

Towards a Personalized Organized Screening in Breast Cancer : Practical Determination of Thresholds of Risk in Women at Average-Risk

Emmanuel Bonnet (✉ emmanuel.bonnet@umontpellier.fr)

University of Montpellier

Jean Pierre Daures

University of Montpellier

Paul Landais

University of Montpellier

Research Article

Keywords: breast cancer (BC), French Hérault region, menopause

Posted Date: January 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-147580/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Towards a personalized organized screening in breast cancer : practical determination of thresholds of risk in women at average-risk

Emmanuel Bonnet^{1,2,*}, Jean-Pierre Daures^{1,2}, and Paul Landais¹

¹Montpellier University, EA2415, Institut Universitaire de recherche clinique, 34093 Montpellier CEDEX 5, FRANCE

²Languedoc Mutualité, Nouvelles Technologies, AESIO, Montpellier, France

* emmanuel.bonnet@umontpellier.fr

ABSTRACT

In France, more than 10 million women at "average" risk of breast cancer (BC), are included in the organized BC screening. Existing predictive models of BC risk are not adapted to the French population. Thus, we set up a new score in the French Hérault region and looked for a graded level of risk in women at "average" risk. We recruited a retrospective cohort of women, aged 50 to 60, who underwent the organized BC screening, and included 2241 non-cancer women and 527 who developed a BC during a 12-year follow-up period (2006-2018). The risk factors identified were high breast density (ACR BI-RADS grading)(B vs A: HR 1.41, 95%CI [1.05; 1.9], p=0.023; C vs A: HR=1.65 [1.2; 2.27], p=0.02 ; D vs A: HR=2.11 [1.25;3.58],p=0.006), a history of maternal breast cancer (HR=1.61 [1.24; 2.09], p < 0.001), and socioeconomic difficulties (HR 1.23 [1.09; 1.55], p=0.003). While early menopause (HR=0.36 [0.13; 0.99], p=0.003) and an age at menarche after 12 years (HR=0.77 [0.63; 0.95], p=0.047) were protective factors. We identified 3 groups at risk: lower, average, and higher, respectively. A low threshold was characterized at 1.9% of risk and a high threshold at 4.5%. Mean risks in the 3 groups of risk were 1.37%, 2.68%, and 5.84%, respectively. Thus, 12% of women presented a level of risk different from the average risk group, corresponding to 600,000 women involved in the French organized BC screening, enabling to propose a new strategy for performing an organized and personalized national BC screening.

Introduction

Breast cancer is the most common cancer in women and the leading cause of death from cancer in women in France. The earlier breast cancer is detected, the greater the chances of recovery. For a breast cancer detected at an early stage, 5-year survival is 99%, and 27% for a metastatic cancer¹. The difference in the detection stage also influences the cost of medical care². While high risk women³ receive an individualized monitoring, the majority of women are included in the nationwide organized breast cancer screening. In France, this screening consists of a mammography (2 views) every two years with a double reading of mammographic images.

In April 2017, the Ministry of Health published an action plan to renovate the organized breast cancer screening⁴ based on the recommendations of the National Cancer Institute. This reorganization aimed to enable a more personalized organized screening, and encouraged research projects on tools and methods for assessing the level of risk, including scoring. We wondered whether, among the 10 million women at "average-risk" targeted for the French organized screening⁵, some women would be at lower-risk than average risk and others at higher risk.

Several predictive models already exist, such as Gail⁶, BRCA PRO⁷, BODICEA⁸, or Tyrer-Cuzick⁹ models. But they are not adapted to the European population¹⁰ and most of them include BRCA1 and BRCA2 mutations. However, in France, women presenting with these mutations are not included in the organized nationwide screening. Moreover, it is difficult to estimate an individual risk precisely^{11,12}. We, therefore, intended to set up a new risk model, including new variables, adapted to the French population, and enabling to explore whether in this average risk population it was possible to identify women at lower risk of breast cancer (or higher risk, respectively) in order to adapt the organized screening accordingly.

Material and methods

Database

We used the database of the Center that coordinates breast cancer screening (CRCDC-OC) in the Hérault department, France. This database registers women with precise, verified, and consolidated information on the development or not of breast cancer during the follow-up of their screening. We enrolled women, aged 50 to 60 at inclusion, who experienced a 12 year breast

cancer screening from 2006. We extracted a database of 33,858 women, including 805 women who developed a breast cancer in 3-12 years following inclusion. We excluded women who have been diagnosed with breast cancer during the first two year of follow-up to avoid including women who might have already developed a cancer but not yet diagnosed at inclusion. Data were recovered by a telephone survey. It was not possible to interview 33,000 women. Therefore, on one hand we selected all the women who developed a breast cancer during the follow-up period, and on the other hand, we randomly selected 5000 women among those who did not develop breast cancer.

Power

We hypothesized a risk ratio ≥ 2 , with a ratio of 4 to 1 between women who did not develop breast cancer and those who did during the follow-up period, together with a power of 80%. It gave a required sample size of 543 women with breast cancer during the follow-up period and 2174 women without breast cancer.

Study questionnaire

We looked for the risk factors for breast cancer in average risk women reported in the literature prior to creating the study questionnaire. We elaborated a text mining algorithm-assisted search (TEMAS) (available at shiny.temas-bonnet.site) to optimize the bibliographic search on very large volumes of scientific articles. We also included the variables used in Gail's model relative to women at average risk. The study questionnaire is provided in Supplementary material 1.

