.44)	1.18 [1.06; 1.32]
Body shape ≥ 4	13 519 (21.12)	96 (19.32)	Reference	543 (17.30)	Reference
Breastfeeding (months)					
None	24 032 (37.55)	187 (37.63)	1.09 [0.82; 1.45]	1231 (39.23)	1.02 [0.91; 1.14]
(0; 3)	17 901 (27.97)	134 (26.96)	0.94 [0.71; 1.25]	858 (27.34)	1.07 [0.95; 1.19]
[3; 6)	10 279 (16.06)	95 (19.11)	1.03 [0.76; 1.38]	513 (16.35)	1.11 [0.99; 1.26]
≥ 6	11 787 (18.42)	81 (16.30)	Reference	536 (17.08)	Reference
Dietary pattern^b					
Low Healthy and Western scores	15 831 (24.73)	100 (20.12)	1.05 [0.78; 1.40]	761 (24.25)	1.10 [0.99; 1.22]
Low Healthy score and high Western score	16 158 (25.25)	154 (30.99)	1.05 [0.81; 1.37]	813 (25.91)	1.20 [1.08; 1.33]
High Healthy score and low Western score	16 292 (25.46)	90 (18.11)	Reference	743 (23.68)	Reference
High Healthy and Western scores	15 718 (24.56)	153 (30.78)	1.36 [1.04; 1.79]	821 (26.16)	1.15 [1.04; 1.28]
Alcohol consumption (g/day)					
[0; 10)	38 441 (60.06)	302 (60.76)	Reference	1750 (55.77)	Reference
[10; 20)	13 060 (20.41)	106 (21.33)	1.01 [0.81; 1.27]	659 (21.00)	1.05 [0.96; 1.15]
≥ 20	12 498 (19.53)	89 (17.91)	0.89 [0.69; 1.14]	729 (23.23)	1.18 [1.07; 1.29]
Vitamin D intake (µg) and UVRd (kJ/m²/day)^c					
Vitamin D intake ≥ 2.38 and UVRd ≥ 2.5	15 329 (23.95)	123 (24.75)	Reference	720 (22.94)	Reference
Vitamin D intake ≥ 2.38 and UVRd < 2.5	16 631 (25.99)	133 (26.76)	0.93 [0.73; 1.20]	881 (28.08)	1.14 [1.03; 1.26]
Vitamin D intake < 2.38 and UVRd ≥ 2.5	17 457 (27.28)	119 (23.94)	0.97 [0.74; 1.27]	834 (26.58)	1.05 [0.94; 1.16]
Vitamin D intake < 2.38 and UVRd < 2.5	14 582 (22.78)	122 (24.55)	1.11 [0.85; 1.46]	703 (22.40)	1.06 [0.95; 1.18]
Oral contraceptive or progestagen alone use					
Recent ^d	10 763 (16.82)	136 (27.36)	1.07 [0.87; 1.33]	549 (17.50)	1.38 [1.18; 1.61]

Past, last use < 10 years ago	16 147 (25.23)	138 (27.77)	1.09 [0.87; 1.37]	870 (27.72)	1.06 [0.97; 1.15]
Past, last use ≥ 10 years ago	14 489 (22.64)	145 (29.17)	Reference	598 (19.06)	Reference
Never	17 132 (26.77)	57 (11.47)		858 (27.34)	
Unknown recency	5468 (8.54)	21 (4.23)	0.78 [0.46; 1.29]	263 (8.38)	0.95 [0.82; 1.09]
Body mass index (kg/m²)					
< 18.5	2234 (3.49)	18 (3.62)	1.03 [0.59; 1.80]	84 (2.68)	Reference
[18.5; 25)	48 383 (75.60)	404 (81.29)	1.15 [0.91; 1.45]	2310 (73.61)	
[25; 30)	10 994 (17.18)	62 (12.47)	Reference	610 (19.44)	1.19 [1.10; 1.30]
≥ 30	2388 (3.73)	13 (2.62)		134 (4.27)	1.25 [1.07; 1.46]
Recent^d MHT use (for postmenopausal women)^e					
Oral oestrogens only	1746 (2.73)	-		108 (3.44)	1.07 [0.92; 1.23]
Oral oestrogens with progesterone/dydrogesterone	4620 (7.22)	-		296 (9.43)	1.20 [1.09; 1.32]
Oral oestrogens with other progestagen ^f	5251 (8.20)	-		412 (13.13)	1.72 [1.57; 1.88]
Other ^g	900 (1.41)	-		57 (1.82)	1.17 [0.96; 1.44]
No recent MHT	49 695 (77.65)	-		2265 (72.18)	Reference

NOTE: Bold values means p-value<0.05.

a: Body shapes were based on drawings in a series of Sorensen's body shapes.

b: Dietary patterns were computed from a factor analysis, based on 57 predefined food groups, using a principal component analysis. The first two factors were extracted and respectively labelled "Healthy pattern" and "Western pattern". A four-modality variable was created using the median value of both of the dietary scores as thresholds. Weak adherence means a score under the median value, great adherence a score over the median value.

c: The average daily vitamin D was estimated from a food composition table. Baseline regions of residence were linked to a database containing mean daily ultraviolet radiation doses (UVR_d in kJ/m²/day). Daily dietary vitamin D and ultraviolet radiation doses were compiled into one four-modality variable, using the median value of daily vitamin D and ultraviolet radiation doses as threshold.

d: Recent use of oral contraceptive, progestagens alone or MHT was defined as current use or past use within the 12 preceding months.

e: Non cases included 1787 (2.79%) of women who were still premenopausal at the end of follow-up.

f: Other progestagen included chlormadinone acetate, cyproterone acetate, demegestone, dienogest, drospirenone, ethynodiol acetate, gestodene, levonorgestrel, lynestrenol, medrogestone, medroxyprogesterone acetate, megestrol acetate, norethisterone acetate, and promegestone.

g: Other MHT included tibolone and MHT containing an androgen, intramuscularly administered, or with no specified formulation.

Table 3 Individual and combined PAFs associated with breast cancer risk factors for ER+ and ER- premenopausal and postmenopausal breast cancer, E3N cohort (N=67 634)

	Premenopausal breast cancer		Postmenopausal breast cancer	
	ER+ N=314	ER- N=97	ER+ N=2050	ER- N=466
Non-behavioural factors	PAF		PAF	
First-degree family history of breast cancer (ref=none)	-0.26	5.82	4.25*	2.20
Level of education (ref=undergraduate)	5.33	-19.00	6.45	4.73
Height at adulthood (ref=less than 160 cm)	7.91	-26.20	8.70*	-0.35
History of benign breast disease (ref=none)	16.21*	11.79	14.51*	15.31*
Age at menarche (ref=more than 14)	14.14	38.62*	14.47*	3.51
Birth weight (ref=low)	21.37	61.01*	5.81	24.63
Age at menopause, for postmenopausal women only (ref=less than 48)	-	-	12.83*	-7.53
Combination of all non-behavioural factors	52.18	70.32	50.82*	38.07
Behavioural factors				
Tobacco smoking (ref=never)	8.87	-1.14	-0.20	1.80
Number of children and age at FFTP (ref=more than one child, the first before 30)	4.84	11.16	7.05*	-6.95
Physical activity level (ref=highly active)	-4.31	12.69	3.08	-0.34
Body shape at menarche (ref=plump)	9.66	13.03	15.25*	6.82
Breastfeeding (ref=more than 6 months)	2.26	-1.43	7.82	-0.65
Dietary pattern (ref=high Healthy score and low Western score)	12.64	14.72	12.85*	7.82
Alcohol consumption (ref=less than one drink a day)	-0.92	-6.45	3.60	8.48
Vitamin D intake and UVRd (ref=high levels for both)	3.57	-10.90	8.50	4.64
Oral contraceptive or progestagen alone use (ref=last use more than 10 years ago or never used)	5.91	-1.08	3.81	2.17
Body mass index (ref=more than 25 (premenopausal) or less than 25 (postmenopausal))	25.01*	-12.40	6.47*	-1.05
Recent MHT use, for postmenopausal women only (ref=no recent MHT)	-	-	17.51*	5.00
Combination of all behavioural factors	50.56	21.80	59.81*	25.28
Combination of all factors	76.30	76.83	80.21*	53.61

* p-value < 0.05

Table 4 Individual and combined PAFs associated with breast cancer risk factors for several types of postmenopausal breast cancer, E3N cohort (N=67 634)

	Postmenopausal breast cancer					
	HER2+	HER2-	ER+/HER2+	ER+/HER2-	ER-/HER2+	ER-/HER2-
	N=178	N=747	N=113	N=629	N=63	N=107
Non-behavioural factors:	PAF		PAF			
First-degree family history of breast cancer (ref=none)	0.09	4.26*	4.23	4.14*	-7.19	6.24
Level of education (ref=undergraduate)	-26.00	25.54*	-20.90	27.10*	-24.70	9.36
Height at adulthood (ref=less than 160 cm)	19.35	12.71	27.51	13.11	7.94	14.72
History of benign breast disease (ref=none)	16.55*	11.70*	15.62	10.91*	16.06	15.75
Age at menarche (ref=more than 14)	5.47	9.72	2.21	11.91	14.58	-7.19
Birth weight (ref=low)	3.69	14.40	-12.20	11.73	28.38	23.12
Age at menopause, for postmenopausal women only (ref=less than 48)	39.10*	-1.52	40.69	4.60	34.23	-31.00
Combination of all non-behavioural factors	53.04	56.74	53.87	59.73	58.51	33.93
Behavioural factors:						
Tobacco smoking (ref=never)	8.06	-0.68	11.16	-1.89	5.08	11.28
Number of children and age at FFTP (ref=more than one child, the first before 30)	-1.08	4.00	7.31	7.01	-18.30	-16.20
Physical activity level (ref=highly active)	-4.03	4.08	1.16	4.69	-17.10	-5.21
Body shape at menarche (ref=plump)	8.72	15.99*	12.38	19.85*	0.11	-0.62
Breastfeeding (ref=more than 6 months)	4.48	15.15	3.77	14.02	5.33	17.36
Dietary pattern (ref=high Healthy score and low Western score)	3.59	16.06*	0.53	16.56*	12.09	13.81
Alcohol consumption (ref=less than one drink a day)	4.78	9.54*	8.22	8.94	0.62	14.23
Vitamin D intake and UVRd (ref=high levels for both)	-10.10	8.35	-15.00	8.28	2.59	14.15
Oral contraceptive or progestagen alone use (ref=last use more than 10 years ago or never used)	-4.69	2.11	-13.40	3.20	13.12	-5.60
Body mass index (ref=more than 25 (premenopausal) or less than 25 (postmenopausal))	12.21	8.12*	11.33	10.45*	13.33	-6.05
Recent MHT use, for postmenopausal women only (ref=no recent MHT)	0.63	12.99*	8.04	15.00*	-13.50	-0.55
Combination of all behavioural factors	22.39	63.91	33.75	68.00	11.42	36.07
Combination of all factors	63.57	84.33	69.46	87.08	63.00	57.78

* p-value < 0.05

Table 4 Individual and combined PAFs associated with breast cancer risk factors for several types of postmenopausal breast cancer, E3N cohort (N=67 634)

	Postmenopausal breast cancer					
	HER2+	HER2-	ER+/HER2+	ER+/HER2-	ER-/HER2+	ER-/HER2-
	N=178	N=747	N=113	N=629	N=63	N=107
Non-behavioural factors:	PAF		PAF			
First-degree family history of breast cancer (ref=none)	0.09	4.26*	4.23	4.14*	-7.19	6.24
Level of education (ref=undergraduate)	-26.00	26.54*	-20.90	27.10*	-24.70	9.36
Height at adulthood (ref=less than 160 cm)	19.35	12.71	27.51	13.11	7.94	14.72
History of benign breast disease (ref=none)	16.55*	11.70*	15.62	10.91*	16.06	15.75
Age at menarche (ref=more than 14)	5.47	9.72	2.21	11.91	14.58	-7.19
Birth weight (ref=low)	3.69	14.40	-12.20	11.73	28.38	23.12
Age at menopause, for postmenopausal women only (ref=less than 48)	39.10*	-1.52	40.69	4.60	34.23	-31.00
Combination of all non-behavioural factors	53.04	56.74	53.87	59.73	58.51	33.93
Behavioural factors:						
Tobacco smoking (ref=never)	8.06	-0.68	11.16	-1.89	5.08	11.28
Number of children and age at FFTP (ref=more than one child, the first before 30)	-1.08	4.00	7.31	7.01	-18.30	-16.20
Physical activity level (ref=highly active)	-4.03	4.08	1.16	4.69	-17.10	-5.21
Body shape at menarche (ref=plump)	8.72	15.99*	12.38	19.85*	0.11	-0.62
Breastfeeding (ref=more than 6 months)	4.48	15.15	3.77	14.02	5.33	17.36
Dietary pattern (ref=high Healthy score and low Western score)	3.59	16.06*	0.53	16.56*	12.09	13.81
Alcohol consumption (ref=less than one drink a day)	4.78	9.54*	8.22	8.94	0.62	14.23
Vitamin D intake and UVRd (ref=high levels for both)	-10.10	8.35	-15.00	8.28	2.59	14.15
Oral contraceptive or progestagen alone use (ref=last use more than 10 years ago or never used)	-4.69	2.11	-13.40	3.20	13.12	-5.60
Body mass index (ref=more than 25 (premenopausal) or less than 25 (postmenopausal))	12.21	8.12*	11.33	10.46*	13.33	-6.05
Recent MHT use, for postmenopausal women only (ref=no recent MHT)	0.63	12.99*	8.04	15.00*	-13.50	-0.55
Combination of all behavioural factors	22.39	63.91	33.75	68.00	11.42	36.07
Combination of all factors	63.57	84.33	69.46	87.08	63.00	57.78

* p-value < 0.05

Reference List

- 1
2
3
4 (1) Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C et al. GLOBOCAN 2012 v1.0,
5 Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France:
6 International Agency for Research on Cancer 2013; Available from: <http://globocan.iarc.fr>, accessed
7 on 22/04/2014.
- 8
9 (2) Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol* 2001; 2(3):133-140.
- 10
11 (3) Iwasaki M, Tsugane S. Risk factors for breast cancer: epidemiological evidence from Japanese
12 studies. *Cancer Sci* 2011; 102(9):1607-1614.
- 13
14 (4) National breast and ovarian cancer centre. Breast cancer risk factors: a review of the evidence.
15 National breast and ovarian cancer centre, Surry Hills, NSW 2009.
- 16
17 (5) Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J*
18 *Public Health* 1998; 88(1):15-19.
- 19
20 (6) Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the
21 United States explained by well-established risk factors. *J Natl Cancer Inst* 1995; 87(22):1681-1685.
- 22
23 (7) Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for
24 multiple risk factors using case-control data. *Am J Epidemiol* 1985; 122(5):904-914.
- 25
26 (8) Sprague BL, Trentham-Dietz A, Egan KM, Titus-Ernstoff L, Hampton JM, Newcomb PA.
27 Proportion of invasive breast cancer attributable to risk factors modifiable after menopause. *Am J*
28 *Epidemiol* 2008; 168(4):404-411.
- 29
30 (9) Hines LM, Risendal B, Slattery ML, Baumgartner KB, Giuliano AR, Sweeney C et al. Comparative
31 analysis of breast cancer risk factors among Hispanic and non-Hispanic white women. *Cancer* 2010;
32 116(13):3215-3223.
- 33
34 (10) Barnes BB, Steindorf K, Hein R, Flesch-Janys D, Chang-Claude J. Population attributable risk of
35 invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable
36 risk factors. *Cancer Epidemiol* 2011; 35(4):345-352.
- 37
38 (11) Ghiasvand R, Bahmanyar S, Zendehdel K, Tahmasebi S, Talei A, Adami HO et al. Postmenopausal
39 breast cancer in Iran; risk factors and their population attributable fractions. *BMC Cancer* 2012;
40 12:414.
- 41
42 (12) Mezzetti M, La Vecchia C, Decarli A, Boyle P, Talamini R, Franceschi S. Population attributable
43 risk for breast cancer: diet, nutrition, and physical exercise. *J Natl Cancer Inst* 1998; 90(5):389-394.
- 44
45 (13) Tavani A, Braga C, La Vecchia C, Negri E, Russo A, Franceschi S. Attributable risks for breast
46 cancer in Italy: education, family history and reproductive and hormonal factors. *Int J Cancer* 1997;
47 70(2):159-163.
- 48
49 (14) Granstrom C, Sundquist J, Hemminki K. Population attributable risks for breast cancer in Swedish
50 women by morphological type. *Breast Cancer Res Treat* 2008; 111(3):559-568.
- 51
52 (15) Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and
53 environmental factors in the UK in 2010. *Br J Cancer* 2011; 105 Suppl 2:S77-S81.
- 54
55 (16) Chen WY, Colditz GA. Risk factors and hormone-receptor status: epidemiology, risk-prediction
56 models and treatment implications for breast cancer. *Nat Clin Pract Oncol* 2007; 4(7):415-423.
- 57
58
59
60

