Eur Radiol (2013) 23:2413–2419
DOI 10.1007/s00330-013-2836-8

BREAST

# Breast cancer risk prediction model: a nomogram based on common mammographic screening findings

**J. M. H. Timmers · A. L. M. Verbeek · J. IntHout · R. M. Pijnappel · M. J. M. Broeders · G. J. den Heeten**

Received: 20 December 2012 / Revised: 7 February 2013 / Accepted: 25 February 2013 / Published online: 18 April 2013
© European Society of Radiology 2013

**Abstract**
*Objectives* To develop a prediction model for breast cancer based on common mammographic findings on screening mammograms aiming to reduce reader variability in assigning BI-RADS.
*Methods* We retrospectively reviewed 352 positive screening mammograms of women participating in the Dutch screening programme (Nijmegen region, 2006–2008). The following mammographic findings were assessed by consensus reading of three expert radiologists: masses and mass density, calcifications, architectural distortion, focal asymmetry and mammographic density, and BI-RADS. Data on age, diagnostic workup and final diagnosis were collected from patient records. Multivariate logistic regression analyses were used to build a breast cancer prediction model, presented as a nomogram.
*Results* Breast cancer was diagnosed in 108 cases (31 %). The highest positive predictive value (PPV) was found for spiculated masses (96 %) and the lowest for well-defined masses (10 %). Characteristics included in the nomogram are age, mass, calcifications, architectural distortion and focal asymmetry.
*Conclusion* With our nomogram we developed a tool assisting screening radiologists in determining the chance of malignancy based on mammographic findings. We propose cutoff values for assigning BI-RADS in the Dutch programme based on our nomogram, which will need to be validated in future research. These values can easily be adapted for use in other screening programmes.
*Key points*
- *There is substantial reader variability in assigning BI-RADS in mammographic screening.*
- *There are no strict guidelines linking mammographic findings to BI-RADS categories.*
- *We developed a model (nomogram) predicting the presence of breast cancer.*
- *Our nomogram is based on common findings on positive screening mammograms.*
- *The nomogram aims to assist screening radiologists in assigning BI-RADS categories.*

**Keywords** Breast cancer · Mass screening · Mammography · Decision support techniques · Nomograms

J. M. H. Timmers (✉) · R. M. Pijnappel · M. J. M. Broeders · G. J. den Heeten
National Expert and Training Centre for Breast Cancer Screening, PO Box 6873, 6503 GJ Nijmegen, The Netherlands
e-mail: j.timmers@lrcb.nl

J. M. H. Timmers · A. L. M. Verbeek · J. IntHout · M. J. M. Broeders
Department for Health Evidence, Radboud University Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands

R. M. Pijnappel
Department of Radiology, University Medical Centre Utrecht, PO Box 85500, 3508 GA Utrecht, The Netherlands

G. J. den Heeten
Department of Radiology, Academic Medical Centre, University of Amsterdam, PO Box 22660, 1100 DD Amsterdam, The Netherlands

## Introduction

The Breast Imaging Reporting and Data System (BI-RADS) [1] lexicon does not provide strict screening guidelines as to which BI-RADS category should be assigned based on which mammographic findings. At training courses at the

Dutch National Expert and Training Centre for Breast Cancer Screening (NETCB), BI-RADS has been a continuous source of discussion among screening radiologists as well as during site visits with the radiologists of the NETCB site visit team. Improving knowledge on the positive predictive values of commonly found lesions would assist screening radiologists in assigning the appropriate BI-RADS category with the aim of reducing observer variability. The BI-RADS lexicon was introduced into the Dutch screening programme in 2010 together with the introduction of digital mammography. This lexicon, developed by the American College of Radiology (ACR), provides standardised mammographic reporting and breast imaging terminology, a report organisation and a classification system [1]. Studies have shown that BI-RADS in a clinical setting is useful in predicting the presence of breast cancer and improving the choice and efficiency of further diagnostic workup [2–4]. The positive predictive value (PPV) for calcifications [5] or other mammographic findings [6, 7] could help to predict the risk of breast cancer by using BI-RADS. However, these studies do not describe the breast lesions in detail and often represent a clinical setting including only patients who have undergone a biopsy procedure [8].