Data collection

Data were collected by telephone by two trained investigators. An explanatory leaflet (Supplementary material 3) was sent to all participants prior to the phone call and informed consent was obtained from all individual participants prior to the telephone interview. Responses to the study questionnaire were collected between November 2018 and May 2019. The inclusion flowchart is shown on Figure 1. This survey enabled collecting 3077 questionnaires, including 597 questionnaires from women who developed a breast cancer. The Case Report Form was developed with the ACCESS software. The analysis database was provided by the CRCDC-OC while having anonymized the data beforehand. This study complies the MR003 CNIL's (National Commission for Informatics and Liberties) methodology (Declaration: 2200402v0). The study was also agreed by the ethics committee of the CPP (Personal Protection Committee) (A0246039). All methods were performed in accordance with the relevant guidelines and regulations.

Figure 1. Flowchart of the retrospective cohort study

Statistical analysis

Data management was performed on the fly. Statistical analysis was carried out using R software¹³.

Descriptive analysis

We discretized the variables according to the literature. For duration of oral contraceptive¹⁴ we retained the classes 0, [1,5[, [5,10[, and ≥ 10 . For the number of biopsies, we complied with Gail's classes⁶ 0, 1, ≥ 2 . Breast density was categorized in four classes according to the ACR BI-RADS (Breast Imaging-Reporting and Data System) Atlas¹⁵ : A: the breasts are almost entirely fatty, B: there are scattered areas of fibroglandular density, C: the breasts are heterogeneously dense, and D: the breasts are extremely dense. For the number of hours per week of physical activity we followed WHO recommendations¹⁶ : 0, [1-5[, and ≥ 5 . Alcohol consumption (in glasses per week) was discretized as follows : 0, [1,10[, ≥ 10 according to Santé Publique France (French public health service) recommendations¹⁷. Age at menarche was divided as follow: < 12 and ≥ 12 ¹⁸. Age at first live birth was discretized in 3 classes¹⁹: nulliparous, ≤ 24 , and > 24 . For number of children, we used the average number of children in the generation of interest²⁰: 0, [1,3[, and ≥ 3 . For the total duration of breastfeeding we followed WHO and UNICEF recommendations²¹: 0, [1,7[, and ≥ 7 . For age at menopause, we used the definition of the national college of French gynecologists of early menopause²²: < 41, ≥ 41 . We created a socio-economic hardship binary variable that was "yes" if the interviewee presented financial hardship and / or was not going on vacation, "no" otherwise. We also studied overweight (i.e. $BMI \geq 25$). We also studied detailed family history of breast cancer : mother, sister, daughter, and aunt. All the quantitative variables were also tested as continuous variables.

Cox model

All variables with a P-value < 0.2 in the univariate models were included in a global Cox model, then a selection with the stepwise Akaike information criterion (AIC) was achieved (age was forced in the model) in order to obtain $\hat{\beta}$, the estimated coefficients of the Cox proportional hazards regression model. Schoenfeld residuals were computed to check the proportional hazards assumption.

Our goal was to define an individual score. Therefore, to estimate $\hat{\Lambda}_0$, the cumulative base-line hazard function, we used Breslow's estimator²³ defined below.

$$\hat{\Lambda}_0(t) = \sum_{i:T_i \leq t} \frac{d_i}{\sum_{j \in R(T_i)} \exp(\hat{\beta}' Z_j)} \quad (1)$$

with:

- $T_1 < T_2 < \dots < T_{12}$ distinct times to event
- d_i the number of events detected at T_i
- Z_i the value of the covariates for patient i
- $R(T_i)$ all individuals still at risk at time T_i^- (just before T_i)
- $\hat{\beta}$ the estimated coefficients of the Cox proportional hazards regression model

It enabled to estimate a risk function F for each individual as a function of its covariates :

$$\hat{F}(t|Z) = 1 - \hat{S}(t|Z) = 1 - \exp(-\hat{\Lambda}_0(t) \exp(\hat{\beta}' Z))$$

Bootstrap

As we performed a random sampling among women without breast cancer, in our database, the distribution between women who developed cancer and the others was no longer representative of the starting population. To remedy this, we performed bootstrap samples. We computed 100,000 bootstrap databases using stratified random sampling with replacement among women for whom we obtained answers to all the covariates of the final Cox model, in order to obtain a database close to the initial CRCDC database (i.e. 33,858 women, including 896 who developed a breast cancer within the 12 years of follow-up)

In these databases, mean risk was identical to the one of the CRCDC database. Then, we estimated, for each woman, $\Lambda_0(t)$, and calculated $F(12|Z)$ to obtain a risk score of developing a breast cancer during the follow-up period.

