



HAL
open science

Cancer du sein, 25 000, 33 867, 41 845 nouveaux cas par an : d'où viennent ces chiffres ?

J. Esteve

► To cite this version:

J. Esteve. Cancer du sein, 25 000, 33 867, 41 845 nouveaux cas par an : d'où viennent ces chiffres ?. 27^e Journées de la Société française de sénologie et de pathologie mammaire (SFSPM), Deauville, 2005. Dogmes et doutes, Nov 2005, Deauville, France. pp.114-121. hal-03574925

HAL Id: hal-03574925

<https://hal.science/hal-03574925v1>

Submitted on 15 Feb 2022

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

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



Cancer du sein, 25 000, 33 867, 41 845 nouveaux cas par an : d'où viennent ces chiffres ?

*Breast cancer, 25 000, 33 867, 41 845 new cases per year:
where do these figures come from?*

Mots clés : Cancer du sein, Registre du cancer, Épidémiologie, Modèles statistiques.
Keywords : *Breast cancer, Cancer Registry, Epidemiology, Statistical modeling.*

J. Estève⁽¹⁾

Le titre de cet article m'a été proposé par les organisateurs de ce congrès et les chiffres qui y figurent suggèrent qu'il existe une certaine incompréhension des données de la littérature épidémiologique. La formule elle-même, "nombre de nouveaux cas par an", permet de penser que la croissance de l'incidence du cancer du sein n'a pas été perçue dans toute son ampleur par les sénologues. Il devrait être clair que l'on ne peut parler de nombre de cas incidents en France que si l'on précise l'année à laquelle on se réfère. Chacun des chiffres du titre correspond en effet à une année particulière pour laquelle une estimation a été publiée [1-5]. Ces chiffres ont été validés dans la dernière de ces études, qui reprend l'ensemble des données disponibles depuis 1978. Les chiffres d'incidence et de mortalité pour la France entière sont rappelés dans le **tableau I**, et l'on peut y constater que les chiffres du titre correspondent assez parfaitement à la tendance de l'incidence entre 1985 et 2000.

Le titre du congrès nous invite à porter un œil critique sur les références utilisées en sénologie. Si l'article de L. Remontet et al. constitue, pour quelque temps, la référence sur l'incidence du cancer, il est légitime de discuter sa validité; sinon, l'article pourrait s'achever ici puisqu'on a répondu à la question qui m'avait été posée: "D'où viennent ces chiffres?" Nous allons donc essayer de répondre à une question un peu différente: "Comment ces chiffres ont-ils été obtenus?"

1. Université Claude-Bernard, Lyon.

TABLEAU I. Évolution de l'incidence et de la mortalité par cancer du sein en France entre 1980 et 2000.

Année	1980	1985	1990	1995	2000
Incidence					
Cas	21 211	24 908	29 617	35 471	41 845
Taux ¹	55,5	63,2	71,5	80,1	88,9
Mortalité					
Décès	8 629	9 299	10 059	10 892	11 637
Taux ¹	18,7	19,1	19,4	19,6	19,7

1. Taux standard pour la population mondiale (pour 100 000 personnes-années).

L'information utilisée et son origine

La contribution essentielle à l'estimation de l'incidence du cancer en France est fournie par l'association des registres de cancer (*FRANCIM*). À la fin des années 1970, en l'absence de politique nationale, les premiers registres de morbidité ont été mis en place sur la base d'initiatives individuelles dans trois domaines : le cancer (1975), les malformations congénitales (1979) et les cardiopathies ischémiques (1984). Le but d'un registre est d'obtenir des informations sur *tous les cas* survenant dans une population de référence. En France, c'est le plus souvent le département, base géographique pour laquelle les données démographiques et les données de mortalité sont disponibles, qui est choisi comme population de référence. Pour obtenir un enregistrement exhaustif, il est impératif de consulter plusieurs sources d'information : anatomopathologistes, biologistes, services des hôpitaux et des centres de lutte contre le cancer, spécialistes et chirurgiens du secteur privé, médecins des caisses d'assurance maladie et, depuis quelques années, les départements d'information médicale (PMSI en France). Enfin, les certificats de décès peuvent constituer une source complémentaire pour retrouver des cas qui auraient échappé à l'approche précédente.

