

# AN INDIVIDUALIZED BREAST CANCER RISK PREDICTION MODEL TO PERSONALIZE MAMMOGRAPHIC SCREENING

Javier Louro<sup>1,2</sup>, Xavier Castells<sup>1,2</sup>, M Jesús Quintana<sup>3,4</sup>, Margarita Posso<sup>1,2</sup>, María Sala<sup>1,2</sup>, Marta Román<sup>1,2</sup>

1. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.

2. Research Network on Health Services in Chronic Diseases (REDISSEC), Barcelona, Spain.

3. Department of Clinical Epidemiology and Public Health, University Hospital de la Santa Creu i Sant Pau. Barcelona, Spain.

4. CIBER of Epidemiology and Public Health (CIBERESP), Spain.

## BACKGROUND

Several studies have proposed personalized strategies based on woman's individual breast cancer risk to improve the effectiveness of breast cancer screening. We developed and validated a risk prediction model to estimate the biennial risk of breast cancer for women targeted to breast cancer screening during the 20 year screening lifespan.

## METHODS

We conducted a retrospective cohort study of 121 969 screened at the Spanish Breast Cancer Screening Program between 1995 and 2015, and followed up until 2017. We used partly conditional Cox hazards regression to estimate the hazard ratios and individual risks in women with family history of breast cancer, previous benign breast disease, and previous mammographic features. As proposed by Zheng and Heagerty, we used a general hazard function to predict the risk of breast cancer diagnosis based on follow-up time, prediction time, and woman's risk profile (1). We calculated the expected-observed ratio and the AUROC using two different methods; the standard method (2) and the time-dependent AUROC proposed by Li et al (3).

## RESULTS

Table 1. Baseline characteristics of the study population

|                                 | No breast cancer<br>n = 119,911 | Breast cancer<br>n = 2,058 | p-value |
|---------------------------------|---------------------------------|----------------------------|---------|
| Mean follow-up                  | 7.6 years                       | 5.8 years                  | <0.05*  |
| Age                             |                                 |                            |         |
| 50-54 years old                 | 63,507 (52.96%)                 | 1,149 (55.83%)             | <0.05** |
| 55-59 years old                 | 25,738 (21.46%)                 | 542 (26.34%)               | <0.05** |
| 60-64 years old                 | 22,796 (19.01%)                 | 325 (15.79%)               | <0.05** |
| 65-69 years old                 | 7,870 (6.56%)                   | 42 (2.04%)                 | <0.05** |
| Family history of breast cancer |                                 |                            |         |
| No                              | 106,144 (88.52%)                | 1,734 (84.26%)             | <0.05** |
| Yes                             | 13,767 (11.48%)                 | 324 (15.74%)               | <0.05** |
| Previous benign breast disease  |                                 |                            |         |
| No                              | 97,312 (81.15%)                 | 1 565 (76.04%)             | <0.05** |
| Prior biopsy, unknown diagnosis | 20,069 (16.74%)                 | 410 (19.92%)               | <0.05** |
| Non proliferative               | 2,226 (1.86%)                   | 67 (3.26%)                 | <0.05** |
| Proliferative without atypia    | 263 (0.22%)                     | 14 (0.68%)                 | <0.05** |
| Proliferative with atypia       | 41 (0.03%)                      | 2 (0.10%)                  | <0.05** |
| Previous mammographic feature   |                                 |                            |         |
| No                              | 98,407 (82.07%)                 | 1,540 (74.83%)             | <0.05** |
| Mass                            | 14,015 (11.69%)                 | 296 (14.38%)               | <0.05** |
| Calcification                   | 1,944 (1.62%)                   | 88 (4.28%)                 | <0.05** |
| Asimmetry                       | 1,680 (1.40%)                   | 25 (1.21%)                 |         |
| Distortion                      | 701 (0.58%)                     | 11 (0.53%)                 |         |
| Multiple feature                | 3,164 (2.64%)                   | 98 (4.76%)                 |         |

\* Different at p < 0.05 in Mann-Whitney U test. \*\* Different at p < 0.05 in a two-sided test of equality for column proportions (z-test). Tests adjusted for all pairwise comparisons within each tumor characteristic using the Bonferroni correction.

Figure 1. Predicted absolute breast cancer risk estimates based on individual risk factors