Thresholds and likelihood ratios

In France, the target population of the organized screening is greater than 10 million women. Current participation is about 50%⁵ with a target of 65%. The number of women actually screened is therefore between 5 and 6.5 million. Among women at average risk, we hypothesized that 10% of them might have a risk different from the "average" risk, either lower or higher. This average risk was 2.6% in this population. From a clinical standpoint, we looked for women that might have a risk two times lower than this "average" risk. This group was considered as at lower-risk. We also looked for women that might have a risk two times higher than this "average" risk. This group was considered as at higher-risk. We then searched for the corresponding thresholds. We looked for s_1 , the higher risk threshold, and s_2 , the lower risk threshold, such as:

$$\frac{\mathbb{P}(M|R > s)}{\mathbb{P}(M|R < s)} \geq 2 \quad \forall s \geq s_1$$

and

$$\frac{\mathbb{P}(M|R < s)}{\mathbb{P}(M|R > s)} \leq 0.5 \quad \forall s \leq s_2$$

where:

- M represents the event "developing a breast cancer during the 12 years follow-up";
- R represents the risk as determined by the score.

This method then enabled to delineate 3 groups according to the level of risk: lower risk, average risk, and higher risk. Given that, for each threshold s , the likelihood ratios (LR) either positive (+) or negative (-) where as follows:

$$LR + (s) = \frac{\text{sensitivity}(s)}{1 - \text{specificity}(s)} \quad LR - (s) = \frac{1 - \text{sensitivity}(s)}{\text{specificity}(s)}$$

We plotted the curves $LR+(s)$ and $LR-(s)$ curves, and looked for s_1 and s_2 according to the calculations available in Supplementary material 2.

Women with a risk score below the "low threshold" will be considered at lower risk of breast cancer, women with a risk level between "low threshold" and "high threshold" will be considered at average risk and finally, women with a risk score above the "high threshold" will be considered at higher risk of breast cancer in this population of women at average risk.

Results

Personalized organized breast cancer screening retrospective cohort: descriptive analysis

Descriptive analysis results are presented in **Table 1**. The significant risk factors in the univariate analysis were the number of biopsies ($p = 0.042$), breast density ($p = 0.012$), total duration of breastfeeding ($p = 0.026$), financial difficulties and/or not going on vacation ($p = 0.002$), mother history of breast cancer ($p = 0.001$), and aunt history of breast cancer ($p = 0.029$).

We explored whether a selection bias might have occurred. We analyzed the distribution of the two variables (age and breast density) we had for women who could not be reached or who refused to answer our questions. The results are shown in Tables **2** and **3**. The only significant difference was observed for breast density of women who did not develop cancer during follow-up. The women who responded to the survey had significantly higher breast density on average. So the women selected, in the group without breast cancer, were more at risk than those who did not respond. Under these conservative conditions, we considered that the selection bias would not penalize the analysis. Moreover, we did not retain the fact that breast density as such was a criterion of choice for women to answer the questionnaire or not.

Cox model

The final model, presented in Table **4**, has been built with 2768 women for whom we had answers to all the covariates of the model, including 2241 non-cancer patients and 527 women who developed cancer during the follow-up period.

The risk factors selected by the model were high breast density (ACR BI-RADS grading) (B vs A: HR 1.41, 95% CI [1.05; 1.9], $p=0.023$; C vs A: HR=1.65, 95% CI [1.2; 2.27], $p=0.02$; D vs A: HR=2.11, CI [1.25; 3.58], $p=0.006$), history of maternal breast cancer (HR=1.61, 95% CI [1.24; 2.09], $p < 0.001$), and socioeconomic difficulties (HR 1.23, 95% CI [1.09; 1.55], $p=0.003$). While early menopause (HR=0.36, 95% CI [0.13; 0.99], $p=0.003$) and an age at menarche after 12 years (HR=0.77, 95% CI [0.63; 0.95], $p=0.047$). Age (HR 1.024, 95% CI [0.997; 1.05], $p = 0.084$) and interaction between age at menarche and menopause (HR 2.73, 95% CI [0.91; 8.19], $p = 0.073$) have been forced in the model.

Bootstrap

We computed 100,000 bootstrap databases using random sampling with replacement among the 2768 women for whom we had answers to all the covariates of the final Cox model. The risk function $F(t)$ was calculated for all women in the 100,000 databases.

Threshold determination

Pre-test prevalence of the disease was the percentage of breast cancer in the CRCDC-OC database i.e. 2.6%.

So, as described in Supplementary material 2, we looked for :

- s_1 such as $\forall s \geq s_1$:

$$LR_+(s) \geq 2.054$$

- s_2 such as $\forall s \leq s_2$:

$$LR_-(s) \leq 0.493$$

We then plot $LR+(s)$ and $LR-(s)$ in Figure **2**. This method gave a low threshold at 1.9% and a high threshold at 4.5%.

Figure 2. Determination of the three risk groups with optimal thresholds in women at average risk of breast cancer: "Lower risk", "Average risk" and "Higher risk".

On one hand, 9.39% of women presented a risk lower than 1.9%. Among this group, the mean risk was 1.37%. On the other hand, 2.85% of women presented a risk higher than 4.5% with a mean risk of 5.84%.

Discussion

In a retrospective cohort of French women, aged 50 to 60, at average risk of breast cancer and followed-up for 12 years, from 2006 to 2018, who underwent the organized screening, we identified two additional levels of risk, either twice lower than the average-risk, or twice higher.

Our partition into three groups at risk of breast cancer in these women showed that the level of risk in the so-called "average-risk" population is not homogeneous.

Depending on the presence or level of clinical, genetic or biological markers, treatment strategies are personalized to be as effective as possible. This approach can be extended to public health and in particular for primary and secondary prevention. Indeed, offering personalized prevention strategies could increase adherence to the program, and effectiveness as well. Thus, identifying new groups at risk among women at average risk offers new perspectives.