- 1 (17) Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL et al. Associations of
2 breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association
3 Consortium studies. *J Natl Cancer Inst* 2011; 103(3):250-263.
- 4 (18) Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a
5 review of the literature. *Breast Cancer Res Treat* 2014; 144(1):1-10.
- 6 (19) Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective
7 studies. *Breast Cancer Res Treat* 2013; 137(3):869-882.
- 8 (20) Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by
9 estrogen and progesterone receptor status--a meta-analysis. *Int J Cancer* 2009; 124(3):698-712.
- 10 (21) Zhang SM, Lee IM, Manson JE, Cook NR, Willett WC, Buring JE. Alcohol consumption and breast
11 cancer risk in the Women's Health Study. *Am J Epidemiol* 2007; 165(6):667-676.
- 12 (22) Clavel-Chapelon F, van Liere MJ, Giubout C, Niravong MY, Goulard H, Le Corre C et al. E3N, a
13 French cohort study on cancer risk factors. E3N Group. Etude Epidemiologique aupres de femmes de
14 l'Education Nationale. *Eur J Cancer Prev* 1997; 6(5):473-478.
- 15 (23) Fournier A, Berrino F, Clavel-Chapelon F. Unequal risks for breast cancer associated with different
16 hormone replacement therapies: results from the E3N cohort study. *Breast Cancer Res Treat* 2008;
17 107(1):103-111.
- 18 (24) Ekblom A, Erlandsson G, Hsieh C, Trichopoulos D, Adami HO, Cnattingius S. Risk of breast cancer
19 in prematurely born women. *J Natl Cancer Inst* 2000; 92(10):840-841.
- 20 (25) Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C et al. 2011
21 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports
22 Exerc* 2011; 43(8):1575-1581.
- 23 (26) Sorensen TI, Stunkard AJ, Teasdale TW, Higgins MW. The accuracy of reports of weight: children's
24 recall of their parents' weights 15 years earlier. *Int J Obes* 1983; 7(2):115-122.
- 25 (27) van Liere MJ, Lucas F, Clavel F, Slimani N, Villemainot S. Relative validity and reproducibility of a
26 French dietary history questionnaire. *Int J Epidemiol* 1997; 26 Suppl 1:S128-S136.
- 27 (28) Cottet V, Touvier M, Fournier A, Touillaud MS, Lafay L, Clavel-Chapelon F et al. Postmenopausal
28 breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. *Am J Epidemiol*
29 2009; 170(10):1257-1267.
- 30 (29) Favier J.-C., Ireland-Ripert J., Toque C., Feinberg M., INRA (Paris France), CNEVA (Paris France)
31 et al. Répertoire général des aliments : table de composition (Composition tables). Tec & Doc -
32 Lavoisier Paris, 897p 1995.
- 33 (30) Gschwind B, Ménard L, Albuissou M, Wald L. Converting a successful research project into a
34 sustainable service: the case of the SoDa Web service. *Environmental Modelling and Software* 2006;
35 21:1555-1561.
- 36 (31) Engel P, Fagherazzi G, Mesrine S, Boutron-Ruault MC, Clavel-Chapelon F. Joint effects of dietary
37 vitamin D and sun exposure on breast cancer risk: results from the French E3N cohort. *Cancer
38 Epidemiol Biomarkers Prev* 2011; 20(1):187-198.
- 39 (32) Tehard B, van Liere MJ, Com NC, Clavel-Chapelon F. Anthropometric measurements and body
40 silhouette of women: validity and perception. *J Am Diet Assoc* 2002; 102(12):1779-1784.
- 41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (33) Fournier A, Berrino F, Riboli E, Avenel V, Clavel-Chapelon F. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. *Int J Cancer* 2005; 114(3):448-454.
 - (34) Rosenberg LU, Magnusson C, Lindstrom E, Wedren S, Hall P, Dickman PW. Menopausal hormone therapy and other breast cancer risk factors in relation to the risk of different histological subtypes of breast cancer: a case-control study. *Breast Cancer Res* 2006; 8(1):R11.
 - (35) Thiebaut AC, Benichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Stat Med* 2004; 23(24):3803-3820.
 - (36) Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; 371(9612):569-578.
 - (37) Fagherazzi G, Chabbert-Buffet N, Fabre A, Guillas G, Boutron-Ruault MC, Mesrine S et al. Hip circumference is associated with the risk of premenopausal ER-/PR- breast cancer. *Int J Obes (Lond)* 2012; 36(3):431-439.
 - (38) Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008; 56(3):159-175.
 - (39) Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control* 2007; 18(5):571-579.
 - (40) Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. *Am J Epidemiol* 1998; 147(9):826-833.
 - (41) Ahlgren M, Melbye M, Wohlfahrt J, Sorensen TI. Growth patterns and the risk of breast cancer in women. *N Engl J Med* 2004; 351(16):1619-1626.
 - (42) Fagherazzi G, Guillas G, Boutron-Ruault MC, Clavel-Chapelon F, Mesrine S. Body shape throughout life and the risk for breast cancer at adulthood in the French E3N cohort. *Eur J Cancer Prev* 2013; 22(1):29-37.
 - (43) Baer HJ, Colditz GA, Rosner B, Michels KB, Rich-Edwards JW, Hunter DJ et al. Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. *Breast Cancer Res* 2005; 7(3):R314-R325.
 - (44) Weiderpass E, Braaten T, Magnusson C, Kumle M, Vainio H, Lund E et al. A prospective study of body size in different periods of life and risk of premenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004; 13(7):1121-1127.
 - (45) Michels KB, Xue F. Role of birthweight in the etiology of breast cancer. *Int J Cancer* 2006; 119(9):2007-2025.
 - (46) Xu X, Dailey AB, Peoples-Sheps M, Talbott EO, Li N, Roth J. Birth weight as a risk factor for breast cancer: a meta-analysis of 18 epidemiological studies. *J Womens Health (Larchmt)* 2009; 18(8):1169-1178.
 - (47) Coombs NJ, Taylor R, Wilcken N, Boyages J. HRT and breast cancer: impact on population risk and incidence. *Eur J Cancer* 2005; 41(12):1775-1781.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- (48) Clarke CA, Purdie DM, Glaser SL. Population attributable risk of breast cancer in white women associated with immediately modifiable risk factors. *BMC Cancer* 2006; 6:170.
- (49) Hayes J, Richardson A, Frampton C. Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women. *Intern Med J* 2013; 43(11):1198-1204.
- (50) Neutel CI, Morrison H. Could recent decreases in breast cancer incidence really be due to lower HRT use? Trends in attributable risk for modifiable breast cancer risk factors in Canadian women. *Can J Public Health* 2010; 101(5):405-409.
- (51) La Vecchia C, Negri E, Franceschi S, Talamini R, Bruzzi P, Palli D et al. Body mass index and postmenopausal breast cancer: an age-specific analysis. *Br J Cancer* 1997; 75(3):441-444.
- (52) Nichols HB, Trentham-Dietz A, Love RR, Hampton JM, Hoang Anh PT, Allred DC et al. Differences in breast cancer risk factors by tumor marker subtypes among premenopausal Vietnamese and Chinese women. *Cancer Epidemiol Biomarkers Prev* 2005; 14(1):41-47.
- (53) Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst* 2004; 96(3):218-228.
- (54) Setiawan VW, Monroe KR, Wilkens LR, Kolonel LN, Pike MC, Henderson BE. Breast cancer risk factors defined by estrogen and progesterone receptor status: the multiethnic cohort study. *Am J Epidemiol* 2009; 169(10):1251-1259.
- (55) Islam T, Matsuo K, Ito H, Hosono S, Watanabe M, Iwata H et al. Reproductive and hormonal risk factors for luminal, HER2-overexpressing, and triple-negative breast cancer in Japanese women. *Ann Oncol* 2012; 23(9):2435-2441.
- (56) Li J, Humphreys K, Eriksson L, Czene K, Liu J, Hall P. Effects of childhood body size on breast cancer tumour characteristics. *Breast Cancer Res* 2010; 12(2):R23.
- (57) Canchola AJ, Anton-Culver H, Bernstein L, Clarke CA, Henderson K, Ma H et al. Body size and the risk of postmenopausal breast cancer subtypes in the California Teachers Study cohort. *Cancer Causes Control* 2012.
- (58) Rosner B, Glynn RJ, Tamimi RM, Chen WY, Colditz GA, Willett WC et al. Breast cancer risk prediction with heterogeneous risk profiles according to breast cancer tumor markers. *Am J Epidemiol* 2013; 178(2):296-308.
- (59) Rosenberg LU, Einarsdottir K, Friman EI, Wedren S, Dickman PW, Hall P et al. Risk factors for hormone receptor-defined breast cancer in postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 2006; 15(12):2482-2488.
- (60) Slinger TE, Chang-Claude JC, Obi N, Kropp S, Berger J, Vettorazzi E et al. Menopausal hormone therapy and risk of clinical breast cancer subtypes. *Cancer Epidemiol Biomarkers Prev* 2009; 18(4):1188-1196.
- (61) Must A, Phillips SM, Naumova EN, Blum M, Harris S, Dawson-Hughes B et al. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am J Epidemiol* 2002; 155(7):672-679.
- (62) Rebora P, Czene K, Reilly M. Timing of familial breast cancer in sisters. *J Natl Cancer Inst* 2008; 100(10):721-727.

Annexe H : « A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort » - Article en revue au journal « Breast Cancer Research and Treatment »

Breast Cancer Research and Treatment
A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort
 --Manuscript Draft--

Manuscript Number:	
Full Title:	A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort
Article Type:	Epidemiology
Keywords:	breast cancer; Women; Risk score; Proportional hazard Cox regression; Nearest-neighbor algorithm; Discrimination; Calibration; Postmenopausal women; Premenopausal women; menopausal status; Gail model
Corresponding Author:	Françoise Clavel-Chapelon, PhD Inserm, Centre for Research in Epidemiology and Population Health (CESP), U1018, Team 9, F-94805, Villejuif, France 2 Univ Paris Sud, UMRS 1018, F-94805, Villejuif, France 3 Gustave Roussy, F-94805, Villejuif, France Villejuif, FRANCE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Inserm, Centre for Research in Epidemiology and Population Health (CESP), U1018, Team 9, F-94805, Villejuif, France 2 Univ Paris Sud, UMRS 1018, F-94805, Villejuif, France 3 Gustave Roussy, F-94805, Villejuif, France
Corresponding Author's Secondary Institution:	
First Author:	Laureen Dartois, PhD candidate
First Author Secondary Information:	
Order of Authors:	Laureen Dartois, PhD candidate Émilien Gauthier, PhD Julia Heitzmann, MSc Laura Baglietto, PhD Stefan Michiels, PhD Sylvie Mesrine, MD, PhD Marie-Christine Boutron-Ruault, MD, PhD Stéphane Ragusa, PhD Françoise Clavel-Chapelon, PhD Guy Fagherazzi, PhD
Order of Authors Secondary Information:	
Abstract:	<p>Purpose: Breast cancer remains a global health concern which lacks high discriminating prediction models. The k-nearest-neighbor algorithm estimates individual risks using an intuitive tool. This study aims at comparing the performances of this new approach with a Cox and the Gail models for the 5-year breast cancer risk prediction.</p> <p>Methods: The study included 64,995 women from the French E3N prospective cohort. The sample was divided into a learning (N=51,821) series to learn the models, using 5-fold cross-validation, and a validation (N=13,174) series to evaluate them. The area under the receiver operating characteristic curve (AUC) and the expected over observed number of cases (E/O) ratio were estimated. In the two series, 393 and 78 premenopausal and 537 and 98 postmenopausal breast cancers were diagnosed, respectively.</p>

	<p>Results: Discrimination values of the best combinations of predictors obtained from cross-validation ranged from 0.59 to 0.60. In the validation series, AUC values in premenopause and postmenopause were 0.583 [0.520; 0.646] and 0.621 [0.563; 0.679] using the k-nearest-neighbor algorithm and 0.565 [0.500; 0.631] and 0.617 [0.561; 0.673] using the Cox regression model, respectively. The E/O ratios were 1.26 and 1.28 in premenopause; 1.44 and 1.40 in postmenopause. The applied Gail model provided AUC values of 0.614 [0.554; 0.675] and 0.549 [0.495; 0.604] and E/O ratios of 0.78 and 1.12, respectively.</p> <p>Conclusions: This study highlighted that prediction performances differed according to menopausal status when using parametric statistical tool. The new k-nearest-neighbor approach performed well and discrimination was improved in postmenopause compared to the Gail model.</p>	
Suggested Reviewers:	<p>Karel GM Moons Julius Center for Health Sciences and Primary Care, UMC Utrecht k.g.m.moons@umcutrecht.nl</p> <p>Carlo La Vecchia Istituto Di Ricerche Farmacologiche Mario Negri lavecchia@marionegri.it</p> <p>Ruth M Pfeiffer Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America Pfeiffer@mail.nih.gov</p>	
Funding Information:	<p>Fondation de France (2011-00023087)</p> <p>Marie Curie International Incoming Fellowship</p> <p>Agence Nationale de la Recherche</p> <p>Inserm</p> <p>Fondation ARC pour la Recherche sur le Cancer</p>	<p>Laureen Dartois</p> <p>Laura Baglietto</p> <p>Sylvie Mesrine</p> <p>Marie-Christine Boutron-Ruault Françoise Clavel-Chapelon</p> <p>Guy Fagherazzi</p>

Manuscript

[Click here to download Manuscript: Manuscript.doc](#)

[Click here to view linked References](#)

November 7th, 2014

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Marc Lippman,
Editor-in-chief
Breast Cancer Research and Treatment

Dear Dr Lippman,

Please find attached our manuscript "A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort", which we would like you to consider for publication as an original research article in your Journal.

The manuscript aims to compare the performances of the Gail prediction model and two 5-year invasive breast cancer prediction models implemented on the E3N cohort: one based on a data-mining tool, the k-nearest-neighbor algorithm, and one based on a classic statistical tool, the Cox regression model. Predictions were separating according to the menopausal status of the women. Both approaches provided prediction models with close performances, both in terms of discrimination and calibration.

To our knowledge, the performances of the k-nearest-neighbor algorithm, a novel, intuitive and readable approach for a general audience, for the prediction of breast cancer risk have rarely been investigated. This study highlighted the benefit of this new approach and adds evidence to the literature of the relevance of providing separate prediction scores for invasive breast cancer according to menopausal status when using parametric statistical tool. We think that our manuscript might be of interest to the whole audience of your journal and beyond, to public health officials and policy makers.

I am the corresponding author. Please feel free to contact me by telephone, fax, or email as indicated below. I look forward your reply.