In this study we retrospectively assessed positive digital screening mammograms for specific mammographic findings and developed a model (nomogram) to predict the presence of breast cancer. We matched BI-RADS categories to the outcome of the nomogram and proposed cutoff values for assigning BI-RADS. We used data on follow-up from the well-documented Nijmegen screening programme as this information is not routinely collected in the national screening programme [9].

## Materials and methods

### Screening setting

The Nijmegen programme is part of the centrally organised Dutch screening programme, with annual nationwide participation of approximately 1 million women and 120 screening radiologists. Details of both the national programme and the Nijmegen programme have been described previously [9]. In short, women aged 50–75 are invited for a screening examination every 2 years. Two-view (mediolateral oblique and craniocaudal) mammography is used at initial screens; at subsequent screens, a mediolateral oblique view is standard and additional craniocaudal views are performed by the radiographer on indication. At present, an additional view is performed in 90 % of all subsequent screens. This policy is currently under revision with the aim of making this the standard procedure. The screening examinations are independently read by two qualified screening radiologists who must reach consensus about recall for further diagnostic assessment. If consensus is not achieved, a third reader decides. Digital mammography is introduced into the screening programme and was completed nationwide in 2010.

Diagnostic workup is not included in the Dutch screening programme. In the case of an abnormal screening mammogram, the woman is recalled to her general practitioner and sent to a hospital for further assessment. For our study, information on the diagnosis and workup of each individual woman were derived by means of chart review in collaboration with the regional hospitals.

### BI-RADS in the screening

BIRADS was introduced into the screening programme as a quality assessment and stratification tool in 2009–2010, at the same time as the introduction of digital mammography [10]. Its application was recommended in the national guideline “Breast Cancer” [11] and was made compulsory by the National Institute for Public Health and the Environment in 2008. The ACR guidelines [12] define a negative (normal) screening examination as one that is negative or has benign findings (BI-RADS categories 1 and 2) and a positive (abnormal) screening examination as one for which recall is indicated (BI-RADS categories 0, 4 or 5) [10]. BI-RADS 3 implies short-interval follow-up (6 months), which is not available in the Dutch screening setting. This category was therefore excluded at the nationwide introduction of the BI-RADS lexicon into the screening [10] as advised in the ACR 2003 [12] guidelines.

### Study population

Data of 417 women were reviewed. Included were women who participated in the Nijmegen screening programme and were recalled to their general practitioner and a hospital for further assessment on the basis of an abnormal screening mammogram (December 2006 to November 2008). All women consented to the use of their anonymous data for scientific research. Our database consisted of digital mammograms, including previous screening examinations.

The screening examinations that were included in our study were independently reviewed retrospectively by three expert screening radiologists who all had >10 years’ experience and read >5,000 mammograms a year. Consensus was reached afterwards between the expert radiologists on all recalled cases. The following mammographic findings (Table 1) were separately assessed for each abnormality or lesion found: laterality (left, right, left and right), calcifications (small, medium or large cluster), mass (ill-defined, well-defined, spiculated), mass density (low, fibroglandular and high), the presence of architectural distortion and focal asymmetry (presence/absence) and mammographic density.