Women's wishes for a more personalized organized screening²⁴ therefore appears attainable. Indeed, women interviewed during the French citizen consultation on breast cancer screening in 2017, expected the organized screening to be more targeted, thanks to advances in research, and to become more personalized according to individual risk factors. Physicians, also consulted, mentioned the value of determining the level of individual risk^{25,26}. They also ask for a score in French women in order to substantiate the levels of risk. This risk assessment would be used by the screening management structures to offer a personalized approach of the organized screening, adapted to the level of risk of each woman. We developed a user friendly computer application in this perspective.

The value of the organized screening is undeniable. In France, it enables a regular follow-up of invited women, aged 50 to 74, to carry out a free mammogram in the radiology office of their choice every two years. Moreover, if the initial assessment does not present any anomaly, a second reading of the mammography is systematically performed by another radiologist. The second independent reading is a guarantee of reliability, with almost 8% of the breast cancers detected by the independent second reading^{27,28}. The principle of an organized breast cancer screening has been challenged given the claims of over-diagnosis or over-treatment. A reassessment of the organized screening that would allow a personalization according to the individual level of risk, would enable a standardized and quality monitoring of women while producing evaluative data.

We identified three levels of risk with a risk ratio of at least 2 between the group at average risk and the group at lower risk on one hand, and on the other hand with the group at higher risk. It lead to fairly large group sizes. The choice of a risk divided or multiplied by a factor two remains arbitrary though clinically relevant. Moreover, sensitivity analysis could be carried out by varying this ratio according to the complementary hypotheses of epidemiologists or physicians specialized in this field. Furthermore, the choice could be supported by a cost function taking into account monetary cost and years of life adjusted for quality of life.

According to the two thresholds of lower and higher risk, 12% of women were thus identified, corresponding to more than 600,000 women in the French population. In this population at "average" risk of breast cancer, a new screening strategy could therefore be proposed for the groups at lower or higher risk. For the group at average-risk of breast cancer, corresponding to 88% of the presently screened population, the current screening rate, i.e. a mammogram every two years, would be maintained as such.

For the group at lower risk that includes 9.4% of the screened population with a level of risk of 1.37% of developing breast cancer within the 12 years of follow-up, we propose to reduce the frequency of mammograms with a mammogram every three or more years instead of the two years currently in force. This mammogram spacing could reduce the risk of over-diagnosis while maintaining a regular follow-up. Moreover, it would reduce, the risk of radiation-induced cancer, even if it is already very low. However, the risk of interval cancer can't be excluded. This proposal will have thus to be tested with an appropriate design.

For the 2.9% of women at higher risk, increasing the frequency of mammograms is possible. However, high breast density is an important risk factor for this group. In this group, the distribution of high or very high breast density was 60% compared to 28% in the group at average risk and 3% in the group at lower risk. Furthermore, the sensitivity of the mammogram falls to 63%²⁹ for very dense breasts. The sensitivity of cancer detection via a mammogram appeared inversely proportional to the breast density³⁰. Thus, it did not seem that increasing the frequency of mammograms would bring a significant benefit. Based on recent results, an alternative would be to offer these women an abbreviated breast MRI, which appeared associated with a higher rate of breast cancer detection among women with dense breasts³¹, while maintaining a two year frequency for surveillance. This would bring the care strategy for these women at higher-risk closer to the strategy applied to women already identified at high risk (i.e. different from women here identified in the "average" risk population), and benefiting from an individual follow-up³.

Such a personalization of the organized screening might make women feeling more involved. For women at lower-risk, a mammogram every three years may be less binding than every two years. For women at higher risk, offering a more specific examination may increase adhesion to the organized screening. These strategies could then increase the participation of French women to the organized screening which is currently around 50%⁵, though the target is 65%.

As previously stated, the cost of the case finding (including diagnosis and treatment of patients diagnosed) should be

economically balanced in relation to possible expenditure on medical care as a whole³². An analysis of the medico-economic gain must be carried out taking into account, in particular, interval cancers, over-diagnosis and the stage of the cancers detected.

This study proposes a new approach to personalize the organized screening by stratifying the risk within the population at "average" risk and currently considered as homogeneous. However, the results of this study remain to be confirmed in other contexts to validate whether the proposed strategy could be implemented.

Six variables were included our the score of breast cancer risk, two of which (age and breast density) are already included in the databases of the different French CRCDCs. Only four more questions are required to be asked in order to establish the score. This information is easy and simple to collect at the time of mammogram. Moreover, an internal validation on another database, for example of women aged 40 to 50 in the Hérault department, could be performed. A double external geographic validation could be carried out on the databases from other CDCRCs and a database from another country³³.

We collected information of interest for the creation of our score at t_0 , the inclusion time. But, over time, the value of variables retained at t_0 may change and the risk at $t_0 + x$ years may also change. It will therefore be necessary to regularly recalculate women risk of breast cancer in order to assign them to the appropriate group of risk.

A re-evaluation of the score will also be necessary on regular time intervals to reappraise the coefficients estimated by the model in order to evaluate whether they would need to be adjusted.

Our proposal for a new personalized organized screening must be validated by other studies.