Sincerely yours,

Françoise Clavel-Chapelon, PhD
Inserm U1018, Gustave Roussy,
114 rue Edouard Vaillant
94805 Villejuif Cedex, France
Tel: (+33) 1 42 11 41 48 / Fax: (+33) 1 42 11 40 00
Email: francoise.clavel@gustaveroussy.fr

A comparison between different prediction models for invasive breast cancer occurrence in the French E3N cohort

Laureen Dartois^{1,2,3}, Émilien Gauthier⁴, Julia Heitzmann⁴, Laura Baglietto^{5,6}, Stefan Michiels^{2,7}, Sylvie Mesrine^{1,2,3}, Marie-Christine Boutron-Ruault^{1,2,3}, Stéphane Ragusa⁴, Françoise Clavel-Chapelon^{1,2,3}, Guy Fagherazzi^{1,2,3}

1 Inserm (Institut National de la Santé et de la Recherche Médicale), Centre for Research in Epidemiology and Population Health (CESP), U1018, Team 9, F-94805, Villejuif, France

2 Univ Paris Sud, UMRS 1018, F-94805, Villejuif, France

3 Gustave Roussy, F-94805, Villejuif, France

4 Statlife, Cancer Campus, Bat. Maurice Tubiana, Gustave Roussy, F-94805 Villejuif, France

5 Cancer Epidemiology Centre, Cancer Council of Victoria, Melbourne, Victoria, Australia,

6 Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Victoria, Australia

7 Service de biostatistique et d'épidémiologie, Gustave Roussy, F-94805, Villejuif, France

Correspondence to:

F. Clavel-Chapelon, francoise.clavel@gustaveroussy.fr

Inserm U1018 Team 9. 114 rue Édouard Vaillant. F-94805 Villejuif Cedex, France

Phone: +33 1 42 11 41 48

Fax: +33 1 42 11 40 00

Abbreviated running title: Invasive breast cancer risk prediction

1
2
3 **Key words:** Breast cancer; Women; Risk score; Proportional hazard Cox regression; Nearest-neighbor
4 algorithm; Discrimination; Calibration; Postmenopausal women; Premenopausal women; menopausal
5 status; Gail model
6
7

8
9
10
11
12
13 Abstract word count: 248
14

15
16 Text word count (excluding references): 4021
17

18
19 Total number of tables: 5
20

21
22 Total number of supplementary tables: 1
23

24
25 Total number of figures: 3
26

27
28 Total number of supplementary figures: 0
29

30
31 Total number of references: 66
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

ABSTRACT:

Purpose: Breast cancer remains a global health concern which lacks high discriminating prediction models. The k-nearest-neighbor algorithm estimates individual risks using an intuitive tool. This study aims at comparing the performances of this new approach with a Cox and the Gail models for the 5-year breast cancer risk prediction.

Methods: The study included 64,995 women from the French E3N prospective cohort. The sample was divided into a learning (N=51,821) series to learn the models, using 5-fold cross-validation, and a validation (N=13,174) series to evaluate them. The area under the receiver operating characteristic curve (AUC) and the expected over observed number of cases (E/O) ratio were estimated. In the two series, 393 and 78 premenopausal and 537 and 98 postmenopausal breast cancers were diagnosed, respectively.

Results: Discrimination values of the best combinations of predictors obtained from cross-validation ranged from 0.59 to 0.60. In the validation series, AUC values in premenopause and postmenopause were 0.583 [0.520; 0.646] and 0.621 [0.563; 0.679] using the k-nearest-neighbor algorithm and 0.565 [0.500; 0.631] and 0.617 [0.561; 0.673] using the Cox regression model, respectively. The E/O ratios were 1.26 and 1.28 in premenopause; 1.44 and 1.40 in postmenopause. The applied Gail model provided AUC values of 0.614 [0.554; 0.675] and 0.549 [0.495; 0.604] and E/O ratios of 0.78 and 1.12, respectively.

Conclusions: This study highlighted that prediction performances differed according to menopausal status when using parametric statistical tool. The new k-nearest-neighbor approach performed well and discrimination was improved in postmenopause compared to the Gail model.

INTRODUCTION:

Breast cancer remains a major global health concern as the most frequently diagnosed worldwide neoplasm among women with an incidence reaching 43.3 for 100,000 women-year in 2012 [1]. Despite a steady reduction in mortality rates in many countries, the global incidence continues to increase, with pronounced variation in the trend across countries [1, 2]. A risk prediction model is a statistical tool for estimating the individual probability that a currently healthy individual with specific predictors, derived from known epidemiological and clinical risk factors, will develop a future condition within a specific time period. Risk prediction models allow for recommendations for public health professionals about preventive interventions and personalized therapy and for patient about behavioral changes and how they were situated from average disease risk [3].

Risk prediction first received attention for the evaluation of cardiovascular risk. One of the most widely validated and used prediction tools is the Framingham Risk Score [4], designed to predict the 10-year absolute risk for coronary heart disease. The first risk prediction model for breast cancer was described by Gail *et al.* in 1989 [5] and was developed using data from a nested case-control study. The Gail model predicts the risk of developing invasive or in situ breast cancer in a defined age interval, taking into account as predictors age, age at menarche, number of previous breast biopsies, age at first live birth and number of first-degree relatives with breast cancer. Investigators have then further modified the original Gail model by estimating the absolute 5-year risk for invasive breast cancer only and adding non-modifiable risk factors (such as age at menopause, parity, ethnicity or height), modifiable risk factors (such as alcohol consumption, exogenous hormone use or physical activity), biomarkers (such as breast density) and multiple genetic variants [6-9]. In the United States, absolute risk estimates derived from the Gail model are used to identify high-risk women for mammography screening [10] and intervention trials [11].

The accuracy and generalizability of a predictive model are assessed using measures of discrimination (how well the model separates patients who experienced the event of interest from

1 those who did not) and calibration (agreement between observed outcome and prediction) [12, 13].
2 While well calibrated at the population-level, the breast cancer prediction models lack adequate
3 discrimination in predicting risk for individuals [6, 7, 14], which encourage the development of more
4 accurate risk models [3]. Among others, we selected the k-nearest-neighbor algorithm (kNN) [15], a
5 non-parametric statistical approach providing a simple, readable and intuitive tool for a general
6 audience, for the estimation of the 5-year prediction of invasive breast cancer risk. To our knowledge,
7 the derivation of predictive models for breast cancer using this algorithm have rarely been explored
8 [16, 17]. This article aims at comparing the performances of this original approach with the
9 performances of a classic statistical tool based on the Cox proportional hazard regression model and
10 the widely used Gail model. We estimated individual invasive breast cancer risk within 5 years using
11 these three approaches from an extended set of risk factors as potential predictors in the large
12 prospective E3N-EPIC cohort study.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

METHODS

Study population

The E3N (*Étude Épidémiologique auprès des femmes de la Mutuelle Générale de l'Éducation Nationale (MGEN)*) prospective cohort was launched in 1990 to investigate cancer risk factors [18]. The cohort includes 98,995 French women, aged 40-65 at inclusion, who were regularly asked about their health status and lifestyle by self-administered questionnaires, sent every 2 to 3 years. All women signed informed consent, in compliance with the rules of the French National Commission for Data Protection and Privacy from which approval was obtained.

For the present study, follow-up started at the date of return of the 1993 questionnaire, which first recorded dietary habits, thereafter considered as baseline. Responders (N=74,522) contributed person-years of follow-up until five years after baseline or the date of any invasive breast cancer diagnosis, whichever occurred first. We excluded women diagnosed with a cancer, other than non-basal skin, before baseline (N=4,705), those with no follow-up questionnaire (N=769), breast cancer cases with no specified date of diagnosis (N=54) and non-cases women with a follow-up shorter than 5 years (N=1,505). In addition, *in situ* cancer cases (N=270) and invasive cancer cases other than breast (N=864) occurring within the five first years after baseline were excluded. Our final population for analysis consisted of 64,995 women, aged 42-72 at baseline.

Cross-validation

Cross-validation is a method for evaluating the performance of a predictive score [19]. In this article, we first randomly divided (without replacement) the whole dataset into two series: a learning and a validation series, in order to learn and validate the final models, using the ratio 4:1 (N=51,821 and N=13,174, respectively). Differences between the learning and the validation series were tested using a between groups *t*-test for means differences and a Khi-2 test for proportions differences. Five-fold cross-validation was used within the learning sample, randomly divided (without replacement) into five learning partitions of roughly equal size (N=41,404, N=41,465, N=41,637, N=41,389 and

1 N=41,389, respectively). Each tested prediction score was implemented in four partitions, and
2 performances were estimated in the last one, namely the test partition. The five estimations obtained
3 were then averaged to provide the cross-validation performances of the tested models.
4
5

6 7 *Data collection* 8

9
10 Prediction models were derived separately for premenopausal and postmenopausal women for the
11 kNN and the Cox approaches, because of the heterogeneity observed according to age at diagnosis
12 with respect to hormonal receptor, tumor size, grade and nodal status [20-22], the V-relationship
13 between adiposity and the risk of breast cancer before and after menopause in non-Asian women [23,
14 24] and different magnitudes of risk observed according to age or menopausal status for several risk
15 factors, particularly menstrual and reproductive characteristics [25-28]. At baseline, women were
16 considered postmenopausal if they reported amenorrhea for more than 1 year (unless due to
17 hysterectomy) or if they self-reported that they were postmenopausal [29]. Potentially predictors
18 included age, first-degree family history of breast cancer, personal history of benign breast diseases
19 (BBD), age at menarche, number of children, age at first full-term pregnancy (FFTP), height at
20 adulthood, alcohol consumption, body mass index (BMI), age at menopause (for postmenopausal
21 women only) and any systemic use of estrogens alone or combined with a progestagens (for
22 postmenopausal women only). All variables were assessed at baseline.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 For each variable included in the analyses, missing values represented less than 5% of the
41 observations and were thus imputed to the modal category for categorical variables and to the median
42 value for quantitative variables. For the kNN algorithm, age at FFTP was imputed to 24 for
43 nulliparous women, corresponding to the median value in the learning sample for parous women,
44 while a separate category was considered for nulliparous women in Cox regression models.
45
46
47
48
49
50
51

52 *Ascertainment of cancer cases* 53

54 All questionnaires enquired about any cancer occurrence, type of cancer, addresses of physicians
55 and permission to contact them. Only invasive breast cancers occurring within the five first years after
56 baseline were considered and 98.3% were confirmed by pathology reports.
57
58
59
60
61
62
63
64
65

Prediction at 5 years

The k-nearest-neighbor algorithm

The kNN algorithm [15] is a non-parametric statistical model based on a various set of risk factors combinations (up to 5 different risk factors because of time-consuming compilation). All combinations with a size of 1 to 5 variables were tested both for premenopausal and postmenopausal women, among the 11 variables described above: age (<45, 45-50, 50-55, 55-60, 60-65, ≥65), first-degree family history of breast cancer (0, 1, ≥2), history of BBD (0, ≥1), age at menarche (<10, 10-15, ≥15), number of children (0, 1, 2, 3, ≥4), age at FFTP (<20, 20-25 or nulliparous women, 25-30, 30-35, 35-40, ≥40), alcohol consumption (0-10, 10-20, 20-30, 30-50, 50-70, ≥70), BMI (<18, 18-22.5, 22.5-28, 28-33, 33-38, ≥38), age at menopause (<45, 45-50, 50-55, ≥55), use of menopausal hormone therapy (MHT, no recent use, mixed use, oral estrogen only, oral estrogen with progesterone or dydrogesterone, oral estrogen with other progestagen, other MHT). For each combination of risk factors' values, coordinates were defined as the proportion of breast cancer within the subgroup of women with the same values for the risk factors. Several values for the length of the neighborhood were tested and a length of 20 neighbors was set, as providing the best performances. At least 20 of the nearest records of the database were included in the neighborhood, determined using the Euclidean distance. The length of the neighborhood may be higher than 20 because of individuals at equal distance. The score prediction for a woman was estimated as the proportion of breast cancers within her neighborhood.

The Cox model approach

Multivariate Cox proportional hazards regression models with date as the timescale (t) and the counting process style of input (start date, end date) were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) of breast cancer risk associated with the risk factors above-mentioned and their interactions. The Efron method was used for handling ties [30]. All quantitative variables were centered and reduced using the mean and the variance estimates in the learning sample. The probability of having a breast cancer within 5 years was then estimated by $P(5) = 1 - [S_0(5)]^{\exp(b'X)}$, with X the values of the risk factors and b the coefficient parameters. The

1 log baseline cumulative hazard function, $\ln(H_0) = \ln(-\ln(S_0))$, was estimated using the Breslow
2 method [31] and approximated by a restricted cubic spline function with seven knots [32]. All
3 quantitative variables were included in the models as continuous variables, except age at FFTP which
4 was categorized (≤ 20 , 20-25, 25-30, 30-35, >35 for parous women, and a separate category for
5 nulliparous women) because age at FFTP and number of children shared a common reference. Recent
6 MHT use was categorized as follow: no recent use, oral estrogen only, oral estrogen with progesterone
7 or dydrogesterone, oral estrogen with other progestagen, other MHT. All interaction terms between
8 two variables were tested.

9
10
11 A forward method, based on AUC values, was used to select the 15 variables (including interaction
12 terms) constituting the prediction models, from 1 to 15 predictors, associated with the highest AUC
13 estimates. All combinations of these 15 variables were tested and the best one was selected by
14 optimizing the values of the AUC and the regression coefficient of the prognostic index (PI, *i.e.*,
15 the weighted sum of the variables in the model, where the weights are the regression coefficients
16 estimated within the learning partitions).

17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 *The Gail model*

34
35
36 The 5-year Gail risk was calculated for each premenopausal and postmenopausal woman by using
37 the Breast Cancer Risk Assessment Tool, available at <http://brca.nci.nih.gov/brc/>.

38 39 40 41 42 *Scores comparison*

43
44 Pearson's correlation coefficients and concordances were estimated between the best premenopausal
45 and postmenopausal breast cancer risk scores provided by both approaches. For concordance
46 estimates, the scores were divided according to their quartiles of distribution within the validation
47 sample.

48
49 For illustration, ten hypothetical scenarios of predictors' values were considered in order to estimate
50 the prediction scores of breast cancer derived from the three approaches for these ten hypothetical
51 women.

Discrimination and calibration

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

The discrimination of a prediction model indicates how well the model separates patients who experienced the event of interest from those who did not at 5 years [33] and was evaluated using the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) [34]. Discriminations of prediction scores implemented using Cox regressions were also assessed by estimating the regression coefficient of the PI [35].

Model calibration describes how well the estimates of survival from the model correspond to the survival from the observed data [33]. Calibration was evaluated using the E/O ratio, with E representing the expected number of breast cancer cases and O the observed number of cases. E was estimated as the sum of all the prediction probability estimates within the test partitions. Calibration was illustrated by plotting the observed proportion of breast cancer cases versus the mean prediction score within four subsamples of the validation sample defining according to the quartiles of the prediction score.

KNN model development was performed using in-house Java-1.6 software. All analyses were conducted using SAS software, version 9.3 (SAS Institute Inc., Cary, North California, USA).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

RESULTS

Population characteristics

The 51,821 women included in the learning series were aged between 42 and 72 at study entry, with a mean age of 52.7 years (6.6 SD). During 256,843 person-years (median follow-up 2.6 years for cases and 5.0 years for non-cases), 930 women were diagnosed with invasive breast cancer (393 in premenopause and 537 in postmenopause). The main characteristics of the women in the learning series are shown in Table 1. Women diagnosed with a breast cancer during the follow-up were more likely to have at least one family history of breast cancer, to have at least one personal history of BBD, to have their first full-term pregnancy at an advanced age and to use MHT (for postmenopausal women only). The main characteristics of the women in the validation series are shown in Supplementary table 1. Minor differences in women characteristics between the training and validation series were observed. The average of the number of children differed across samples and was higher in the learning sample (p-value=0.01). In addition, a higher number of women had at least two family histories of breast cancer in the validation series (1.1%), than in the learning series (0.9%, p-value=0.02), and a higher number of women aged between 45 and 50 (32.0% vs. 30.9%, p-value=0.02), without statistically significant differences in the mean estimates. At last, a statistically significant higher number of diagnosed breast cancers in the learning series (1.8%) than in the validation series (1.3%) was observed (p=0.0003).

Premenopausal women

Best combinations, with the corresponding discrimination and calibration estimations, selected during the cross-validation in the learning series using the kNN and the Cox approaches, are presented in Table 2.

The best combination for the premenopausal women, derived from the kNN approach, was age, family history of breast cancer, history of BBD and age at menarche, with AUC value of 0.590 and E/O ratio of 1.00. Using a Cox regression model, the best combination included the same variables,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

except age at menarche, and the interaction term between family history of breast cancer and personal history of BBD, producing an AUC value of 0.603, a regression coefficient associated with the PI of 0.686 and an E/O ratio of 1.04. When implementing the kNN and the Cox best premenopausal combinations in the whole learning series and validating them on the premenopausal women of the validation series (Table 3), AUC values were 0.583 [0.520; 0.646] for the kNN score and 0.565 [0.500; 0.631] for the Cox model score; E/O ratios were 1.26 and 1.28, respectively. Calibration curves for premenopausal scores within the validation series were presented in Figure 1 for the Cox regression approach and Figure 2 for the kNN approach. Both prediction scores overestimated the number of diagnosed breast cancer cases in the validation series, particularly among high-risk women. When the postmenopausal risk scores were applied to the premenopausal women (Table 3), performances were similar for the kNN approach (AUC value of 0.590 [0.522; 0.658] and E/O ratio of 0.93). For the Cox approach, discrimination was similar (AUC value of 0.581 [0.514; 0.649]), but calibration was poorer (E/O ratio of 5.67).