Une politique nationale a été mise en place, en 1986, avec la création du Comité national des registres (CNR), placé sous la tutelle conjointe de la Direction générale de la Santé (DGS) et de l'Institut national de la santé et de la recherche médicale, mettant l'accent sur la double mission des registres : recherche épidémiologique et santé publique. Plus récemment, la mission confiée à la DGS a été transférée à l'Institut national de veille sanitaire, qui encadre et pilote la création de nouveaux registres et le renforcement des anciens. *FRANCIM* est une association créée par les premiers registres de cancer. Elle assure une action importante de coordination et de mise en commun des méthodes et des données. En particulier, une base de données communes a été créée et sa gestion a été confiée au service de biostatistique des Hospices civils de Lyon (HCL). Ce dernier conduit, par ailleurs, une recherche méthodologique dont bénéficie directement le

réseau. Des liens étroits existent aussi avec le CépiDc⁽¹⁾ de l'INSERM, chargé des statistiques nationales des causes médicales de décès en collaboration avec l'INSEE. Dans le cas du cancer du sein, les données disponibles sont celles de dix registres départementaux dont la date de création s'étend de 1975 à 1994. Ces données, disponibles jusqu'en 1996 ou 1997, sont à la base de l'estimation dont les résultats sont présentés dans le **tableau I**. Elles représentent environ 50 000 cas de cancer du sein pour environ 50 millions de personnes-années.

Les difficultés d'une estimation nationale

Si les dix registres départementaux utilisés avaient été créés à la même date et étaient représentatifs de la population française, le calcul du taux d'incidence par année d'âge dans la population des registres pour chaque année d'observation et son application aux personnes-années de la population française, calculées à partir des estimations de population de l'INSEE, auraient fourni les estimations cherchées. Ces deux conditions n'étant pas satisfaites, on est obligé d'avoir recours à une modélisation plus complexe. Il est tout d'abord nécessaire d'examiner la tendance par registre pour éviter que les registres observés récemment biaisent l'estimation de l'évolution des taux et du nombre de cas, en raison des différences de niveau d'incidence entre les anciens et les nouveaux registres. La **figure 1** montre la tendance du taux d'incidence standardisé pour la population mondiale pour chacun des dix registres. On observe, en effet, des niveaux d'incidence assez différents, mais une tendance très homogène. Cette figure est obtenue à partir d'un modèle statistique qui estime la progression, pour chaque registre, du taux d'incidence avec l'âge et l'augmentation du niveau moyen d'incidence avec chaque cohorte de naissance. Il permet de prédire avec une fiabilité raisonnable [6] le taux d'incidence pour chaque année d'âge et pour chacun des dix départements pour l'année 2000. On en déduit facilement le taux

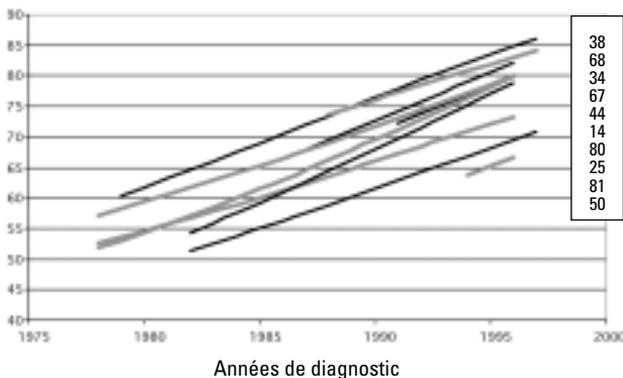


FIGURE 1.

Évolution du taux standardisé d'incidence dans les dix départements où un registre existe (les numéros des départements sont listés par ordre décroissant du taux en 1996/1997).

1. <http://www.cepidc.vesinet.inserm.fr>

d'incidence pour la population des registres en 2000 en pondérant chaque département par sa population⁽²⁾. On a ainsi résolu le problème d'absence de simultanéité des observations dans les dix départements disponibles. À titre indicatif, l'utilisation de ce taux d'incidence donnerait une estimation de 40 764 cas en 2000 pour la France entière. Cette approche ne nous libère pas, toutefois, de l'absence de représentativité des registres. Il nous faut donc redresser notre estimation en tenant compte d'autres informations comparant la population des registres et la population complémentaire. L'estimation faite récemment [7] utilise la mortalité par cancer du sein comme information complémentaire. La démarche décrite dans cette dernière référence est assez sophistiquée. Pour mieux la saisir, on donne ci-dessous quelques indications sur l'évolution de la mortalité par cancer du sein dans les départements où existe un registre et on compare la mortalité dans la population constituée de ces départements à celle de la population française.