References

1. Lefevre, D., Catajar, N., Le, C., Benjamin, B. & Bousquet, P. J. Dépistage du cancer du sein : impact sur les trajectoires de soins. Tech. Rep., Institut national du cancer (2018).
2. Jay, N., Nuemi, G., Gadreau, M. & Quantin, C. A data mining approach for grouping and analyzing trajectories of care using claim data: The example of breast cancer. *BMC Med. Informatics Decis. Mak.* **13**, DOI: [10.1186/1472-6947-13-130](https://doi.org/10.1186/1472-6947-13-130) (2013).
3. Haute Autorité de Santé. Dépistage du cancer du sein en France : identification des femmes à haut risque et modalités de dépistage. Tech. Rep., HAS (2014).
4. INCa. Plan d'action pour la rénovation du dépistage organisé du cancer du sein. Tech. Rep., Institut National du Cancer (2017).
5. Santé publique France. Taux de participation au programme de dépistage organisé du cancer du sein 2018-2019 et évolution depuis 2005. Tech. Rep., Santé publique France (2020).
6. Gail, M. H. *et al.* Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.* **81**, 1879–1886, DOI: [10.1093/jnci/81.24.1879](https://doi.org/10.1093/jnci/81.24.1879) (1989).
7. Euhus, D. M. *et al.* Pretest prediction of BRCA1 or BRCA2 mutation by risk counselors and the computer model BRCAPRO. *J. Natl. Cancer Inst.* **94**, 844–851 (2002).
8. Antoniou, A. C., Pharoah, P. P., Smith, P. & Easton, D. F. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br. J. Cancer* **91**, 1580–1590, DOI: [10.1038/sj.bjc.6602175](https://doi.org/10.1038/sj.bjc.6602175) (2004).
9. Tyrer, J., Duffy, S. W. & Cuzick, J. A breast cancer prediction model incorporating familial and personal risk factors. *Stat. Medicine* **23**, 1111–1130, DOI: [10.1002/sim.1668](https://doi.org/10.1002/sim.1668) (2004).
10. Pastor-Barriuso, R. *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.* **138**, 249–259, DOI: [10.1007/s10549-013-2428-y](https://doi.org/10.1007/s10549-013-2428-y) (2013).
11. McTiernan, A., Gilligan, M. A. & Redmond, C. Assessing individual risk for breast cancer: Risky business. *J. Clin. Epidemiol.* **50**, 547–556, DOI: [10.1016/S0895-4356\(97\)00013-9](https://doi.org/10.1016/S0895-4356(97)00013-9) (1997).
12. Wang, X. *et al.* Assessment of performance of the Gail model for predicting breast cancer risk: A systematic review and meta-analysis with trial sequential analysis. *Breast Cancer Res.* **20**, DOI: [10.1186/s13058-018-0947-5](https://doi.org/10.1186/s13058-018-0947-5) (2018).
13. R Core Team. R: A Language and Environment for Statistical Computing (2019).
14. Mørch, L. S. *et al.* Contemporary hormonal contraception and the risk of breast cancer. *New Engl. J. Medicine* **377**, 2228–2239, DOI: [10.1056/NEJMoa1700732](https://doi.org/10.1056/NEJMoa1700732) (2017).
15. D'Orsi, S. E., C J & al. 2013 ACR BI-RADS Atlas: Breast Imaging Reporting and Data System (American College of Radiology, 2014).
16. WHO. *Global recommendations on physical activity for health* (WHO Library Cataloguing-in-Publication Data, 2010).

17. Santé publique France (MILDECA). Nouvelles recommandations sur l'alimentation, y compris l'alcool, l'activité physique et la sédentarité | Mildeca (2019).
18. Lalys, L. & Pineau, J. C. Age at menarche in a group of French schoolgirls. *Pediatr. Int.* **56**, 601–604, DOI: [10.1111/ped.12296](https://doi.org/10.1111/ped.12296) (2014).
19. Volant, S. Un premier enfant à 28,5 ans en 2015 : 4,5 ans plus tard qu'en 1974. *INSEE Première* **1642**, 2015–2018 (2017).
20. Robert-Bobée, I. 2,1 enfants par femme pour les générations nées entre 1947 et 1963. *Insee Focus. Avril* (2015).
21. WHO. Planning Guide for national implementation of the Global Strategy for Infant and Young Child Feeding. Tech. Rep., WHO (2007).
22. Collège National des Gynécologues et Obstétriciens Français. La ménopause.
23. BRESLOW & E., N. Contribution to discussion of paper by D. R. Cox. *J. Roy. Stat. Soc., Ser. B* **34**, 216–217 (1972).
24. INCa. Avis de la conférence des citoyennes. Tech. Rep., INCa (2016).
25. INCa. Avis de la conférence des professionnels. Tech. Rep., INCa (2016).
26. Cases, C. *et al.* Rapport du comité d'orientation sur le dépistage du cancer du sein. Tech. Rep., INCA (2016).
27. Doutriaux-Dumoulin, I., Allioux, A., Campion, L., Meingan, P. & Molina, L. Cancers détectés par le deuxième lecteur : analyse des données de la campagne de dépistage du cancer du sein en Loire-Atlantique, 2003-2005 (nouveau cahier des charges). *J. Radiol* **1137**, 1839–1910 (2007).
28. Thurfjell, E. L., Lernevall, K. A. & Taube, A. A. Benefit of independent double reading in a population-based mammography screening program. *Radiology* **191**, 241–244, DOI: [10.1148/radiology.191.1.8134580](https://doi.org/10.1148/radiology.191.1.8134580) (1994).
29. Carney, P. A. *et al.* Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Annals Intern. Medicine* **138**, 168–175, DOI: [10.7326/0003-4819-138-3-200302040-00008](https://doi.org/10.7326/0003-4819-138-3-200302040-00008) (2003).
30. Drukesteinis, J. S., Mooney, B. P., Flowers, C. I. & Gatenby, R. A. Beyond mammography: New frontiers in breast cancer screening, DOI: [10.1016/j.amjmed.2012.11.025](https://doi.org/10.1016/j.amjmed.2012.11.025) (2013).
31. Comstock, C. E. *et al.* Comparison of Abbreviated Breast MRI vs Digital Breast Tomosynthesis for Breast Cancer Detection among Women with Dense Breasts Undergoing Screening. *JAMA* **323**, 746–756, DOI: [10.1001/jama.2020.0572](https://doi.org/10.1001/jama.2020.0572) (2020).
32. Wilson, J. M. G., Jungner, G. & World Health Organization. Principles and practice of screening for disease. *Public health papers* **34** (1968).
33. Iatrakis, G. *et al.* Manosmed University's Risk factor calculator for female breast cancer: Preliminary data. *Rev. Clin. Pharmacol. Pharmacokinet. Int. Ed.* **32**, 23–27 (2018).