Postmenopausal women

The best combination for postmenopausal women, derived from the kNN approach, was age, family history of breast cancer, history of BBD, age at menopause and recent MHT use, with AUC value of 0.592 and E/O ratio of 1.03. Using a Cox regression model, the best combination included these 5 variables in addition of alcohol consumption, BMI, number of children and the interaction terms between age and age at menopause and between age at menopause and personal history of BBD, producing an AUC value of 0.602, a regression coefficient associated with the PI of 0.561 and an E/O ratio of 1.05.

When implementing the kNN and the Cox best postmenopausal combinations in the whole learning series and validating them on the postmenopausal women of the validation series (Table 3), AUC values were 0.621 [0.563; 0.679] for the kNN score and 0.617 [0.561; 0.673] for the Cox model score; E/O ratios were 1.44 and 1.40, respectively. Calibration curves for postmenopausal scores within the validation series were presented in Figure 1 for the Cox regression approach and Figure 2 for the kNN

1 approach. Both prediction scores overestimated the number of diagnosed breast cancer cases,
2 particularly among high-risk women. When the premenopausal risk scores were applied to the
3 postmenopausal women, performances were similar for the kNN approach (AUC value of 0.627
4 [0.569; 0.685] and E/O ratio of 1.30). For the Cox approach, discrimination was lower (AUC value of
5 0.544 [0.487; 0.602]) but calibration was better (E/O ratio of 1.25).
6
7
8
9

10 *Comparison of the three approaches*

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

When comparing the best combination of the kNN and the Cox approaches and the Gail model applied to the premenopausal and the postmenopausal women (Table 3), we observed that correlations were statistically different from 0 (from 0.53 to 0.77 in premenopause and from 0.37 to 0.65 in postmenopause), but concordances were only modest (from 0.10 to 0.50). Performances of the kNN and the Cox approaches were very close. The Gail model provided a better discrimination (AUC value of 0.614 [0.554; 0.675]) and an inversed calibration (E/O ratio of 0.78) in premenopause and a lower discrimination (AUC value of 0.549 [0.495; 0.604]) and a better calibration (E/O ratio of 1.12) in postmenopause. Calibration curves for the Gail model (Figure 3) showed an underestimation of the number of premenopausal breast cancer cases and an overestimation of postmenopausal breast cancer cases for high-risk women.

66 *Hypothetical women*

67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Table 5 presents the estimated invasive breast cancer risk of 10 hypothetical women. In premenopause, risks of women between 45 and 55 years old ranked from 1.1% to 7.8%. In postmenopause, risks of women between 45 and 70 years old ranked from 0.5% to 11.2%. The trends of the three risk scores were the same using the three methods; however the range of predicted values were larger using the Cox regression or the Gail models than using the kNN algorithm.

DISCUSSION

Our results

When implemented the 5-year invasive breast cancer risk for women aged between 50 and 70 using the kNN algorithm and the Cox regression model, performances of the selected premenopausal and postmenopausal scores were close, despite some differences in the selected predictors. Predictors selected using the kNN algorithm were age, family history of breast cancer, personal history of BBD and age at menarche in premenopause, and age, family history of breast cancer, personal history of BBD, age at menopause and recent MHT use in postmenopause. Using the Cox regression model, the selected predictors were similar, except that age at menarche was not selected in premenopause and three additional variables (alcohol consumption, number of children and BMI) were selected in postmenopause. The Gail model, applied to the premenopausal and postmenopausal women of our population analysis produced higher discrimination value in premenopause, but lower in postmenopause. Both in premenopause and postmenopause, the prediction scores had only modest discrimination values (from 0.55 to 0.62). Performances of the three scores were different according to menopausal status: both the kNN and the Cox scores presented higher discrimination in postmenopause, while the Gail model presented higher discrimination in premenopause. Moreover, using the Cox approach, scores performances were better when the scores were applied in the targeted population (*i.e.*, the premenopausal scores on premenopausal women and the postmenopausal scores on postmenopausal women). On the contrary, scores performances were similar using the kNN algorithm when applying both scores on the targeted and the non-targeted populations.

Literature confrontation

Behavioral factors seemed to hardly impact the 5-year premenopausal breast cancer risk, as none was retained, and to have a greater impact in postmenopause than in premenopause. This result was consistent with previously published studies which concluded that the proportion of breast cancers attributable to behavioral factors were lower in premenopause than in postmenopause [36]. In the

1 literature, previously published breast cancer risk scores produced AUC values ranged from 0.54 to
2 0.68 [37-53] and E/O ratios from 0.85 to 1.10 [37-45]. Performances of the premenopausal and
3 postmenopausal prediction scores from box kNN and Cox approaches were within the range of the
4 literature for performances obtained using the cross-validation method within the learning sample.
5
6 However, E/O ratios of the kNN and the Cox prediction scores were higher than the range of the
7 literature when validate the final models in the validation series, producing scores which
8 overestimated the number of predicted invasive breast cancer cases. This may result from differences
9 between the two series (particularly, the statistically significant higher number of diagnosed breast
10 cancers observed in the learning series than in the validation series). The number of predictors retained
11 for the previously published prediction risk scores ranged from 3 to 15 [5, 8, 37-61]. To our
12 knowledge, only two previously published risk scores implemented different scores according to the
13 menopausal status of the women [48, 54], and both selected a higher number of predictors in
14 postmenopause than in premenopause. While some previously published breast cancer risk prediction
15 scores insisted in the importance of reproductive factors [46, 50, 55, 62], prediction scores obtained in
16 this study showed only modest prediction power of age at menarche, age at FFTP and number of
17 children. Performances of the breast cancer risk scores may be improved by the addition of other
18 predictors, particularly breast density, as previously published scores using this variable produced
19 higher AUC estimates (from 0.62 to 0.68) [41, 47, 48, 53].
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41 *Strengths and limitations*

42
43 This study has some limitations which should be considered. Categorizations used in the kNN and
44 the Cox approaches were not similar. However comparison between these two approaches produced
45 similar models performances. A low number of predictors (up to 5) was taking into account in the
46 kNN approach because of time consuming implementation. Age at menarche was retrospectively
47 collected from women aged 40-65 years and thus requires long-term memory, which may generate
48 measurement errors. However, an independent validation study showed that this information was
49 reliable (correlations between original and recalled was 0.8) [63]. We considered in the analyses the
50 number of children instead of the number of deliveries (for 906 women in the learning series and 240
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

in the validation series, these two variables were different because of multiple births); however, sensitivity analyses replacing the number of children by the number of deliveries provided similar results for both approaches. While a wide set of potential predictors of breast cancer risk was taken into account, some risk factors were lacking. Sensitivity analyses testing the addition of one risk factor associated with modest association with breast cancer (birth weight, body shape at menarche, breastfeeding, history of alcohol consumption and number of melanocytic nevi [64]) and not listed among the potential predictors led to similar model performances. However, testing a wider set of risk factors might lead to better performances, particularly breast density or genetic factors.

The main strength of these analyses is the size of the cohort enabling a high statistical power and a wide set of available risk factors. The study design also enabled to explore the heterogeneity of subgroups of breast cancer defined by its menopausal status. Histological confirmation was obtained for the vast majority of breast cancer cases. Because of the prospective design, information on exposure collected prior to cancer diagnosis resulted in low risk of recall bias. In addition, dietary and anthropometric data were validated [65, 66], limiting declaration bias and demonstrating the reliability of the reported data.

Conclusion

This study highlighted that the new kNN approach provided similar performances than a prediction score implemented using a Cox regression model and that scores discrimination was improved compared to the Gail model in postmenopause. Further research is needed in order to increase the number of predictors and the number of categories to take into account in the algorithm and then increase the range of estimated individual risk. This study adds evidence to the literature of the importance of separating prediction risk scores of breast cancer according to menopausal status when using a parametric statistical tool, which appeared less appropriate when using the kNN algorithm. In addition, this study highlighted the relevance of taking into account a wide set of potentially predictors, including risk factors associated with only modest increased risk of breast cancer in the literature.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgments: The authors are indebted to all participants for providing data and to practitioners for providing pathology reports. They are grateful to all members of the E3N-EPIC study group, particularly to Gaëlle Gusto for her help on statistical design, Rafika Chaït for her management on cancer database, and Marie Fangon, Pascale Gerbouin-Rérolle, Lyan Hoang, Céline Kernaléguen, Camille Laplanche, Maryvonne Niravong and Maxime Valdenaire for their technical assistance.

Contribution: Laureen Dartois, Guy Fagherazzi, Émilien Gauthier and I designed the analysis plan. Laureen Dartois and Julia Heitzmann performed the statistical analyses, Laureen Dartois wrote the manuscript. Laureen Dartois, Julia Heitzmann, Émilien Gauthier, Laura Baglietto, Stefan Michiels, Marie-Christine Boutron-Ruault, Stéphane Ragusa, Sylvie Mesrine, Guy Fagherazzi and I critically revised the manuscript and approved the final version.

Sponsoring: The E3N study was funded by the MGEN (Mutuelle Générale de l'Éducation Nationale), the Ligue contre le Cancer, Gustave Roussy and the Institut National de la Santé et de la Recherche Médicale (Inserm), and received grants from the "Cohortes Santé TGIR Program", Direction générale de la santé (DGS), European Union Research executive Agency, Inserm-IRESP, Fondation cœur et artères (Fondacoœur), Fondation pour la recherche médicale (FRM), Agence Nationale de la Recherche (ANR), Institut National du Cancer (INCa), Fondation ARC pour la recherche sur le cancer, World Cancer Research Fund (WCRF), Ligue contre le Cancer, Fondation de France, Agence Nationale de Sécurité du Médicament et des produits de santé (ANSM), the French Ministry of Health and the Université Paris Sud. For the submitted work, MCBR and FCC were supported by Inserm, LD by the Fondation de France (2011-00023087), SMe by ANR, LB by a Marie Curie International Incoming Fellowship within the European Community Framework Programme and GF by the Fondation ARC.

Ethical approval: The E3N study received approval from the national commission for computed data and liberties ("Commission Nationale de l'Informatique et des Libertés"), classification AG662, 12489 (September 25th, 1988). All women signed an informed consent.

1 **Data sharing:** The dataset can be requested from the principal investigator of the E3N study at
2 francoise.clavel@gustaveroussy.fr
3
4

5 **Conflicts of interest:** EG is an employee of Statlife, JH was an employee of Statlife and SR is a
6 shareholder of Statlife. The other authors declare no conflict of interest.
7
8
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, *et al.* GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer 2013; Available from: <http://globocan.iarc.fr>, accessed on 22/04/2014.
2. Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, *et al.* Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* 2011; **378(9801)**:1461-1484.
3. Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA, Ballard-Barbash R, *et al.* Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst* 2005; **97(10)**:715-723.
4. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *Am J Cardiol* 1976; **38(1)**:46-51.
5. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989; **81(24)**:1879-1886.
6. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. *Breast Cancer Res Treat* 2012; **132(2)**:365-377.
7. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinian A. Risk prediction models of breast cancer: a systematic review of model performances. *Breast Cancer Res Treat* 2012; **133(1)**:1-10.
8. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, *et al.* Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010; **362(11)**:986-993.
9. van ZM, van der Net JB, Kundu S, Freedman AN, van Duijn CM, Janssens AC. Genome-based prediction of breast cancer risk in the general population: a modeling study based on meta-analyses of genetic associations. *Cancer Epidemiol Biomarkers Prev* 2011; **20(1)**:9-22.
10. Smith RA, Manassaram-Baptiste D, Brooks D, Cokkinides V, Doroshenk M, Saslow D, *et al.* Cancer screening in the United States, 2014: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2014; **64(1)**:30-51.
11. Visvanathan K, Hurley P, Bantug E, Brown P, Col NF, Cuzick J, *et al.* Use of pharmacologic interventions for breast cancer risk reduction: American Society of Clinical Oncology clinical practice guideline. *J Clin Oncol* 2013; **31(23)**:2942-2962.
12. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, *et al.* Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98(9)**:691-698.
13. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21(1)**:128-138.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
14. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J Natl Cancer Inst* 2010; **102**(10):680-691.
15. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 1967; **13**(1):21-27.
16. Gauthier E, Brisson L, Lenca P, Ragusa S. Breast cancer risk score: a data-mining approach to improve readability. *The International Conference on Data Mining CSREA Press* 2011:15-21.
17. Gauthier E, Brisson L, Lenca P, Clavel-Chapelon F, Ragusa S. Challenges to building a platform for a breast cancer risk score. *Sixth International Conference on Research Challenges in Information Science IEEE* 2012:1-10.
18. Clavel-Chapelon F. Cohort Profile: The French E3N Cohort Study. *Int J Epidemiol* 2014.
19. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Statistics Surveys* 2010; **4**:40-79.
20. Jatoi I, Anderson WF. Qualitative age interactions in breast cancer studies: a mini-review. *Future Oncol* 2010; **6**(11):1781-1788.
21. Garcia-Closas M, Brinton LA, Lissowska J, Chatterjee N, Peplonska B, Anderson WF, *et al*. Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer* 2006; **95**(1):123-129.
22. Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. *Breast Cancer Res Treat* 2002; **76**(1):27-36.
23. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008; **371**(9612):569-578.
24. Fagherazzi G, Chabbert-Buffet N, Fabre A, Guillas G, Boutron-Ruault MC, Mesrine S, *et al*. Hip circumference is associated with the risk of premenopausal ER-/PR- breast cancer. *Int J Obes (Lond)* 2012; **36**(3):431-439.
25. Clavel-Chapelon F, Launoy G, Auquier A, Gairard B, Bremond A, Piana L, *et al*. Reproductive factors and breast cancer risk. Effect of age at diagnosis. *Ann Epidemiol* 1995; **5**(4):315-320.
26. Bouchardy C, Le MG, Hill C. Risk factors for breast cancer according to age at diagnosis in a French case-control study. *J Clin Epidemiol* 1990; **43**(3):267-275.
27. Tryggvadottir L, Tulinius H, Eyfjord JE, Sigurvinsson T. Breast cancer risk factors and age at diagnosis: an Icelandic cohort study. *Int J Cancer* 2002; **98**(4):604-608.
28. Trentham-Dietz A, Sprague BL, Hampton JM, Miglioretti DL, Nelson HD, Titus LJ, *et al*. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. *Breast Cancer Res Treat* 2014; **145**(1):165-175.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
29. Fournier A, Berrino F, Riboli E, Avenel V, Clavel-Chapelon F. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. *Int J Cancer* 2005; **114**(3):448-454.
 30. Arnaud L, Georges G, Sorrentini J, Zerrad M, Deumie C, Amra C. An enhanced contrast to detect bulk objects under arbitrary rough surfaces. *Opt Express* 2009; **17**(7):5758-5773.
 31. Breslow NE. Discussion of the paper by D. R. Cox. *J R Statist Soc B* 1972; **34**:216-217.
 32. Smith PL. Splines as a useful and convenient statistical tool. *The American Statistician* 1979; **33**:57-62.
 33. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**:b605.
 34. Harrel FE Jr. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. *New York: Springer-Verlag* 2001.
 35. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13**:33.
 36. Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. Association between five lifestyle habits and cancer risk: results from the E3N cohort. *Cancer Prev Res (Phila)* 2014; **7**(5):516-525.
 37. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, *et al*. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst* 2007; **99**(23):1782-1792.
 38. Matsuno RK, Costantino JP, Ziegler RG, Anderson GL, Li H, Pee D, *et al*. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J Natl Cancer Inst* 2011; **103**(12):951-961.
 39. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT, *et al*. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS Med* 2013; **10**(7):e1001492.
 40. Boyle P, Mezzetti M, La Vecchia C., Franceschi S, Decarli A, Robertson C. Contribution of three components to individual cancer risk predicting breast cancer risk in Italy. *Eur J Cancer Prev* 2004; **13**(3):183-191.
 41. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med* 2008; **148**(5):337-347.
 42. Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. *J Natl Cancer Inst* 2006; **98**(23):1686-1693.
 43. Pastor-Barriuso R, Ascunce N, Ederra M, Erdozain N, Murillo A, Ales-Martinez JE, *et al*. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. *Breast Cancer Res Treat* 2013; **138**(1):249-259.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
44. Gao F, Machin D, Chow KY, Sim YF, Duffy SW, Matchar DB, *et al.* Assessing risk of breast cancer in an ethnically South-East Asia population (results of a multiple ethnic groups study). *BMC Cancer* 2012; **12**:529.
 45. Petracci E, Decarli A, Schairer C, Pfeiffer RM, Pee D, Masala G, *et al.* Risk factor modification and projections of absolute breast cancer risk. *J Natl Cancer Inst* 2011; **103**(13):1037-1048.
 46. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J Natl Cancer Inst* 2004; **96**(3):218-228.
 47. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treat* 2005; **94**(2):115-122.
 48. Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, *et al.* Prospective breast cancer risk prediction model for women undergoing screening mammography. *J Natl Cancer Inst* 2006; **98**(17):1204-1214.
 49. Chlebowski RT, Anderson GL, Lane DS, Aragaki AK, Rohan T, Yasmeen S, *et al.* Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst* 2007; **99**(22):1695-1705.
 50. Rosner B, Colditz GA, Iglehart JD, Hankinson SE. Risk prediction models with incomplete data with application to prediction of estrogen receptor-positive breast cancer: prospective data from the Nurses' Health Study. *Breast Cancer Res* 2008; **10**(4):R55.
 51. Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction model expanded to include category of prior benign breast disease lesion. *Cancer* 2010; **116**(21):4944-4953.
 52. Tice JA, Miike R, Adduci K, Petrakis NL, King E, Wrensch MR. Nipple aspirate fluid cytology and the Gail model for breast cancer risk assessment in a screening population. *Cancer Epidemiol Biomarkers Prev* 2005; **14**(2):324-328.
 53. Chen WY, Colditz GA. Risk factors and hormone-receptor status: epidemiology, risk-prediction models and treatment implications for breast cancer. *Nat Clin Pract Oncol* 2007; **4**(7):415-423.
 54. Ueda K, Tsukuma H, Tanaka H, Ajiki W, Oshima A. Estimation of individualized probabilities of developing breast cancer for Japanese women. *Breast Cancer* 2003; **10**(1):54-62.
 55. 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):950-964.
 56. Lee EO, Ahn SH, You C, Lee DS, Han W, Choe KJ, *et al.* Determining the main risk factors and high-risk groups of breast cancer using a predictive model for breast cancer risk assessment in South Korea. *Cancer Nurs* 2004; **27**(5):400-406.
 57. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat Med* 2004; **23**(7):1111-1130.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
58. Novotny J, Pecen L, Petruzelka L, Svobodnik A, Dusek L, Danes J, *et al.* Breast cancer risk assessment in the Czech female population—an adjustment of the original Gail model. *Breast Cancer Res Treat* 2006; **95**(1):29-35.
59. Gjorgov NA. Breast cancer risk assessments to barrier contraception exposure. A new approach. *Prilozi* 2009; **30**(1):217-232.
60. Cook NR, Rosner BA, Hankinson SE, Colditz GA. Mammographic screening and risk factors for breast cancer. *Am J Epidemiol* 2009; **170**(11):1422-1432.
61. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst* 1996; **88**(6):359-364.
62. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J Natl Cancer Inst* 1996; **88**(6):359-364.
63. Must A, Phillips SM, Naumova EN, Blum M, Harris S, Dawson-Hughes B, *et al.* Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am J Epidemiol* 2002; **155**(7):672-679.
64. Kvaskoff M, Bijon A, Mesrine S, Vilier A, Baglietto L, Fournier A, *et al.* Association between melanocytic nevi and risk of breast diseases: The French E3N prospective cohort. *PLoS Med* 2014; **11**(6):e1001660.
65. Tehard B, van Liere MJ, Com NC, Clavel-Chapelon F. Anthropometric measurements and body silhouette of women: validity and perception. *J Am Diet Assoc* 2002; **102**(12):1779-1784.
66. van Liere MJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. *Int J Epidemiol* 1997; **26** Suppl 1:S128-S136.