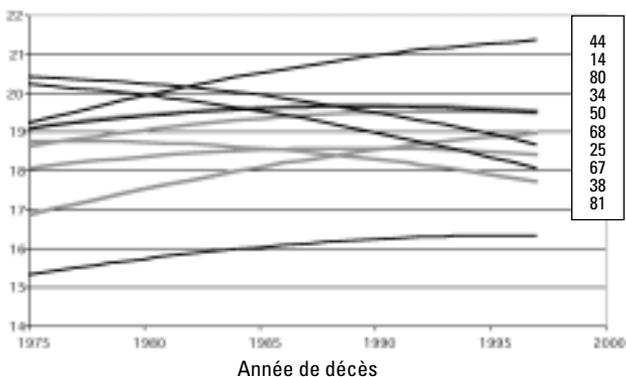


FIGURE 2.
Évolution du taux standardisé de mortalité dans les dix départements où un registre existe (les numéros des départements sont listés par ordre décroissant du taux en 1997).

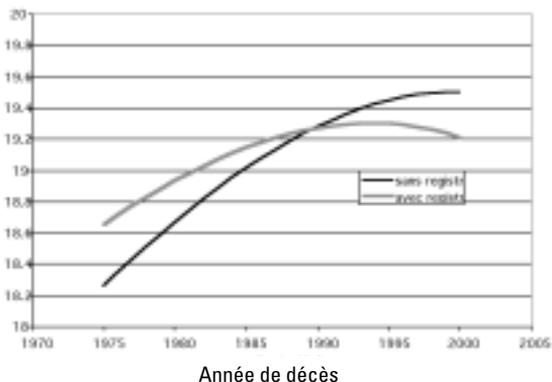


FIGURE 3.
Évolution de la mortalité dans la population des départements avec et sans registre.

2. Cette estimation peut être faite de façon fiable pour toutes les années entre 1980 et 2000. Cette dernière a été choisie, car on souhaite discuter l'estimation pour l'année 2000.

La tendance de la mortalité par cancer du sein en France

Contrairement à celle de l'incidence, la tendance de la mortalité est passablement hétérogène (*figure 2*). Certains départements ont vu leur mortalité substantiellement diminuer, alors que pour d'autres, l'augmentation s'est poursuivie. Cela donne en moyenne pour la population des registres une augmentation modérée. Dans la population complémentaire, cette importante hétérogénéité existe également, mais en moyenne, les deux populations ont une tendance modérée mais contradictoire (*figure 3*). En particulier, la population des registres a, pour l'année 2000, une mortalité plus faible que celle de la population complémentaire : le taux brut de mortalité de la population des registres est de 35,9 alors que celui de la France entière est de 37,1. Il semble alors acceptable de corriger l'estimation de l'incidence nationale du facteur correspondant (1,033). En procédant ainsi, on obtient une estimation de 42 109 cas en France ($40\,764 \times 1,033$) pour l'année 2000, un chiffre assez proche de celui publié et rappelé dans le *tableau I*.

Dogmes et doutes

Le dogme affirme qu'il y a 42 000 cas de cancer du sein en France en 2000. En d'autres termes, le nombre de cas a doublé entre 1980 et 2000. Une partie de cette augmentation est due aux changements démographiques liés à la taille et à la structure de la population. Le taux standardisé corrigé de ces facteurs a augmenté de 60 %, une croissance qui reste considérable. On peut concevoir que ces chiffres surprennent et inspirent quelques doutes quant à leur validité. Une tentative de quantification de ces doutes a été proposée dans la publication de la *Revue d'épidémiologie et de santé publique* [5], et largement ignorée par les lecteurs de cet article. Les auteurs ont fourni un intervalle de valeurs crédibles pour le nombre de cas en 2000 fondé essentiellement sur l'incertitude liée à l'hétérogénéité des tendances dans les registres et à l'imperfection de la liaison entre incidence et mortalité. Cet intervalle est obtenu en évaluant la variabilité de l'estimation, que l'on obtient lorsqu'on enlève tour à tour chacun des registres pour obtenir le nombre de cas incidents en 2000. La fourchette pour le cancer du sein était de 37 731-45 959. Cet intervalle est, à mon avis, un peu pessimiste, mais il ne laisse aucun doute quant à l'ampleur de l'accroissement de l'incidence entre les années 1980 et 2000. En tout état de cause, celui-ci est apprécié par la tendance observée dans les registres existants qui, comme on l'a vu, est remarquablement homogène. Il n'y a aucune raison de penser que la situation est différente dans les autres départements. Les doutes ne subsistent que pour le niveau exact de l'incidence, mais la similarité de la mortalité entre la population des registres et la population complémentaire laisse penser que ces doutes sont probablement moins importants que suggérés par la fourchette ci-dessus. On comprend, en effet, que l'omission de données au niveau du registre pour évaluer la variabilité de l'estimateur utilisé est une méthode un peu brutale en raison de l'hétérogénéité de la contribution de chacun