Declarations

Funding

This work was funded by the University of Montpellier and Nîmes University Hospital which supported Emmanuel Bonnet PhD thesis. A complementary funding was obtained from the Fondation de l'Avenir for the implementation of the telephone survey.

Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

Ethics approval

We obtained the authorization n°: 2200402v0 from the CNIL's (National Commission for Informatics and Liberties). We also received the agreement (n°: A0246039) of the ethics committee of the CPP (Personal Protection Committee). All methods were performed in accordance with the relevant guidelines and regulations.

Consent to participate

An explanatory leaflet (Supplementary material 3) was sent to all participants prior to the phone call and informed consent was obtained from all individual participants prior to the telephone interview.

Consent for publication

Not applicable

Availability of data and material

The data used and analyzed during the current study are available from the corresponding author (EB).

Code availability

The code used during the current study is available from the corresponding author (EB).

Authors' contributions

All authors contributed to the study conception and design. Material preparation and analysis were performed by E.B. The first draft of the manuscript was written by E.B. and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Figure legends

Figure 1: Flowchart of the retrospective cohort study.

Figure 2: Determination of the three risk groups with optimal thresholds in women at average risk of breast cancer: Lower risk, Average risk and Higher risk.

Tables

Table 1. Descriptive analysis

Breast cancer		No n (%)	Yes n (%)	P-value
Age (years)	[49,53[626 (25.2)	156 (26.1)	0.164
	[53,57[915 (36.9)	196 (32.8)	
	[57,60]	939 (37.9)	245 (41.0)	
Oral contraceptive use	No	605 (24.5)	165 (27.7)	0.111
	Yes	1867 (75.5)	430 (72.3)	
Contraceptive pill duration of use (years)	0	605 (24.8)	165 (28.5)	0.156
	[1,5[520 (21.3)	127 (21.9)	
	[5,10[327 (13.4)	80 (13.8)	
	≥ 10	986 (40.4)	207 (35.8)	
Number of breast biopsies	0	2182 (88.8)	506 (86.2)	0.042
	1	231 (9.4)	61 (10.4)	
	≥ 2	45 (1.8)	20 (3.4)	
Breast density (BI-RADS classification)	A	338 (13.6)	60 (10.1)	0.012
	B	1483 (59.8)	348 (58.3)	
	C	604 (24.4)	168 (28.1)	
	D	55 (2.2)	21 (3.5)	
Physical activity (hours/week)	0	972 (39.3)	225 (37.7)	0.093
	[1,5[1185 (47.9)	275 (46.1)	
	≥ 5	318 (12.8)	97 (16.2)	
Alcohol consumption (glasses/week)	0	1329 (53.6)	344 (57.7)	0.16
	[1,10[920 (37.1)	197 (33.1)	
	≥ 10	229 (9.2)	55 (9.2)	
Age at menarche (years)	<12	483 (20.1)	133 (23.1)	0.121
	≥ 12	1919 (79.9)	442 (76.9)	
Age at first live birth (years)	≤ 24	1361 (55.0)	334 (56.4)	0.226
	>24	911 (36.8)	200 (33.8)	
	nulliparous	201 (8.1)	58 (9.8)	
Number of children	0	245 (9.9)	72 (12.1)	0.068
	[1,3[1694 (68.3)	379 (63.5)	
	≥ 3	541 (21.8)	146 (24.5)	
Total duration of breastfeeding (months)	0	1383 (55.8)	343 (57.6)	0.026
	[1,7[739 (29.8)	149 (25.0)	
	≥ 7	355 (14.3)	104 (17.4)	
Overweight (BMI ≥ 25)	No	2120 (87.7)	488 (85.2)	0.122
	Yes	298 (12.3)	85 (14.8)	
Financial difficulties and/or not going on vacation	No	1518 (61.9)	323 (54.8)	0.002
	Yes	935 (38.1)	266 (45.2)	
Age at menopause (years)	≥ 41	2229 (94.1)	532 (95.3)	0.311
	< 41	139 (5.9)	26 (4.7)	
Mother history of breast cancer	No	2251 (92.4)	517 (87.9)	0.001
	Yes	184 (7.6)	71 (12.1)	
Sister history of breast cancer	No	2261 (91.6)	536 (90.1)	0.268
	Yes	207 (8.4)	59 (9.9)	
Daughter history of breast cancer	No	2451 (99.1)	588 (98.7)	0.499
	Yes	23 (0.9)	8 (1.3)	
Aunt history of breast cancer	No	2271 (93.9)	521 (91.2)	0.029
	Yes	148 (6.1)	50 (8.8)	