Figure

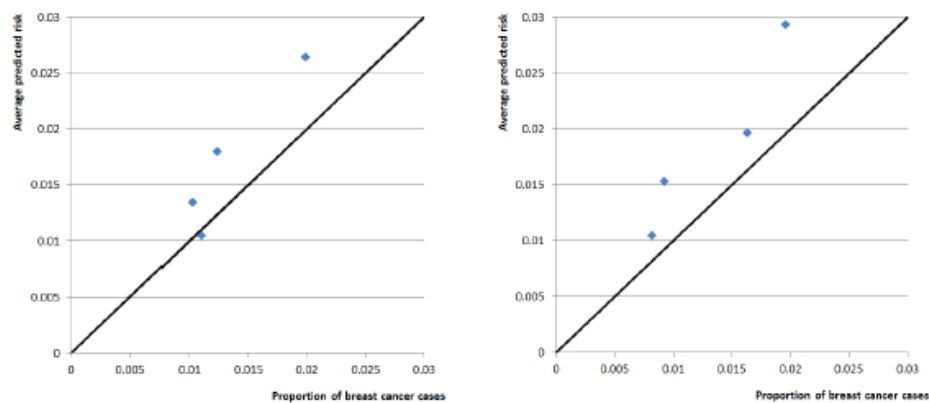
[Click here to download Figure: Figures.doc](#)

Figure 1 Calibration curves for the premenopausal (left) and the postmenopausal (right) best score obtained from the Cox regression approach, validation sample (N=13,174)

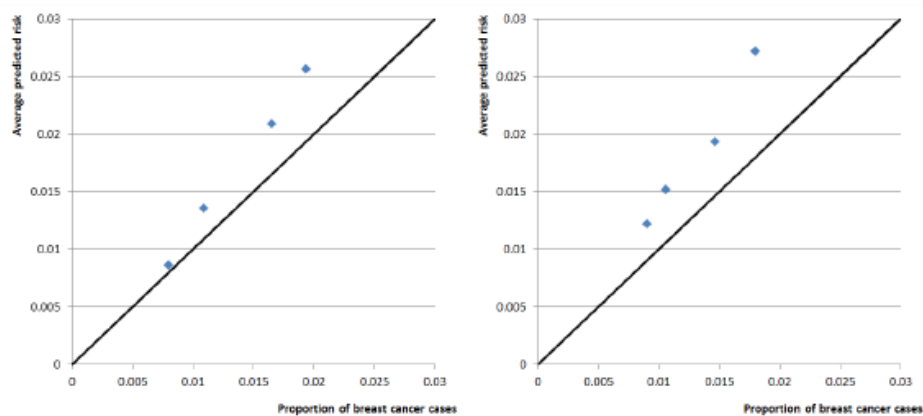


Figure 2 Calibration curves for the premenopausal (left) and the postmenopausal (right) best score obtained from the kNN approach, validation sample (N=13,174)

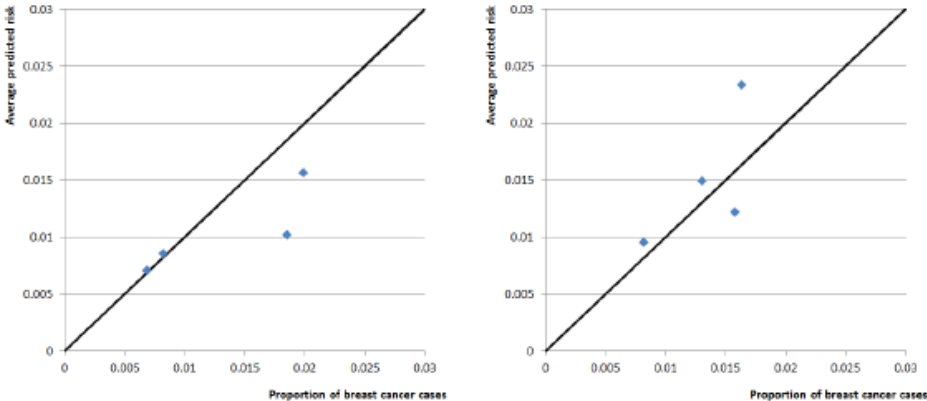


Figure 3 Calibration curves for the premenopausal (left) and the postmenopausal (right) score obtained from the Gail model, validation sample (N=13,174)

Table

[Click here to download Table: Tables.doc](#)

Table 1 Baseline characteristics of the learning sample, E3N cohort (N=51,821).

	Invasive breast cancer cases		
	Non cases N=50,891	Premenopausal N=393	Postmenopausal N=537
Qualitative factors, N (%)			
Menopausal status			
Premenopausal	22,969 (45.1%)	393 (100%)	-
Postmenopausal	27,922 (54.9%)	-	537 (100%)
First-degree family history of breast cancer ^a			
0	45,017 (88.5%)	330 (84.0%)	439 (81.8%)
1	5,414 (10.6%)	58 (14.7%)	87 (16.2%)
≥ 2	460 (0.9%)	5 (1.3%)	11 (2.0%)
History of BBD ^b			
0	36,075 (70.9%)	225 (57.3%)	338 (62.9%)
≥ 1	14,816 (29.1%)	168 (42.7%)	199 (37.1%)
Recent MHT use ^c (for postmenopausal women)			
Oral oestrogen only	1,493 (2.9%)	-	27 (5.0%)
Oral oestrogen with progesterone/dydrogesterone	5,159 (10.1%)	-	117 (21.8%)
Oral oestrogen with other progestagen ^d	4,225 (8.3%)	-	88 (16.4%)
Other ^e	773 (1.5%)	-	18 (3.3%)
No recent MHT ^f	17,033 (61.0%)	-	301 (56.1%)
Quantitative factors, mean (STD)			
Age (years)	53.54 (6.34)	48.03 (2.83)	57.57 (5.02)
Height at adulthood (cm)	161.82 (5.59)	162.37 (5.68)	161.42 (5.49)
Age at menarche (years)	12.78 (1.34)	12.73 (1.27)	12.82 (1.39)
Number of children	1.93 (1.15)	1.87 (1.05)	1.97 (1.21)
Age at first full-term pregnancy (for parous women)	24.90 (3.96)	25.61 (4.60)	25.05 (3.93)
Alcohol consumption (g/day)	12.22 (13.84)	11.18 (13.63)	12.97 (13.95)
Body mass index (kg/m ²)	23.04 (3.29)	22.34 (2.94)	23.55 (3.44)
Age at menopause (years, for postmenopausal women)	49.46 (4.14)	-	50.09 (3.54)

a: Family history of breast cancer included history among first-degree relatives (mother, sisters and daughters).

b: BBD included adenomas and fibrocystic diseases.

c: Recent MHT use was defined as any MHT use within the 12 preceding months.

d: Other progestagen included chlormadinone acetate, cyproterone acetate, demegestone, dienogest, drospirenone, ethynodiol acetate, gestodene, levonorgestrel, lynestrenol, medrogestone, medroxyprogesterone acetate, megestrol acetate, nomegestrol acetate, norethisterone acetate, and promegestone.

e: Other MHT included tibolone and MHT containing an androgen, intramuscularly administered, or with no specified formulation.

f: 720 women (706 (1.4%) non cases women and 14 (2.6%) postmenopausal cases women) used mixed MHT.

Table 2 Best combinations according to performance values from cross-validation, using both approaches, E3N cohort (N=51,821).

	Number of predictors	AUC	E/O	Regression coefficient of the prognostic index
Premenopausal score (N=23,362)				
kNN : Age, family history of breast cancer, personal history of BBD, age at menarche	4	0.590	1.00	- ^a
Cox : Age, family history of breast cancer, personal history of BBD (with interaction term between family history of breast cancer and personal history of BBD)	3	0.603	1.04	0.686
Postmenopausal score (N=28,459)				
kNN : Age, family history of breast cancer, personal history of BBD, age at menopause and recent MHT use	5	0.595	1.04	- ^a
Cox : Age, family history of breast cancer, personal history of BBD, age at menopause, recent MHT use, alcohol consumption, number of children, BMI (with interaction terms between age and age at menopause, age at menopause and personal history of BBD)	8	0.602	1.05	0.561

a. The regression coefficient of the prognostic index can be estimated only in a regression model.

Table 3 Comparison between the best model obtained from both approaches in the validation data set, E3N cohort (N=13,174).

	AUC			E/O		
	kNN	Cox	Gail	kNN	Cox	Gail
Premenopausal women (N=5,843)						
Pre-menopausal score prediction ^b	0.583 [0.520; 0.646]	0.565 [0.500; 0.631]	0.614 [0.554; 0.675]	1.26	1.28	0.78
Postmenopausal score prediction ^c	0.590 [0.522; 0.658]	0.581 [0.514; 0.649]	- ^a	0.93	5.67	- ^a
Postmenopausal women (N=7,331)						
Pre-menopausal score prediction	0.627 [0.569; 0.685]	0.544 [0.487; 0.602]	- ^a	1.30	1.25	- ^a
Postmenopausal score prediction	0.621 [0.563; 0.679]	0.617 [0.561; 0.673]	0.549 [0.495; 0.604]	1.44	1.40	1.12

a: The Gail model provides a unique score prediction for both premenopausal and postmenopausal women.

b: Predictors of the premenopausal scores were age, family history of breast cancer, personal history of BBD and age at menarche for the kNN algorithm, age, family history of breast cancer, personal history of BBD for the Cox regression model and age, age at menarche, number of previous breast biopsies, age at first live birth and family history of breast cancer for the Gail model.

c: Predictors of the postmenopausal scores were age, family history of breast cancer, personal history of BBD, age at menopause and recent MHT use for the kNN algorithm, age, family history of breast cancer, personal history of BBD, age at menopause, recent MHT use, alcohol consumption, number of children and BMI for the Cox regression model and age, age at menarche, number of previous breast biopsies, age at first live birth and family history of breast cancer for the Gail model.

Table 4 Correlation and concordance estimates between the best model obtained from both approaches and the Gail model in the validation data set, E3N cohort (N=13,174).

Correlation Concordance	Best Cox	Best kNN	Gail model
	Premenopausal woman (N=5,843)		
Best Cox	1	0.7747* 0.4977	0.5506* 0.1550
Best kNN		1	0.5274* 0.2102
Gail model			1
Postmenopausal woman (N=7,331)			
Best Cox	1	0.6540* 0.2596	0.5137* 0.1497
Best kNN		1	0.3691* 0.1021
Gail model			1

*: p<0.001

Table 5 Estimated breast cancer risk for 10 different scenarios of predictors' values.

Menopausal status	Age of menarche	Age	Family history of breast cancer	History of EBD	Age at menopause	Alcohol consumption	BMI	Number of children	MHT	5-year prediction risk (%)		
										Cox regression	k-nearest neighbors	Gail model
0	15	45	0	0	-	-	-	-	-	1.06	1.23	0.54
0	12	50	1	1	-	-	-	-	-	2.75	2.65	2.29
0	10	55	2	1	-	-	-	-	-	4.19	3.80	7.78
1	-	45	0	0	45	0	20	4	No MHT	0.49	0.93	0.54
1	-	50	0	0	45	0	22	2	No MHT	0.73	1.25	0.63
1	-	55	0	1	50	10	25	2	Oral estrogen	2.18	2.04	1.08
1	-	60	1	0	50	20	25	1	Other MHT ^a	3.18	2.00	2.93
1	-	65	1	0	55	20	30	1	Oral estrogen with progesterone/ ^b dydrogesterone	4.73	4.06	3.38
1	-	65	2	1	60	20	30	0	Oral estrogen with other progestins ^b	7.75	4.48	10.78
1	-	70	3	1	65	30	35	0	Oral estrogen with progesterone/ ^b dydrogesterone	10.42	4.80	11.20

a. Other MHT included tibolone and MFT containing an androgen, intramuscularly administered, or with no specified formulation.
 b. Other progestagen included chlormadinone acetate, cyproterone acetate, demegestone, dienogest, drospirenone, ethynodiol acetate, gestodene, levonorgestrel, lynestrol, medrogestone, medroxyprogesterone acetate, megestrol acetate, norgestrol acetate, norethisterone acetate, and promegestone

RÉFÉRENCES

1. Place AE, Jin HS, Polyak K. The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Res.* 2011;13:227.
2. Ferlay J, Soerjomataram I, Ervik M et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. *Lyon, France: International Agency for Research on Cancer* 2013; Available from: <http://globocan.iarc.fr>, consulté le 22/04/2014.
3. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 - Tumeurs solides. *Saint Maurice : Institut de veille sanitaire* 2013;122 p.
4. Monnereau A, Remontet L, Maynadie M et al. Estimation nationale de l'incidence des cancers en France entre 1980 et 2012. Partie 2 - Hémopathies malignes. *Saint Maurice : Institut de veille sanitaire* 2013;88 p.
5. Molinie F, Vanier A, Woronoff AS et al. Trends in breast cancer incidence and mortality in France 1990-2008. *Breast Cancer Res.Treat.* 2014;147:167-75.
6. Anderson WF, Rosenberg PS, Prat A, Perou CM, Sherman ME. How many etiological subtypes of breast cancer: two, three, four, or more? *J Natl.Cancer Inst.* 2014;106.
7. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177-82.
8. Perou CM, Sorlie T, Eisen MB et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.
9. Sorlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc.Natl.Acad.Sci.U.S.A* 2001;98:10869-74.
10. Weigelt B, Horlings HM, Kreike B et al. Refinement of breast cancer classification by molecular characterization of histological special types. *J Pathol.* 2008;216:141-50.
11. Herschkowitz JI, Simin K, Weigman VJ et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* 2007;8:R76.
12. 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;358:1389-99.
13. Evans JP, Skrzynia C, Susswein L, Harlan M. Genetics and the young woman with breast cancer. *Breast Dis.* 2005;23:17-29.
14. Ban KA, Godellas CV. Epidemiology of Breast Cancer. *Surg.Oncol.Clin.N.Am.* 2014;23:409-22.
15. Nelson HD, Huffman LH, Fu R, Harris EL. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force. *Ann.Intern.Med.* 2005;143:362-79.
16. Keen JC, Davidson NE. The biology of breast carcinoma. *Cancer* 2003;97:825-33.