d'eux. L'omission de l'Isère ou du Bas-Rhin, dont les données sont disponibles sur toute la période, n'a, en effet, rien à voir avec celle de la Manche ou de la Loire-Atlantique, départements pour lesquels peu de données ont été recueillies.

Validité de la modélisation

Dans l'étude des tendances de l'incidence et de la mortalité, les épidémiologistes ont l'habitude d'étudier et de représenter les taux standardisés "observés" en fonction du temps. Si l'on utilise cette approche pour des données départementales, on obtient des graphiques chaotiques sur lesquels le "bruit" aléatoire a complètement noyé le signal. De façon générale, cette approche est critiquable : elle ne rend pas compte correctement de l'évolution de l'incidence ou de la mortalité, et l'extrapolation des taux standardisés est une très mauvaise méthode de prédiction [8]. C'est la raison pour laquelle nous avons choisi la modélisation pour extraire le signal du bruit. Le modèle que nous avons préféré est connu sous le nom de "age cohorte". Comme il a été rappelé plus haut, le principal avantage de ce modèle est qu'il mesure la progression du risque avec l'âge chez les mêmes personnes⁽³⁾ d'une année d'âge à la suivante, dans l'intervalle d'observation disponible pour une cohorte de naissance donnée. Les règles statistiques nous disent que la qualité prédictive des modèles est bonne. L'ajout d'une variable prenant en compte l'année de décès améliorerait, cependant, l'adéquation du modèle pour la mortalité. Toutefois, cette variable, bien que significative, ne conduirait vraisemblablement pas à une meilleure prédiction [5]. Dans une approche critique de la modélisation effectuée, il est important de réaliser que le modèle qui a permis de construire les courbes présentées dans les **figures 1** et **2** a considérablement simplifié la description de la tendance, mais a permis d'en retenir l'essentiel : croissance quasi linéaire du taux standardisé d'incidence et importante hétérogénéité interdépartementale des tendances de la mortalité. Cette remarque vaut aussi pour la comparaison entre les sous-populations avec et sans registre. Ce compromis entre qualité descriptive des données déjà observées et qualité prédictive pour les données futures est à la base de toute démarche de modélisation. Cette réflexion sur le choix judicieux d'un modèle est particulièrement pertinente pour la modélisation de l'incidence et de la mortalité, qui utilise des données entachées d'erreurs, systématiques pour certaines données d'incidence, aléatoires mais importantes pour les données de mortalité.

Discussion et conclusion

Ainsi, on a compris que l'estimation de l'incidence du cancer du sein en France résulte essentiellement des données d'incidence des registres et que la mortalité ne joue qu'un rôle minime de redressement de l'échantillon. L'utilisation de la modélisation statistique

3. Pour la grande majorité des personnes suivies.

TABLEAU II. "Résidus"⁽¹⁾ selon la période de diagnostic et le groupe d'âge.

Période de diagnostic	1975-79	1980-84	1985-89	1990-94	1995-97
Âge					
< 50	0,97	1,01	1,04	1,02	0,96
50-69	1,02	0,95	1,00	1,03	0,97
70+	0,98	0,96	1,01	1,03	0,95