Table 2. Distribution of breast density according to the answer or not to the study questionnaire

	Breast density	Unreachable or refusal	Answers	P-value
No breast cancer	A	17.2%	13.7%	0.0007
	B	58.1%	59.8%	
	C	21.7%	24.3%	
	D	3.0%	2.2%	
Breast cancer	A	11.7%	10.5%	0.6035
	B	61.4%	58.1%	
	C	23.3%	28.0%	
	D	3.6%	3.4%	

Table 3. Distribution of age according to the answer or not to the study questionnaire

	Unreachable or refusal	Answers	P-value	
		mean (sd)		mean(sd)
No breast cancer	55 (3.2)	55.1 (3.2)	0.384	
Breast cancer	55.7 (3.1)	55.3 (3.3)	0.125	

Table 4. Cox model for the occurrence of breast cancer

		Hazard Ratio	95%CI	P-value
Age		1.024	0.997-1.05	0.084
Breast density	A	-	-	-
	B	1.411	1.05-1.9	0.023
	C	1.651	1.2-2.27	0.002
	D	2.112	1.25-3.58	0.006
Mother history of breast cancer	No	-	-	-
	Yes	1.612	1.24-2.09	< 0.001
Financial difficulties and/or not going on vacation	No	-	-	-
	Yes	1.299	1.09-1.55	0.003
Age at menopause	≥ 41	-	-	-
	< 41	0.364	0.13-0.99	0.047
Age at menarche	< 12	-	-	-
	≥ 12	0.774	0.63-0.95	0.015
Age at menopause < 41 and Age at menarche ≥ 12		2.73	0.91-8.19	0.073

HR for Age at menopause < 41 and Age at menarche ≥ 12 = 0.364 × 0.774 × 2.73 = 0.769