RÉFÉRENCES

17. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. *Oncogene* 2006;25:5898-905.
18. Thompson D, Easton D. The genetic epidemiology of breast cancer genes. *J.Mammary.Gland.Biol.Neoplasia*. 2004;9:221-36.
19. Song M, Lee KM, Kang D. Breast cancer prevention based on gene-environment interaction. *Mol.Carcinog*. 2011;50:280-90.
20. Key TJ, Pike MC. The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. *Eur.J.Cancer Clin.Oncol*. 1988;24:29-43.
21. Begg L, Kuller LH, Gutai JP, Caggiula AG, Wolmark N, Watson CG. Endogenous sex hormone levels and breast cancer risk. *Genet.Epidemiol*. 1987;4:233-47.
22. Korenman SG. The endocrinology of breast cancer. *Cancer* 1980;46:874-8.
23. Clavel-Chapelon F, Gerber M. Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? *Breast Cancer Res.Treat*. 2002;72:107-15.
24. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol*. 2012;13:1141-51.
25. Clavel-Chapelon F. Cumulative number of menstrual cycles and breast cancer risk: results from the E3N cohort study of French women. *Cancer Causes Control* 2002;13:831-8.
26. Titus-Ernstoff L, Longnecker MP, Newcomb PA et al. Menstrual factors in relation to breast cancer risk. *Cancer Epidemiol.Biomarkers Prev*. 1998;7:783-9.
27. Liu Q, Wu J, Lambe M, Hsieh SF, Ekblom A, Hsieh CC. Transient increase in breast cancer risk after giving birth: postpartum period with the highest risk (Sweden). *Cancer Causes Control* 2002;13:299-305.
28. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. *Lancet* 2002;360:187-95.
29. Lee SH, Akuete K, Fulton J, Chelmow D, Chung MA, Cady B. An increased risk of breast cancer after delayed first parity. *Am.J.Surg*. 2003;186:409-12.
30. Beral V, Bull D, Doll R, Peto R, Reeves G. Breast cancer and abortion: collaborative reanalysis of data from 53 epidemiological studies, including 83 000 women with breast cancer from 16 countries. *Lancet* 2004;363:1007-16.
31. Key TJ, Appleby PN, Reeves GK et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. *Lancet Oncol*. 2013;14:1009-19.
32. Key TJ, Appleby PN, Reeves GK et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br.J.Cancer* 2011;105:709-22.
33. Kaaks R, Berrino F, Key T et al. Serum sex steroids in premenopausal women and breast cancer risk within the European Prospective Investigation into Cancer and Nutrition (EPIC). *J.Natl.Cancer Inst*. 2005;97:755-65.
34. Hankinson SE, Eliassen AH. Endogenous estrogen, testosterone and progesterone levels in relation to breast cancer risk. *J.Steroid Biochem.Mol.Biol*. 2007;106:24-30.

35. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J.Natl.Cancer Inst.* 2004;96:1856-65.
36. Noda M. International Agency for Research on Cancer (<http://www.iarc.fr/>). *Jpn.J.Clin.Oncol.* 1999;29:592.
37. Kahlenborn C, Modugno F, Potter DM, Severs WB. Oral contraceptive use as a risk factor for premenopausal breast cancer: a meta-analysis. *Mayo Clin.Proc.* 2006;81:1290-302.
38. Zhu H, Lei X, Feng J, Wang Y. Oral contraceptive use and risk of breast cancer: a meta-analysis of prospective cohort studies. *Eur.J.Contracept.Reprod.Health Care* 2012;17:402-14.
39. Anothaisintawee T, Wiratkapun C, Lerdsitthichai P et al. Risk factors of breast cancer: a systematic review and meta-analysis. *Asia Pac.J.Public Health* 2013;25:368-87.
40. Breast cancer and hormonal contraceptives: further results. Collaborative Group on Hormonal Factors in Breast Cancer. *Contraception* 1996;54:1S-106S.
41. Fournier A, Dossus L, Mesrine S et al. Risks of endometrial cancer associated with different hormone replacement therapies in the E3N cohort, 1992-2008. *Am.J.Epidemiol.* 2014;180:508-17.
42. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* 1997;350:1047-59.
43. Fournier A, Berrino F, Clavel-Chapelon F. Unequal risks for breast cancer associated with different hormone replacement therapies: results from the E3N cohort study. *Breast Cancer Res.Treat.* 2008;107:103-11.
44. Fournier A, Berrino F, Riboli E, Avenel V, Clavel-Chapelon F. Breast cancer risk in relation to different types of hormone replacement therapy in the E3N-EPIC cohort. *Int.J.Cancer* 2005;114:448-54.
45. Fournier A, Mesrine S, Boutron-Ruault MC, Clavel-Chapelon F. Estrogen-progestagen menopausal hormone therapy and breast cancer: does delay from menopause onset to treatment initiation influence risks? *J.Clin.Oncol.* 2009;27:5138-43.
46. Fournier A, Mesrine S, Dossus L, Boutron-Ruault MC, Clavel-Chapelon F, Chabbert-Buffet N. Risk of breast cancer after stopping menopausal hormone therapy in the E3N cohort. *Breast Cancer Res.Treat.* 2014;145:535-43.
47. World Cancer Research Fund / American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Washington, DC* 2007.
48. Fagherazzi G, Vilier A, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. Alcohol consumption and breast cancer risk subtypes in the E3N-EPIC cohort. *Eur.J.Cancer Prev.* 2014.
49. Hamajima N, Hirose K, Tajima K et al. 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;87:1234-45.
50. Cottet V, Touvier M, Fournier A et al. Postmenopausal breast cancer risk and dietary patterns in the E3N-EPIC prospective cohort study. *Am.J.Epidemiol.* 2009;170:1257-67.

51. World Cancer Research Fund / American Institute for Cancer Research. Continuous update project report. Food, nutrition and physical activity and the prevention of breast cancer. *Washington, DC* 2010.
52. Tehard B, Friedenreich CM, Oppert JM, Clavel-Chapelon F. Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. *Cancer Epidemiol.Biomarkers Prev.* 2006;15:57-64.
53. Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. *Br.J.Sports Med.* 2008;42:636-47.
54. Fournier A, Dos Santos G., Guillas G et al. Recent recreational physical activity and breast cancer risk in postmenopausal women in the E3N cohort. *Cancer Epidemiol.Biomarkers Prev.* 2014.
55. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. *J.Natl.Cancer Inst.* 2013;105:515-25.
56. DeRoo LA, Cummings P, Mueller BA. Smoking before the first pregnancy and the risk of breast cancer: a meta-analysis. *Am.J.Epidemiol.* 2011;174:390-402.
57. Dossus L, Boutron-Ruault MC, Kaaks R et al. Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. *Int.J.Cancer* 2014;134:1871-88.
58. Yang Y, Zhang F, Skrip L, Wang Y, Liu S. Lack of an association between passive smoking and incidence of female breast cancer in non-smokers: evidence from 10 prospective cohort studies. *PLoS.One.* 2013;8:e77029.
59. Xu X, Dailey AB, Peoples-Sheps M, Talbott EO, Li N, Roth J. Birth weight as a risk factor for breast cancer: a meta-analysis of 18 epidemiological studies. *J.Womens Health (Larchmt.)* 2009;18:1169-78.
60. Fagherazzi G, Guillas G, Boutron-Ruault MC, Clavel-Chapelon F, Mesrine S. Body shape throughout life and the risk for breast cancer at adulthood in the French E3N cohort. *Eur.J.Cancer Prev.* 2013;22:29-37.
61. Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 2008;371:569-78.
62. Fagherazzi G, Chabbert-Buffet N, Fabre A et al. Hip circumference is associated with the risk of premenopausal ER-/PR- breast cancer. *Int.J.Obes.(Lond)* 2012;36:431-9.
63. Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *Lancet Oncol.* 2011;12:785-94.
64. Organisation mondiale de la Santé (OMS). Obésité : prévention et prise en charge de l'épidémie mondiale. *Genève, Suisse: OMS, Série de Rapports techniques 894* 2003.
65. Antoni S, Sasco AJ, Dos Santos Silva I, McCormack V. Is mammographic density differentially associated with breast cancer according to receptor status? A meta-analysis. *Breast Cancer Res.Treat.* 2013;137:337-47.
66. McCormack VA, Dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol.Biomarkers Prev.* 2006;15:1159-69.

67. Boyd NF, Rommens JM, Vogt K et al. Mammographic breast density as an intermediate phenotype for breast cancer. *Lancet Oncol.* 2005;6:798-808.
68. Dupont WD, Page DL. Breast cancer risk associated with proliferative disease, age at first birth, and a family history of breast cancer. *Am.J.Epidemiol.* 1987;125:769-79.
69. Zhou WB, Xue DQ, Liu XA, Ding Q, Wang S. The influence of family history and histological stratification on breast cancer risk in women with benign breast disease: a meta-analysis. *J Cancer Res.Clin.Oncol.* 2011;137:1053-60.
70. Hartmann LC, Sellers TA, Frost MH et al. Benign breast disease and the risk of breast cancer. *N.Engl.J.Med.* 2005;353:229-37.
71. Goldacre MJ, Abisgold JD, Yeates DG, Vessey MP. Benign breast disease and subsequent breast cancer: English record linkage studies. *J.Public Health (Oxf)* 2010;32:565-71.
72. Xue F, Michels KB. Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence. *Am.J.Clin.Nutr.* 2007;86:s823-s835.
73. De Bruijn KM, Arends LR, Hansen BE, Leeftang S, Ruiter R, van Eijck CH. Systematic review and meta-analysis of the association between diabetes mellitus and incidence and mortality in breast and colorectal cancer. *Br.J.Surg.* 2013;100:1421-9.
74. Larsson SC, Mantzoros CS, Wolk A. Diabetes mellitus and risk of breast cancer: a meta-analysis. *Int J Cancer* 2007;121:856-62.
75. Boyle P, Boniol M, Koechlin A et al. Diabetes and breast cancer risk: a meta-analysis. *Br.J Cancer* 2012;107:1608-17.
76. Binefa G, Rodriguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: From prevention to personalized medicine. *World J.Gastroenterol.* 2014;20:6786-808.
77. Bishehsari F, Mahdavinia M, Vacca M, Malekzadeh R, Mariani-Costantini R. Epidemiological transition of colorectal cancer in developing countries: Environmental factors, molecular pathways, and opportunities for prevention. *World J.Gastroenterol.* 2014;20:6055-72.
78. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 2014;383:1490-502.
79. Ridge CA, McErlean AM, Ginsberg MS. Epidemiology of lung cancer. *Semin.Intervent.Radiol.* 2013;30:93-8.
80. Fallah M, Pukkala E, Tryggvadottir L et al. Risk of thyroid cancer in first-degree relatives of patients with non-medullary thyroid cancer by histology type and age at diagnosis: a joint study from five Nordic countries. *J Med.Genet.* 2013;50:373-82.
81. Dal Maso L., Bosetti C, La Vecchia C., Franceschi S. Risk factors for thyroid cancer: an epidemiological review focused on nutritional factors. *Cancer Causes Control* 2009;20:75-86.
82. Fisher SG, Fisher RI. The epidemiology of non-Hodgkin's lymphoma. *Oncogene* 2004;23:6524-34.
83. Gordon R. Skin cancer: an overview of epidemiology and risk factors. *Semin.Oncol.Nurs.* 2013;29:160-9.
84. Karagas MR, Zens MS, Nelson HH et al. Measures of cumulative exposure from a standardized sun exposure history questionnaire: a comparison with histologic assessment of solar skin damage. *Am.J Epidemiol.* 2007;165:719-26.

RÉFÉRENCES

85. Kvaskoff M, Bijon A, Mesrine S, Boutron-Ruault MC, Clavel-Chapelon F. Cutaneous melanoma and endogenous hormonal factors: a large French prospective study. *Am.J Epidemiol.* 2011;173:1192-202.
86. Barry JA, Azizia MM, Hardiman PJ. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum.Reprod.Update.* 2014;20:748-58.
87. Cramer DW. The epidemiology of endometrial and ovarian cancer. *Hematol.Oncol.Clin.North Am.* 2012;26:1-12.
88. Danforth KN, Schairer C, Schatzkin A, Lacey JV. Bone fractures and incident epithelial ovarian cancer in a prospective cohort study. *J Womens Health (Larchmt.)* 2009;18:1777-82.
89. Walboomers JM, Jacobs MV, Manos MM et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol.* 1999;189:12-9.
90. Crosbie EJ, Brabin L. Cervical cancer: problem solved? Vaccinating girls against human papillomavirus. *BJOG.* 2010;117:137-42.
91. Yeo TP, Lowenfels AB. Demographics and epidemiology of pancreatic cancer. *Cancer J.* 2012;18:477-84.
92. Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology* 2013;144:1252-61.
93. Centre d'épidémiologie sur les causes médicales de décès. *Inserm* 2014; www.cepdc.inserm.fr/site4/, consulté le 24/07/2014.
94. How can cancer kill you? *Cancer research UK* 2014; <http://www.cancerresearchuk.org>, consulté le 24/07/2014.
95. Ringa V, Fournier A. [Did the decrease in use of menopausal hormone therapy induce a decrease in the incidence of breast cancer in France (and elsewhere)?]. *Rev.Epidemiol.Sante Publique* 2008;56:297-301.
96. Tilloy E, Cottel D, Ruidavets JB et al. Characteristics of current smokers, former smokers, and second-hand exposure and evolution between 1985 and 2007. *Eur.J.Cardiovasc.Prev.Rehabil.* 2010;17:730-6.
97. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. (Reprinted from *Br Med J* 1954;ii:1451-5). *BMJ* 2004;328:1529-33.
98. Gandini S, Botteri E, Iodice S et al. Tobacco smoking and cancer: a meta-analysis. *Int.J.Cancer* 2008;122:155-64.
99. Hecht SS. Cigarette smoking: cancer risks, carcinogens, and mechanisms. *Langenbecks Arch.Surg.* 2006;391:603-13.
100. Ziolkiewicz A, Ellison PT, Lipson SF, Thune I, Jasienska G. Body fat, energy balance and estradiol levels: a study based on hormonal profiles from complete menstrual cycles. *Hum.Reprod.* 2008;23:2555-63.
101. Kaidar-Person O, Bar-Sela G, Person B. The two major epidemics of the twenty-first century: obesity and cancer. *Obes.Surg.* 2011;21:1792-7.