**Table 1** Predictive values of mammographic findings in 352 recalled women

| BI-RADS classification and mammographic findings | Mammograms in category | Breast cancer | | | |
|---|---|---|---|---|---|
| | (*n*=352) | Absent *n* | Present *n* | PPV* % | 95 % CI % |
| ACR BI-RADS | | | | | |
| BI-RADS 0 | 120 | 109 | 11 | 9 | 4–13 |
| BI-RADS 4 | 198 | 134 | 64 | 32 | 25–39 |
| BI-RADS 5 | 34 | 1 | 33 | 97 | 91–100 |
| Mass | 139 | 81 | 59 | 42 | 34–50 |
| Well defined | 60 | 54 | 6 | 10 | 3–17 |
| Ill defined | 55 | 25 | 30 | 55 | 43–67 |
| Spiculated | 24 | 1 | 23 | 96 | 88–100 |
| Density of mass | | | | | |
| Low density | 13 | 9 | 4 | 29 | 17–41 |
| High density | 24 | 6 | 18 | 75 | 57–93 |
| Fibroglandular density | 102 | 65 | 37 | 36 | 27–45 |
| Cluster calcifications | 181 | 117 | 64 | 35 | 32–38 |
| Architectural distortion | 20 | 14 | 6 | 30 | 20–40 |
| Focal asymmetry | 48 | 43 | 5 | 10 | 1–19 |
| Mammographic density | | | | | |
| ACR 1 | 49 | 30 | 19 | 39 | 15–51 |
| ACR 2 | 161 | 110 | 51 | 32 | 23–39 |
| ACR 3 | 106 | 76 | 30 | 28 | 20–38 |
| ACR 4 | 36 | 28 | 8 | 22 | 9–36 |

**PPV* Positive predictive value

In addition, the BI-RADS used in the screening programme was assigned to all recalled cases (BI-RADS 0, 4 or 5).

Retrospectively, 64 women would not have been recalled by the expert panel. These women were excluded from our analysis because for our study we aimed to set a quality standard determined by the expert panel in this study. We only assessed primary breast cancer; therefore, one woman was excluded because the lesion appeared to be a metastasis of lung cancer. In one participant, two lesions were found in one breast; we only regarded the most suspicious finding as the reason for recall. In total, 352 screening mammograms were analysed in this study.

## Data analysis

A true-positive (TP) final diagnosis is defined as a screen detected carcinoma found and confirmed by biopsy within a year of recall in the screening programme. False-positive diagnoses are defined as positive screening examinations that prove to be benign in the assessment process (using either biopsy or additional imaging procedures) [13]. PPVs and 95 % CIs were calculated for mass and mass density, clusters of calcifications, focal asymmetry, architectural distortion, mammographic density and BI-RADS. To calculate PPVs for the mammographic findings and the BI-RADS categories, we used the number of TP diagnoses among the total number of recalls per category × 100 %.

Multivariate logistic regression with the stepwise backward selection procedure was used to build the model (removal criterion $P>0.10$). Diagnosis of breast cancer was the dependent variable. The following predictors were examined: age (50–54 versus 55–74 years old, dichotomised as a proxy for menopausal status), mass density (no, fibroglandular, low and high mass density), calcification (clusters/no clusters), mass (no mass, well-defined, ill-defined or spiculated mass), focal asymmetry (yes or no), architectural distortion (yes or no) and mammographic density (ACR 1, 2, 3 or 4). Odds ratios (OR) for the predictors including 95 % CIs were calculated. Data analysis was performed using the statistical packages of R (version 2.11.1) [14] and SPSS (version 20; SPSS, Chicago, IL, USA). The area under the receiver-operating characteristics (ROC) curve (AUC) was calculated to determine the accuracy of the model's predictive value. ROC areas of 0.6–0.7, 0.7–0.8, 0.8–0.9 and 0.9–1 represent a poor, fair, good and excellent discriminative model respectively [15, 16]. The model was internally validated by means of the bootstrapping method. The selection and estimation procedure was repeated 500 times.

## Results

The mean age of all participants was 62 (ranging from 53 to 75) with a standard deviation of 7. Breast cancer was diagnosed in 108 of the 352 recalled women (true-positive recalls). The overall recall rate of the Nijmegen screening programme was 2.8 per 1,000. The overall positive predictive value of recall was 31 % (95 % CI 26–36 %).

### Positive predictive values

The mammographic findings are presented in Table 1. Overall, clusters of calcifications ($n$=181) and masses ($n$=139) were the most commonly recorded findings. The highest PPVs were reported for spiculated masses and high mass density (96 % and 75 % respectively, Table 2). Intermediate PPVs were demonstrated for ill-defined masses, architectural distortion, masses with low and fibroglandular densities (55 %, 30 %, 29 % and 36 % respectively). Low PPVs were reported for well-defined masses (10 %) and focal asymmetry (10 %). The PPVs for ACR density 1, 2, 3 and 4 were 39 %, 32 %, 28 % and 22 % respectively. The PPVs of the BI-RADS categories 0, 4 and 5 were 9 %, 32 % and 97 % respectively.