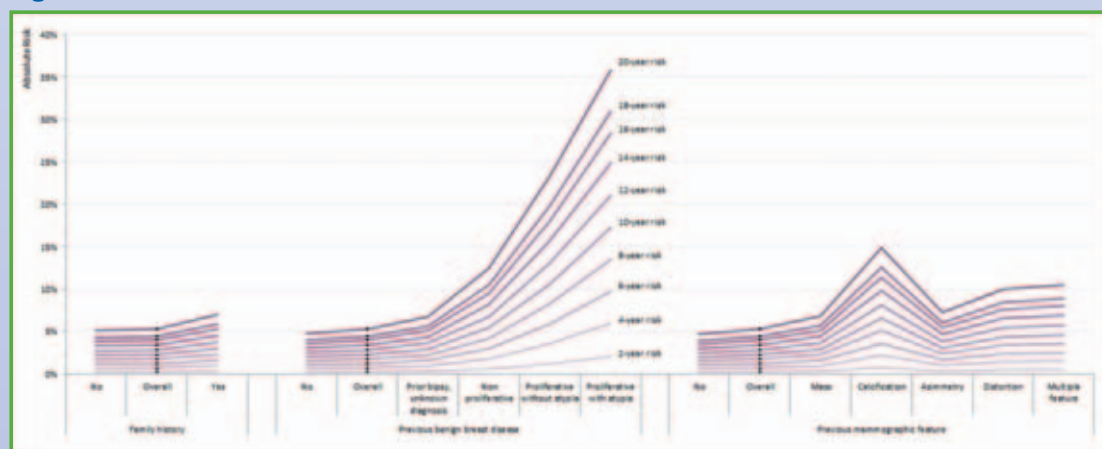


Table 2. Partly conditional cox model hazard ratios

|                                 | Women-year | Cases | aHR* (CI95%)      |
|---------------------------------|------------|-------|-------------------|
| Family history of breast cancer |            |       |                   |
| No                              | 478.147,2  | 1.009 | Ref               |
| Yes                             | 69.705,9   | 205   | 1.34 (1.16-1.56)  |
| Previous benign breast disease  |            |       |                   |
| No                              | 429.760,8  | 840   | Ref               |
| Prior biopsy, unknown diagnosis | 100.091,0  | 271   | 1.17 (1.02-1.35)  |
| Non proliferative               | 15.791,3   | 80    | 1.54 (1.20-1.98)  |
| Proliferative without atypia    | 1.940,3    | 19    | 2.40 (1.49-3.84)  |
| Proliferative with atypia       | 269,6      | 4     | 4.15 (1.63-10.59) |
| Previous mammographic feature   |            |       |                   |
| No                              | 415.112,8  | 738   | Ref               |
| Mass                            | 81.649,0   | 226   | 1.34 (1.14-1.57)  |
| Calcification                   | 12.171,4   | 87    | 3.19 (2.50-4.09)  |
| Asimmetry                       | 12.530,3   | 37    | 1.47 (1.06-2.05)  |
| Distortion                      | 4.648,1    | 19    | 1.96 (1.24-3.10)  |
| Multiple feature                | 21.741,6   | 107   | 2.29 (1.85-2.84)  |

\*aHR: Adjusted Hazard Ratio. Model adjusted by age, family history, previous benign breast disease and previous mammographic feature

Table 3. E/O ratio and AUROC of the model for every time horizon

|              | Observed events | E/O ratio (95%CI) | AUROC (95%CI)    | AUROCt (95%CI)   |
|--------------|-----------------|-------------------|------------------|------------------|
| 2-year risk  | 196             | 0.93 (0.81-1.07)  | 0.58 (0.55-0.62) | 0.60 (0.56-0.64) |
| 4-year risk  | 455             | 1.02 (0.93-1.12)  | 0.58 (0.55-0.60) | 0.58 (0.55-0.61) |
| 6-year risk  | 688             | 1.00 (0.93-1.08)  | 0.58 (0.56-0.61) | 0.57 (0.54-0.60) |
| 8-year risk  | 863             | 1.02 (0.96-1.10)  | 0.59 (0.57-0.62) | 0.58 (0.55-0.60) |
| 10-year risk | 1.000           | 1.03 (0.97-1.10)  | 0.61 (0.59-0.63) | 0.58 (0.56-0.60) |
| 12-year risk | 1.106           | 1.02 (0.97-1.09)  | 0.62 (0.61-0.64) | 0.58 (0.56-0.61) |
| 14-year risk | 1.171           | 1.03 (0.97-1.09)  | 0.65 (0.63-0.67) | 0.59 (0.57-0.61) |
| 16-year risk | 1.201           | 1.03 (0.98-1.09)  | 0.69 (0.67-0.71) | 0.59 (0.57-0.61) |
| 18-year risk | 1.210           | 1.05 (0.99-1.11)  | 0.73 (0.71-0.75) | 0.58 (0.56-0.60) |
| 20-year risk | 1.214           | 1.05 (0.99-1.12)  | 0.76 (0.73-0.79) | 0.58 (0.54-0.61) |

E/O ratio: Expected vs observed ratio, AUROC: Area under the receiver operating characteristic curve, AUROCt: Time-dependent area under the receiver operating characteristic curve, 95%CI: 95% Confidence Interval

## CONCLUSIONS

We developed a risk prediction model to estimate the short and long-term risk of breast cancer for woman targeted to breast cancer screening based on information of age, family history of breast cancer, previous benign breast disease, and previous mammographic features. Our model showed a good calibration, but its discriminative power was modest and should be improved by adding new variables. The model can be key to develop personalized strategies based on woman's individual breast cancer risk.

## REFERENCES

- (1) Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. 2005;61(2):379-91.
- (2) Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-87.
- (3) Li L, Greene T, Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Stat Methods Med Res*. 2018;27(8):2264-78.

Founding: This work was supported by Grants from Instituto de Salud Carlos III FEDER (PI17/00047).