102. Freeman EW, Sammel MD, Lin H, Gracia CR. Obesity and reproductive hormone levels in the transition to menopause. *Menopause*. 2010;17:718-26.
103. Key TJ. Fruit and vegetables and cancer risk. *Br.J.Cancer* 2011;104:6-11.
104. Poortinga W. The prevalence and clustering of four major lifestyle risk factors in an English adult population. *Prev.Med* 2007;44:124-8.
105. Wallstrom P, Wirfalt E, Janzon L et al. Fruit and vegetable consumption in relation to risk factors for cancer: a report from the Malmo Diet and Cancer Study. *Public Health Nutr*. 2000;3:263-71.
106. Friedman GD, Tekawa I, Klatsky AL, Sidney S, Armstrong MA. Alcohol drinking and cigarette smoking: an exploration of the association in middle-aged men and women. *Drug Alcohol Depend*. 1991;27:283-90.
107. Piazza-Gardner AK, Barry AE. Examining physical activity levels and alcohol consumption: are people who drink more active? *Am.J.Health Promot*. 2012;26:e95-104.
108. Kaczynski AT, Manske SR, Mannell RC, Grewal K. Smoking and physical activity: a systematic review. *Am.J.Health Behav*. 2008;32:93-110.
109. Van Rensburg KJ, Taylor A, Hodgson T. The effects of acute exercise on attentional bias towards smoking-related stimuli during temporary abstinence from smoking. *Addiction* 2009;104:1910-7.
110. Swift DL, Johannsen NM, Lavie CJ, Earnest CP, Church TS. The role of exercise and physical activity in weight loss and maintenance. *Prog.Cardiovasc.Dis*. 2014;56:441-7.
111. Audrain-McGovern J, Benowitz NL. Cigarette smoking, nicotine, and body weight. *Clin.Pharmacol.Ther*. 2011;90:164-8.
112. Palaniappan U, Jacobs SL, O'Loughlin J, Gray-Donald K. Fruit and vegetable consumption is lower and saturated fat intake is higher among Canadians reporting smoking. *J.Nutr*. 2001;131:1952-8.
113. Baer WD, Nietert PJ. Patterns of fruit, vegetable, and milk consumption among smoking and nonsmoking female teens. *Am.J.Prev.Med*. 2002;22:240-6.
114. Poisson T, Dallongeville J, Evans A et al. Fruit and vegetable intake and smoking cessation. *Eur.J.Clin.Nutr*. 2012;66:1247-53.
115. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. *BMJ* 2008;337:a1440.
116. Jiao L, Mitrou PN, Reedy J et al. A combined healthy lifestyle score and risk of pancreatic cancer in a large cohort study. *Arch.Intern.Med*. 2009;169:764-70.
117. Cerhan JR, Potter JD, Gilmore JM et al. Adherence to the AICR cancer prevention recommendations and subsequent morbidity and mortality in the Iowa Women's Health Study cohort. *Cancer Epidemiol.Biomarkers Prev*. 2004;13:1114-20.
118. Sasazuki S, Inoue M, Iwasaki M et al. Combined impact of five lifestyle factors and subsequent risk of cancer: the Japan Public Health Center Study. *Prev.Med*. 2012;54:112-6.
119. Romaguera D, Vergnaud AC, Peeters PH et al. Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? Results from the EPIC study. *Am.J.Clin.Nutr*. 2012;96:150-63.

RÉFÉRENCES

120. Kirkegaard H, Johnsen NF, Christensen J, Frederiksen K, Overvad K, Tjønneland A. Association of adherence to lifestyle recommendations and risk of colorectal cancer: a prospective Danish cohort study. *BMJ* 2010;341:c5504.
121. Meng L, Maskarinec G, Lee J, Kolonel LN. Lifestyle factors and chronic diseases: application of a composite risk index. *Prev.Med.* 1999;29:296-304.
122. Ford ES, Bergmann MM, Kroger J, Schienkiewitz A, Weikert C, Boeing H. Healthy living is the best revenge: findings from the European Prospective Investigation Into Cancer and Nutrition-Potsdam study. *Arch.Intern.Med.* 2009;169:1355-62.
123. Harnack L, Nicodemus K, Jacobs DR, Jr., Folsom AR. An evaluation of the Dietary Guidelines for Americans in relation to cancer occurrence. *Am.J.Clin.Nutr.* 2002;76:889-96.
124. Odegaard AO, Koh WP, Yuan JM. Combined lifestyle factors and risk of incident colorectal cancer in a Chinese population. *Cancer Prev.Res.(Phila)* 2013;6:360-7.
125. Hastert TA, Beresford SA, Patterson RE, Kristal AR, White E. Adherence to WCRF/AICR cancer prevention recommendations and risk of postmenopausal breast cancer. *Cancer Epidemiol.Biomarkers Prev.* 2013;22:1498-508.
126. McCullough ML, Patel AV, Kushi LH et al. Following cancer prevention guidelines reduces risk of cancer, cardiovascular disease, and all-cause mortality. *Cancer Epidemiol.Biomarkers Prev.* 2011;20:1089-97.
127. Khaw KT, Wareham N, Bingham S, Welch A, Luben R, Day N. Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. *PLoS.Med* 2008;5:e12.
128. Kvaavik E, Batty GD, Ursin G, Huxley R, Gale CR. Influence of individual and combined health behaviors on total and cause-specific mortality in men and women: the United Kingdom health and lifestyle survey. *Arch.Intern Med* 2010;170:711-8.
129. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Proportion of breast cancer cases in the United States explained by well-established risk factors. *J.Natl.Cancer Inst.* 1995;87:1681-5.
130. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am.J.Epidemiol.* 1985;122:904-14.
131. Sprague BL, Trentham-Dietz A, Egan KM, Titus-Ernstoff L, Hampton JM, Newcomb PA. Proportion of invasive breast cancer attributable to risk factors modifiable after menopause. *Am.J.Epidemiol.* 2008;168:404-11.
132. Hines LM, Risendal B, Slattery ML et al. Comparative analysis of breast cancer risk factors among Hispanic and non-Hispanic white women. *Cancer* 2010;116:3215-23.
133. Barnes BB, Steindorf K, Hein R, Flesch-Janys D, Chang-Claude J. Population attributable risk of invasive postmenopausal breast cancer and breast cancer subtypes for modifiable and non-modifiable risk factors. *Cancer Epidemiol.* 2011;35:345-52.
134. Ghiasvand R, Bahmanyar S, Zendehdel K et al. Postmenopausal breast cancer in Iran; risk factors and their population attributable fractions. *BMC.Cancer* 2012;12:414.
135. Mezzetti M, La Vecchia C, Decarli A, Boyle P, Talamini R, Franceschi S. Population attributable risk for breast cancer: diet, nutrition, and physical exercise. *J.Natl.Cancer Inst.* 1998;90:389-94.

136. Tavani A, Braga C, La Vecchia C, Negri E, Russo A, Franceschi S. Attributable risks for breast cancer in Italy: education, family history and reproductive and hormonal factors. *Int.J.Cancer* 1997;70:159-63.
137. Granstrom C, Sundquist J, Hemminki K. Population attributable risks for breast cancer in Swedish women by morphological type. *Breast Cancer Res.Treat.* 2008;111:559-68.
138. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. *Br.J.Cancer* 2011;105 Suppl 2:S77-S81.
139. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26:725-31.
140. Richy F, Deceulaer F, Ethgen O, Bruyere O, Reginster JY. Development and validation of the ORACLE score to predict risk of osteoporosis. *Mayo Clin.Proc.* 2004;79:1402-8.
141. Kanis JA, McCloskey EV, Johansson H, Oden A, Strom O, Borgstrom F. Development and use of FRAX in osteoporosis. *Osteoporos.Int* 2010;21 Suppl 2:S407-S413.
142. Cho E, Rosner BA, Feskanich D, Colditz GA. Risk factors and individual probabilities of melanoma for whites. *J.Clin.Oncol.* 2005;23:2669-75.
143. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol.* 2006;5:735-41.
144. Driver JA, Gaziano JM, Gelber RP, Lee IM, Buring JE, Kurth T. Development of a risk score for colorectal cancer in men. *Am.J.Med.* 2007;120:257-63.
145. Spitz MR, Hong WK, Amos CI et al. A risk model for prediction of lung cancer. *J.Natl.Cancer Inst.* 2007;99:715-26.
146. Gail MH, Brinton LA, Byar DP et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J.Natl.Cancer Inst.* 1989;81:1879-86.
147. Rosner B, Colditz GA. Nurses' health study: log-incidence mathematical model of breast cancer incidence. *J.Natl.Cancer Inst.* 1996;88:359-64.
148. Rosner B, Colditz GA, Iglehart JD, Hankinson SE. Risk prediction models with incomplete data with application to prediction of estrogen receptor-positive breast cancer: prospective data from the Nurses' Health Study. *Breast Cancer Res.* 2008;10:R55.
149. Ueda K, Tsukuma H, Tanaka H, Ajiki W, Oshima A. Estimation of individualized probabilities of developing breast cancer for Japanese women. *Breast Cancer* 2003;10:54-62.
150. 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;152:950-64.
151. Lee EO, Ahn SH, You C et al. Determining the main risk factors and high-risk groups of breast cancer using a predictive model for breast cancer risk assessment in South Korea. *Cancer Nurs.* 2004;27:400-6.
152. Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat.Med.* 2004;23:1111-30.
153. Barlow WE, White E, Ballard-Barbash R et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J.Natl.Cancer Inst.* 2006;98:1204-14.

154. Chlebowski RT, Anderson GL, Lane DS et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J.Natl.Cancer Inst.* 2007;99:1695-705.
155. Tamimi RM, Rosner B, Colditz GA. Evaluation of a breast cancer risk prediction model expanded to include category of prior benign breast disease lesion. *Cancer* 2010;116:4944-53.
156. Pfeiffer RM, Park Y, Kreimer AR et al. Risk prediction for breast, endometrial, and ovarian cancer in white women aged 50 y or older: derivation and validation from population-based cohort studies. *PLoS.Med.* 2013;10:e1001492.
157. Gail MH, Costantino JP, Pee D et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J.Natl.Cancer Inst.* 2007;99:1782-92.
158. Decarli A, Calza S, Masala G, Specchia C, Palli D, Gail MH. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence-European Prospective Investigation Into Cancer and Nutrition cohort. *J.Natl.Cancer Inst.* 2006;98:1686-93.
159. Matsuno RK, Costantino JP, Ziegler RG et al. Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. *J.Natl.Cancer Inst.* 2011;103:951-61.
160. Pastor-Barriuso R, Ascunce N, Ederra M et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. *Breast Cancer Res.Treat.* 2013;138:249-59.
161. Gao F, Machin D, Chow KY et al. Assessing risk of breast cancer in an ethnically South-East Asia population (results of a multiple ethnic groups study). *BMC.Cancer* 2012;12:529.
162. Chen J, Pee D, Ayyagari R et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. *J.Natl.Cancer Inst.* 2006;98:1215-26.
163. Boyle P, Mezzetti M, La Vecchia C., Franceschi S, Decarli A, Robertson C. Contribution of three components to individual cancer risk predicting breast cancer risk in Italy. *Eur.J.Cancer Prev.* 2004;13:183-91.
164. Novotny J, Pecen L, Petruzalka L et al. Breast cancer risk assessment in the Czech female population--an adjustment of the original Gail model. *Breast Cancer Res.Treat.* 2006;95:29-35.
165. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res.Treat.* 2005;94:115-22.
166. Tice JA, Miike R, Adduci K, Petrakis NL, King E, Wrensch MR. Nipple aspirate fluid cytology and the Gail model for breast cancer risk assessment in a screening population. *Cancer Epidemiol.Biomarkers Prev.* 2005;14:324-8.
167. WHO Report on Global Tobacco Epidemic, 2011: Warning about the dangers of tobacco. Geneva, Switzerland: World Health Organization 2011.
168. Licaj L, Romieu I, Clavel-Chapelon F, Boutron-Ruault MC, Ferrari P. Impact du tabac sur la mortalité totale et sur la mortalité par cause dans l'étude européenne EPIC (European Prospective Investigation into Cancer and nutrition). *Bulletin épidémiologique hebdomadaire (BEH)* 2013;234-8.
169. Oberg M, Jaakkola MS, Woodward A, Peruga A, Pruss-Ustun A. Worldwide burden of disease from exposure to second-hand smoke: a retrospective analysis of data from 192 countries. *Lancet* 2011;377:139-46.
170. Global health risks: mortality and burden of disease attributable to selected major risks. Geneva, Switzerland: World Health Organization 2009.

171. Whitlock G, Lewington S, Sherliker P et al. Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* 2009;373:1083-96.
172. World health statistics. *Geneva, Switzerland: World Health Organization* 2014.
173. Enquête épidémiologique nationale sur le surpoids et l'obésité. *Inserm, Kantar Health, Roche.ObEpi* 2012.
174. Ministère du Travail de l'Emploi et de la Santé. Programme national nutrition santé 2011-2015. *PNNS* 2011.
175. Pearl R. Alcohol and longevity. *New York, Knopf* 1926.
176. Reid MC, Boutros NN, O'Connor PG, Cadariu A, Concato J. The health-related effects of alcohol use in older persons: a systematic review. *Subst.Abus.* 2002;23:149-64.
177. Global status report on alcohol and health. *Geneva, Switzerland: World Health Organization* 2014.
178. Rapport sur la santé dans le monde 2002 - Réduire les risques et promouvoir une vie saine. *Geneva, Switzerland: World Health Organization* 2002.
179. Elmadfa I, Meyer A, owak V et al. European Nutrition and Health Report. *Forum Nutrition* 2009;62:1-405.
180. Situation nutritionnelle en France en 2006 selon les indicateurs d'objectif et les repères du Programme national nutrition santé (PNNS). *Etude nationale nutrition santé ENNS* 2006.
181. Vuillemin A, Escalon H, Bossard C. Activité physique et sédentarité. *Baromètre santé 2008.Saint-Denis : INPES* 2008;239-68.
182. Woodcock J, Franco OH, Orsini N, Roberts I. Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. *Int J Epidemiol.* 2011;40:121-38.
183. Samitz G, Egger M, Zwahlen M. Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies. *Int J Epidemiol.* 2011;40:1382-400.
184. Loeff M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Prev.Med.* 2012;55:163-70.
185. Tsubono Y, Koizumi Y, Nakaya N et al. Health practices and mortality in Japan: combined effects of smoking, drinking, walking and body mass index in the Miyagi Cohort Study. *J.Epidemiol.* 2004;14 Suppl 1:S39-S45.
186. Yun JE, Won S, Kimm H, Jee SH. Effects of a combined lifestyle score on 10-year mortality in Korean men and women: a prospective cohort study. *BMC.Public Health* 2012;12:673.
187. Haveman-Nies A, de Groot L, Burema J, Cruz JA, Osler M, van Staveren WA. Dietary quality and lifestyle factors in relation to 10-year mortality in older Europeans: the SENECA study. *Am.J.Epidemiol.* 2002;156:962-8.
188. van den Brandt PA. The impact of a Mediterranean diet and healthy lifestyle on premature mortality in men and women. *Am.J.Clin.Nutr.* 2011;94:913-20.
189. Behrens G, Fischer B, Kohler S, Park Y, Hollenbeck AR, Leitzmann MF. Healthy lifestyle behaviors and decreased risk of mortality in a large prospective study of U.S. women and men. *Eur.J.Epidemiol.* 2013;28:361-72.

RÉFÉRENCES

190. Knuops KT, de Groot LC, Kromhout D et al. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. *JAMA* 2004;292:1433-9.
191. Ford ES, Zhao G, Tsai J, Li C. Low-risk lifestyle behaviors and all-cause mortality: findings from the National Health and Nutrition Examination Survey III Mortality Study. *Am.J.Public Health* 2011;101:1922-9.
192. Hamer M, Bates CJ, Mishra GD. Multiple health behaviors and mortality risk in older adults. *J.Am.Geriatr.Soc.* 2011;59:370-2.
193. Gopinath B, Flood VM, Burlutsky G, Mitchell P. Combined influence of health behaviors on total and cause-specific mortality. *Arch.Intern.Med.* 2010;170:1605-7.
194. Stamler J, Stamler R, Neaton JD et al. Low risk-factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy: findings for 5 large cohorts of young adult and middle-aged men and women. *JAMA* 1999;282:2012-8.
195. Nechuta SJ, Shu XO, Li HL et al. Combined impact of lifestyle-related factors on total and cause-specific mortality among Chinese women: prospective cohort study. *PLoS.Med.* 2010;7.
196. Iversen L, Hannaford PC, Lee AJ, Elliott AM, Fielding S. Impact of lifestyle in middle-aged women on mortality: evidence from the Royal College of General Practitioners' Oral Contraception Study. *Br.J.Gen.Pract.* 2010;60:563-9.
197. Tamakoshi A, Tamakoshi K, Lin Y, Yagyu K, Kikuchi S. Healthy lifestyle and preventable death: findings from the Japan Collaborative Cohort (JACC) Study. *Prev.Med.* 2009;48:486-92.
198. Byun W, Sieverdes JC, Sui X et al. Effect of positive health factors and all-cause mortality in men. *Med.Sci.Sports Exerv.* 2010;42:1632-8.
199. Eguchi E, Iso H, Tanabe N et al. Healthy lifestyle behaviours and cardiovascular mortality among Japanese men and women: the Japan collaborative cohort study. *Eur.Heart J.* 2012;33:467-77.
200. Clavel-Chapelon F, Jadand C, Goulard H, Guibout-Peigne C. [E3N, a cohort study on cancer risk factors in MGEN women. Description of protocol, main characteristics and population]. *Bull.Cancer* 1996;83:1008-13.
201. Clavel-Chapelon F, van Liere MJ, Giubout C et al. E3N, a French cohort study on cancer risk factors. E3N Group. Étude épidémiologique auprès de femmes de l'Éducation nationale. *Eur.J.Cancer Prev.* 1997;6:473-8.
202. van Liere MJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. *Int.J.Epidemiol.* 1997;26 Suppl 1:S128-S136.
203. Lucas F, Niravong M, Kaaks R, Villeminot S, Clavel-Chapelon F. Estimation of food portion size using photographs: Relative validity, strenghts, weaknesses and recommendations. *J.Hum.Nutr.Diet.* 1995;8:65-74.
204. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann.Oncol.* 1992;3:783-91.
205. 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.
206. Riboli E, Hunt KJ, Slimani N et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* 2002;5:1113-24.