1. Rapport du nombre de cas observés et du nombre attendu selon le modèle âge-cohorte retenu.

a permis d'extraire le signal du bruit de façon adéquate. Il n'y a donc aucune raison de douter de la réalité des chiffres cités dans le titre de cet article, à condition d'accompagner chacun d'eux de l'année à laquelle ils se réfèrent. L'augmentation du nombre de cas de cancer du sein entre 1980 et 2000 est considérable (100 %) et la contribution de l'augmentation de l'espérance de vie ne joue qu'un faible rôle dans cette augmentation puisque, après correction, la croissance reste très élevée (60 %). Ce que l'on observe en France est similaire à ce qui est observé dans d'autres pays développés où l'activité de dépistage est importante. Aux États-Unis, le taux d'incidence standardisé⁽⁴⁾ a augmenté de 50 % entre 1975 et 1995, puis s'est stabilisée à un niveau très élevé (taux standardisé pour la population mondiale supérieur à 90) ; la mortalité a commencé à diminuer à la date où l'incidence a cessé d'augmenter (1992-1995)⁽⁵⁾. Il n'est pas impossible que le même scénario se reproduise en France. Le rôle du dépistage dans ce type d'évolution est indiscutable. Quantifier ce rôle est difficile à partir des données actuellement disponibles dans les registres et dans les bases de données du dépistage [9, 10]. Ce thème est trop complexe pour être discuté dans le cadre limité de cet article, mais on peut remarquer que, selon le modèle retenu, l'augmentation de l'incidence ne se limite pas à la classe d'âge théorique du dépistage (50-69 ans). Pour chaque département, l'incidence augmente dans chaque classe d'âge selon des modalités différentes qui dépendent du département, mais, en général, le taux augmente plus vite chez les femmes plus âgées, ce qui ne suggère pas un effet favorable du dépistage⁽⁶⁾. Pour vérifier l'adéquation de cette conclusion impliquée par la modélisation retenue, nous avons présenté, dans le **tableau II**, les "résidus" du modèle selon la période de diagnostic et le groupe d'âge. Nous avons aussi examiné ces mêmes résidus par département. Ces chiffres suggèrent une légère surestimation dans la dernière période, particulièrement pour les femmes plus âgées, mais aucune déviation systématique qui indiquerait des tendances différentes de

4. Sur la population des États-Unis de l'année 2000, ce qui montre une belle ignorance du reste du monde...

5. http://seer.cancer.gov/csr/1975_2000.

6. Si le dépistage 50-69 ans anticipait le diagnostic des tumeurs, le taux d'incidence devrait diminuer après 70 ans...

ce que prédit le modèle. La mise à jour de ces calculs avec les données les plus récentes permettra de savoir si, comme aux États-Unis, l'augmentation de l'incidence se ralentit effectivement et si la décroissance de la mortalité s'amorce dans tous les départements français.

L'auteur remercie l'association FRANCIM et le CépiDc, qui ont mis à sa disposition les données utilisées dans cet article.

Références bibliographiques

- [1] Benhamou E, Laplanche A, Wartelle M, Favier J, Gignoux M, Menegoz F et al. *Statistiques de santé – Incidence des cancers en France 1978-1982*. Paris : INSERM, 1990.
- [2] Jensen OM, Estève J, Moller H, Renard H. *Cancer in the European Community and its member states*. *Eur J Cancer* 1990;26(11-12): 1167-256.
- [3] De Vathaire F, Koscielny S, Rezvani A, Laplanche A, Estève J, Ferlay J. *Statistiques de Santé – Estimation de l'incidence des cancers en France 1982-1987*. Paris : INSERM, 1996.
- [4] Menegoz F, Black RJ, Arveux P, Magne V, Ferlay J, Buemi A et al. *Cancer incidence and mortality in France in 1975-95*. *Eur J Cancer Prev* 1997;6(5):442-66.
- [5] Remontet L, Estève J, Bouvier AM, Grosclaude P, Launoy G, Menegoz F et al. *Cancer incidence and mortality in France over the period 1978-2000*. *Rev Epidemiol Sante Publique* 2003;51:3-30.
- [6] Bashir S, Estève J. *Projecting cancer incidence and mortality using Bayesian age-period-cohort models*. *J Epidemiol Biostat* 2001; 6(3):287-96.
- [7] Remontet L, Buemi A, Velten M, Jouglà E, Estève J. *Évolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000*. *INVS* 2003;1-217.
- [8] Estève J, Benhamou E, Raymond L. *Statistical methods in cancer research. Descriptive epidemiology*. IARC, Press, 1994.
- [9] Zahl PH, Strand BH, Maehlen J. *Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study*. *BMJ* 2004;328(7445):921-4.
- [10] Paci E, Warwick J, Falini P, Duffy SW. *Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern?* *J Med Screen* 2004;11(1):23-7.