Figures

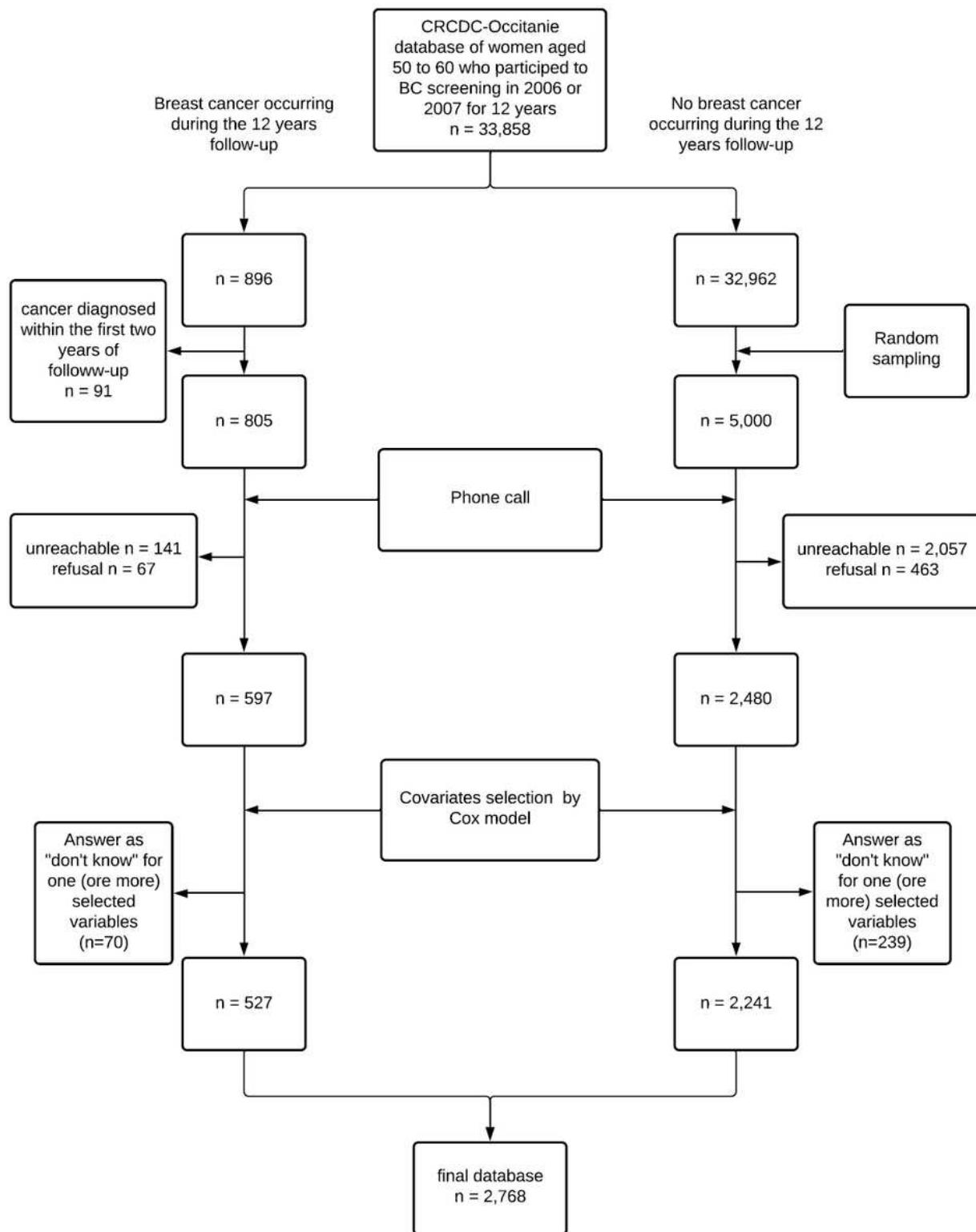


Figure 1

Flowchart of the retrospective cohort study.

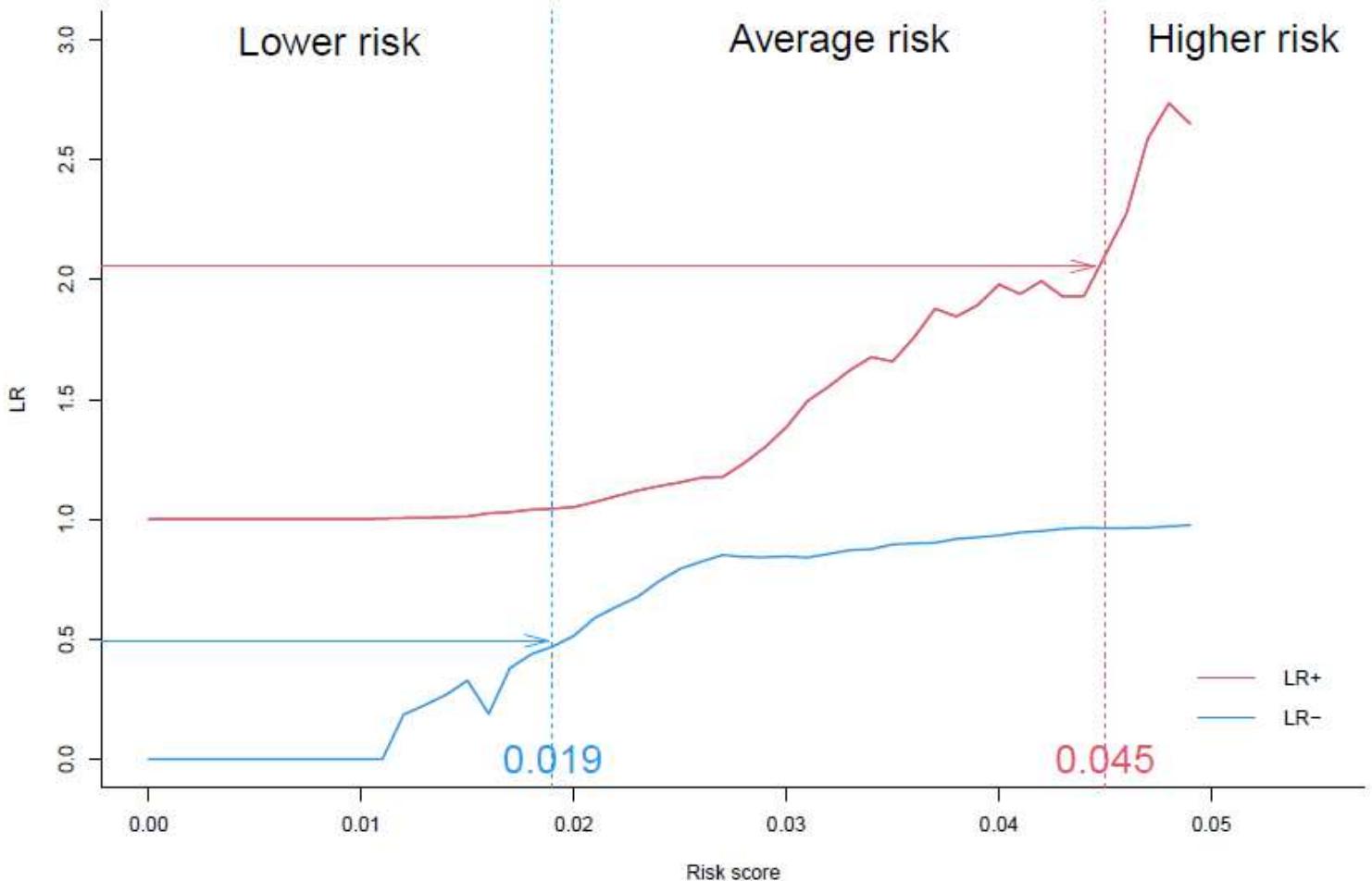


Figure 2

Determination of the three risk groups with optimal thresholds in women at average risk of breast cancer: Lower risk, Average risk and Higher risk.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial1.pdf](#)
- [Supplementarymaterial2.pdf](#)
- [Supplementarymaterial3.pdf](#)