207. Gonzalez CA. The European Prospective Investigation into Cancer and Nutrition (EPIC). *Public Health Nutr.* 2006;9:124-6.
208. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 10th Revision. *WHO Library Cataloguing-in-Publication Data* 2010.
209. Percy C, Fritz A, Jack A et al. International Classification of Diseases for Oncology (ICD-O). Third Edition. *World Health Organization, Geneva* 2000.
210. Sobin LH, Gospodarowicz MK, Wittekind C. TNM Classification of Malignant Tumours, 7th Edition. *UICC International Union Against Cancer* 2009.
211. Inserm, CépiDc. Comparability and quality improvement of European causes of death statistics. *European Commission DG Santo Agreement* 2001; N°EDC DGV/F3 SOC 98 20108.
212. Eurostat, European Commission.; <http://epp.eurostat.ec.europa.eu/>, consulté le 18/08/2014.
213. Fouillet A, Pavillon G, Vicente P et al. La certification électronique des décès, France, 2007-2011. *BEH 1 / 10 janvier* 2012.
214. Favier J-C, Ireland-Ripert J, INRA (Paris F et al. Répertoire général des aliments : table de composition. *Paris, France : Tec & Doc - Lavoisier* 1995.
215. Verdebout J. A European satellite-derived UV climatology available for impact studies. *Radiat.Prot.Dosimetry.* 2004;111:407-11.
216. Sorensen TI, Stunkard AJ, Teasdale TW, Higgins MW. The accuracy of reports of weight: children's recall of their parents' weights 15 years earlier. *Int.J.Obes.* 1983;7:115-22.
217. Tehard B, van Liere MJ, Com NC, Clavel-Chapelon F. Anthropometric measurements and body silhouette of women: validity and perception. *J.Am.Diet.Assoc.* 2002;102:1779-84.
218. Ainsworth BE, Haskell WL, Herrmann SD et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med.Sci.Sports Exerc.* 2011;43:1575-81.
219. Fournier A, Fabre A, Mesrine S, Boutron-Ruault MC, Berrino F, Clavel-Chapelon F. Use of different postmenopausal hormone therapies and risk of histology- and hormone receptor-defined invasive breast cancer. *J.Clin.Oncol.* 2008;26:1260-8.
220. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)* 1972;34:184-220.
221. Collet D. Modelling survival data in medical research - Second Edition. *Chapman & Hall Ed.* 2003.
222. Breslow NE. Discussion of the paper by D. R. Cox. *J R Statist Soc B* 1972;34:216-7.
223. de Boor. A practical guide to Splines. *Springer, New York* 1978.
224. Harrel FEJr. Regression modeling strategies: with applications to linear models, logistic regression and survival analysis. *New York: Springer-Verlag* 2001.
225. Commenges D, Letenneur L, Joly P. Re: "Serum transferrin saturation, stroke incidence, and mortality in women and men. The NHANES I Epidemiologic Followup Study". *Am.J Epidemiol.* 1997;146:683-4.

RÉFÉRENCES

226. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am.J Epidemiol.* 1997;145:72-80.
227. Thiebaut AC, Benichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Stat.Med.* 2004;23:3803-20.
228. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. *New York: John Wiley & Sons.* 1980.
229. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995;51:524-32.
230. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 1967;13:21-7.
231. Gauthier E, Brisson L, Lenca P, Ragusa S. Breast cancer risk score: a data-mining approach to improve readability. *The International Conference on Data Mining.CSREA Press* 2011;15-21.
232. Gauthier E, Brisson L, Lenca P, Clavel-Chapelon F, Ragusa S. Challenges to building a platform for a breast cancer risk score. *Sixth International Conference on Research Challenges in Information Science.IEEE* 2012;1-10.
233. Levin ML. The occurrence of lung cancer in man. *Acta Unio.Int.Contra.Cancrum.* 1953;9:531-41.
234. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am.J.Public Health* 1998;88:15-9.
235. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. *Am.J Epidemiol.* 1974;99:325-32.
236. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control* 2007;18:571-9.
237. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int.J Epidemiol.* 2004;33:1389-97.
238. Agence Nationale de Sécurité Sanitaire Alimentation Environnement Travail. Nutrition et cancer. *ANSES* 2011.
239. World Health Organization. National cancer control programmes. *Policies and managerial guidelines.2nd Edition* 2002.
240. Agence Nationale de Sécurité Sanitaire Alimentation Environnement Travail. Nutrition et cancer. *ANSES* 2011.
241. Levin ML. The occurrence of lung cancer in an. *Acta Unio Int Contra Cancrum* 1953;9:531-41.
242. Agate L, Lorusso L, Elisei R. New and old knowledge on differentiated thyroid cancer epidemiology and risk factors. *J.Endocrinol.Invest* 2012;35:3-9.
243. Zhao ZG, Guo XG, Ba CX et al. Overweight, obesity and thyroid cancer risk: a meta-analysis of cohort studies. *J Int Med.Res.* 2012;40:2041-50.
244. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC.Cancer* 2012;12:385.

245. Lin JS, Eder M, Weinmann S et al. Behavioral counseling to prevent skin cancer: Systematic evidence review to update the 2003 U.S. preventive services task force recommendation. *Ann of Intern Med.* 2011;154:190-201.
246. Bataille V, de VE. Melanoma - Part 1: epidemiology, risk factors, and prevention. *BMJ* 2008;337:a2249.
247. Tardon A, Lee WJ, Delgado-Rodriguez M et al. Leisure-time physical activity and lung cancer: a meta-analysis. *Cancer Causes Control* 2005;16:389-97.
248. Buffart LM, Singh AS, van Loon EC, Vermeulen HI, Brug J, Chinapaw MJ. Physical activity and the risk of developing lung cancer among smokers: A meta-analysis. *J.Sci.Med Sport* 2013; [Epub ahead of print].
249. Seitz HK, Pelucchi C, Bagnardi V, La Vecchia C. Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012. *Alcohol Alcohol* 2012;47:204-12.
250. Bagnardi V, Rota M, Botteri E et al. Light alcohol drinking and cancer: a meta-analysis. *Ann Oncol.* 2013;24:301-8.
251. Amadou A, Ferrari P, Muwonge R et al. Overweight, obesity and risk of premenopausal breast cancer according to ethnicity: a systematic review and dose-response meta-analysis. *Obes.Rev.* 2013;14:665-78.
252. Crosbie EJ, Zwahlen M, Kitchener HC, Egger M, Renehan AG. Body mass index, hormone replacement therapy, and endometrial cancer risk: a meta-analysis. *Cancer Epidemiol.Biomarkers Prev.* 2010;19:3119-30.
253. Larsson SC, Wolk A. Body mass index and risk of non-Hodgkin's and Hodgkin's lymphoma: a meta-analysis of prospective studies. *Eur.J.Cancer* 2011;47:2422-30.
254. Larsson SC, Wolk A. Overweight and obesity and incidence of leukemia: a meta-analysis of cohort studies. *Int.J.Cancer* 2008;122:1418-21.
255. Wallin A, Larsson SC. Body mass index and risk of multiple myeloma: a meta-analysis of prospective studies. *Eur.J.Cancer* 2011;47:1606-15.
256. Merhi M, Raynal H, Cahuzac E, Vinson F, Cravedi JP, Gamet-Payrastre L. Occupational exposure to pesticides and risk of hematopoietic cancers: meta-analysis of case-control studies. *Cancer Causes Control* 2007;18:1209-26.
257. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *Lancet Oncol.* 2001;2:133-40.
258. Iwasaki M, Tsugane S. Risk factors for breast cancer: epidemiological evidence from Japanese studies. *Cancer Sci.* 2011;102:1607-14.
259. National breast and ovarian cancer centre. Breast cancer risk factors: a review of the evidence. *National breast and ovarian cancer centre, Surry Hills, NSW* 2009.
260. Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. *Am.J.Epidemiol.* 1998;147:826-33.
261. Ahlgren M, Melbye M, Wohlfahrt J, Sorensen TI. Growth patterns and the risk of breast cancer in women. *N.Engl.J.Med* 2004;351:1619-26.

RÉFÉRENCES

262. Baer HJ, Colditz GA, Rosner B et al. Body fatness during childhood and adolescence and incidence of breast cancer in premenopausal women: a prospective cohort study. *Breast Cancer Res.* 2005;7:R314-R325.
263. Weiderpass E, Braaten T, Magnusson C et al. A prospective study of body size in different periods of life and risk of premenopausal breast cancer. *Cancer Epidemiol.Biomarkers Prev.* 2004;13:1121-7.
264. Michels KB, Xue F. Role of birthweight in the etiology of breast cancer. *Int.J.Cancer* 2006;119:2007-25.
265. Coombs NJ, Taylor R, Wilcken N, Boyages J. HRT and breast cancer: impact on population risk and incidence. *Eur.J.Cancer* 2005;41:1775-81.
266. Clarke CA, Purdie DM, Glaser SL. Population attributable risk of breast cancer in white women associated with immediately modifiable risk factors. *BMC.Cancer* 2006;6:170.
267. Hayes J, Richardson A, Frampton C. Population attributable risks for modifiable lifestyle factors and breast cancer in New Zealand women. *Intern Med J.* 2013;43:1198-204.
268. Neutel CI, Morrison H. Could recent decreases in breast cancer incidence really be due to lower HRT use? Trends in attributable risk for modifiable breast cancer risk factors in Canadian women. *Can.J.Public Health* 2010;101:405-9.
269. La Vecchia C, Negri E, Franceschi S et al. Body mass index and post-menopausal breast cancer: an age-specific analysis. *Br.J.Cancer* 1997;75:441-4.
270. 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. *J.Natl.Cancer Inst.* 2011;103:250-63.
271. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis. *Int.J.Cancer* 2009;124:698-712.
272. Nichols HB, Trentham-Dietz A, Love RR et al. Differences in breast cancer risk factors by tumor marker subtypes among premenopausal Vietnamese and Chinese women. *Cancer Epidemiol.Biomarkers Prev.* 2005;14:41-7.
273. Anderson KN, Schwab RB, Martinez ME. Reproductive risk factors and breast cancer subtypes: a review of the literature. *Breast Cancer Res.Treat.* 2014;144:1-10.
274. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. *J.Natl.Cancer Inst.* 2004;96:218-28.
275. Setiawan VW, Monroe KR, Wilkens LR, Kolonel LN, Pike MC, Henderson BE. Breast cancer risk factors defined by estrogen and progesterone receptor status: the multiethnic cohort study. *Am.J.Epidemiol.* 2009;169:1251-9.
276. Islam T, Matsuo K, Ito H et al. Reproductive and hormonal risk factors for luminal, HER2-overexpressing, and triple-negative breast cancer in Japanese women. *Ann Oncol.* 2012;23:2435-41.
277. Li J, Humphreys K, Eriksson L, Czene K, Liu J, Hall P. Effects of childhood body size on breast cancer tumour characteristics. *Breast Cancer Res.* 2010;12:R23.
278. Canchola AJ, Anton-Culver H, Bernstein L et al. Body size and the risk of postmenopausal breast cancer subtypes in the California Teachers Study cohort. *Cancer Causes Control* 2012.

279. Rosner B, Glynn RJ, Tamimi RM et al. Breast cancer risk prediction with heterogeneous risk profiles according to breast cancer tumor markers. *Am.J.Epidemiol.* 2013;178:296-308.
280. Rosenberg LU, Einarsdottir K, Friman EI et al. Risk factors for hormone receptor-defined breast cancer in postmenopausal women. *Cancer Epidemiol.Biomarkers Prev.* 2006;15:2482-8.
281. Slanger TE, Chang-Claude JC, Obi N et al. Menopausal hormone therapy and risk of clinical breast cancer subtypes. *Cancer Epidemiol.Biomarkers Prev.* 2009;18:1188-96.
282. Must A, Phillips SM, Naumova EN et al. Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am.J.Epidemiol.* 2002;155:672-9.
283. Forouzanfar MH, Foreman KJ, Delossantos AM et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. *Lancet* 2011;378:1461-84.
284. Freedman AN, Seminara D, Gail MH et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J.Natl.Cancer Inst.* 2005;97:715-23.
285. Dartois L, Fagherazzi G, Boutron-Ruault MC, Mesrine S, Clavel-Chapelon F. Association between five lifestyle habits and cancer risk: results from the E3N cohort. *Cancer Prev.Res.(Phila)* 2014;7:516-25.
286. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann.Intern.Med.* 2008;148:337-47.
287. Petracci E, Decarli A, Schairer C et al. Risk factor modification and projections of absolute breast cancer risk. *J Natl.Cancer Inst.* 2011;103:1037-48.
288. Chen WY, Colditz GA. Risk factors and hormone-receptor status: epidemiology, risk-prediction models and treatment implications for breast cancer. *Nat.Clin.Pract.Oncol.* 2007;4:415-23.
289. Gail MH, Brinton LA, Byar DP et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J.Natl.Cancer Inst.* 1989;81:1879-86.
290. Gjorgov NA. Breast cancer risk assessments to barrier contraception exposure. A new approach. *Prilozi.* 2009;30:217-32.
291. Cook NR, Rosner BA, Hankinson SE, Colditz GA. Mammographic screening and risk factors for breast cancer. *Am.J.Epidemiol.* 2009;170:1422-32.
292. Wacholder S, Hartge P, Prentice R et al. Performance of common genetic variants in breast-cancer risk models. *N.Engl.J.Med.* 2010;362:986-93.
293. Goldstein LB, Adams R, Alberts MJ et al. Primary prevention of ischemic stroke: a guideline from the American Heart Association/American Stroke Association Stroke Council: cosponsored by the Atherosclerotic Peripheral Vascular Disease Interdisciplinary Working Group; Cardiovascular Nursing Council; Clinical Cardiology Council; Nutrition, Physical Activity, and Metabolism Council; and the Quality of Care and Outcomes Research Interdisciplinary Working Group. *Circulation* 2006;113:e873-e923.
294. World Health Organization. Prevention of cardiovascular disease. *Guidelines for assessment and management of cardiovascular risk* 2007.
295. National Institute for Health and Clinical Excellence. Preventing type 2 diabetes - risk identification and interventions for individuals at high risk (PH38). *Public health guidance* 38 212.

RÉFÉRENCES

296. Agence Nationale de Sécurité Sanitaire Alimentation Environnement Travail. Nutrition et cancer. *ANSES* 2011.
297. Beck F, Guignard R, Richard JB, Wilquin JL. Évolutions récentes du tabagisme en France. *Inpes, à paraître* 2010.
298. Inserm, Kantar Health, Roche. Enquête épidémiologique nationale sur le surpoids et l'obésité. Inserm / Kantar Health / Roche. *ObEpi* 2012.
299. Beck F, Richard JB. Épidémiologie de l'alcoolisation. *EMC - Endocrinologie Nutrition* 2012;10:1-9.
300. Blot C. Fruits et légumes : Évolutions et tendances de consommation. *FranceAgriMer, établissement national des produits de l'agriculture et de la mer* 2012.
301. Inserm. Activité physique. Contextes et effets sur la santé. Expertise collective. *Les éditions Inserm* 2008.