### Development of the nomogram

Multivariate logistic regression analysis showed that the relevant predictors were age ≥55 (OR=1.8), well-defined mass (OR=2.4), ill-defined mass (OR=18.8), spiculated mass (OR=373), calcifications (OR=10) and architectural distortion (OR=3.9); see Table 2. Even though it was not significant, we selected focal asymmetry (OR 2.1) as it was considered to be of clinical importance based on expert opinion and the literature [17–19]. Mammographic density was excluded because the variable did not improve the model according to the Akaike information criterion (AIC) [20] and mass density was excluded because it was too strongly correlated to mass. For simplification, the model calcification was dichotomised (cluster versus no cluster) as the model did not improve with the four calcification categories. The ROC curve of the multivariate logistic regression model is shown in Fig. 1. The area under the ROC curve is classified as fair (0.78; 95 % CI 0.72–0.83). The AUC was 0.75 after correction for optimism. The predicted probability of breast cancer was satisfactory with some slight deviations from the perfect calibration line (overestimation) around probabilities of 0.5.

With the aim of providing a practical tool for screening radiologists, we then developed a nomogram based on this model (Fig. 2). As an example, using the points scale in Fig. 1, a woman aged 60 (9 points) has an ill-defined mass (49 points) and a cluster with calcifications (39 points), and therefore scores a total of 97. The lower scale of the nomogram shows that this score corresponds to a probability of breast cancer of about 0.85 (85 %).

**Table 2** Multivariate logistic regression analyses to predict probability of breast cancer

| Predictor | Multivariate logistic regression analysis | |
|---|---|---|
| | Odds ratio | 95 % CI |
| Age, ≥ 55 years | 1.8 | 0.9–3.2 |
| Mass | | |
| Well-defined | 2.4 | 0.7–9.8 |
| Ill-defined | 18.8 | 5.8–60 |
| Spiculated | 373 | 38–3679 |
| Calcifications | 10 | 3–31 |
| Focal asymmetry | 2.1 | 0.6–7.6 |
| Architectural distortion | 3.9 | 1.1–14.2 |

## Discussion

We developed and internally validated a nomogram that estimates the individual risk of breast cancer based on age and common mammographic findings in screening examinations. To our knowledge, this is the first study that offers a tool that helps screening radiologists determine the chance of malignancy based on mammographic findings and,

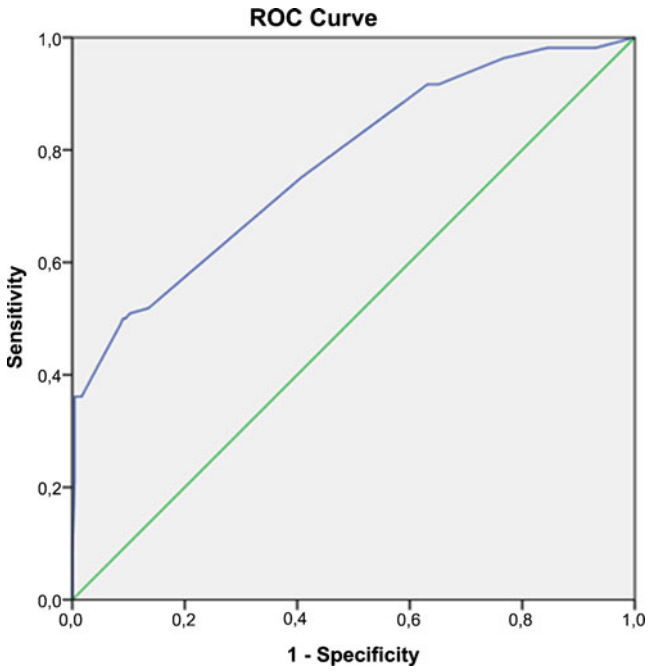


**Fig. 1** Receiver-operating characteristic (ROC) curve of the multivariate logistic model predicting the probablilty of breast cancer using common mammographic findings. AUC=0.78 (95 % CI 0.72–0.83)

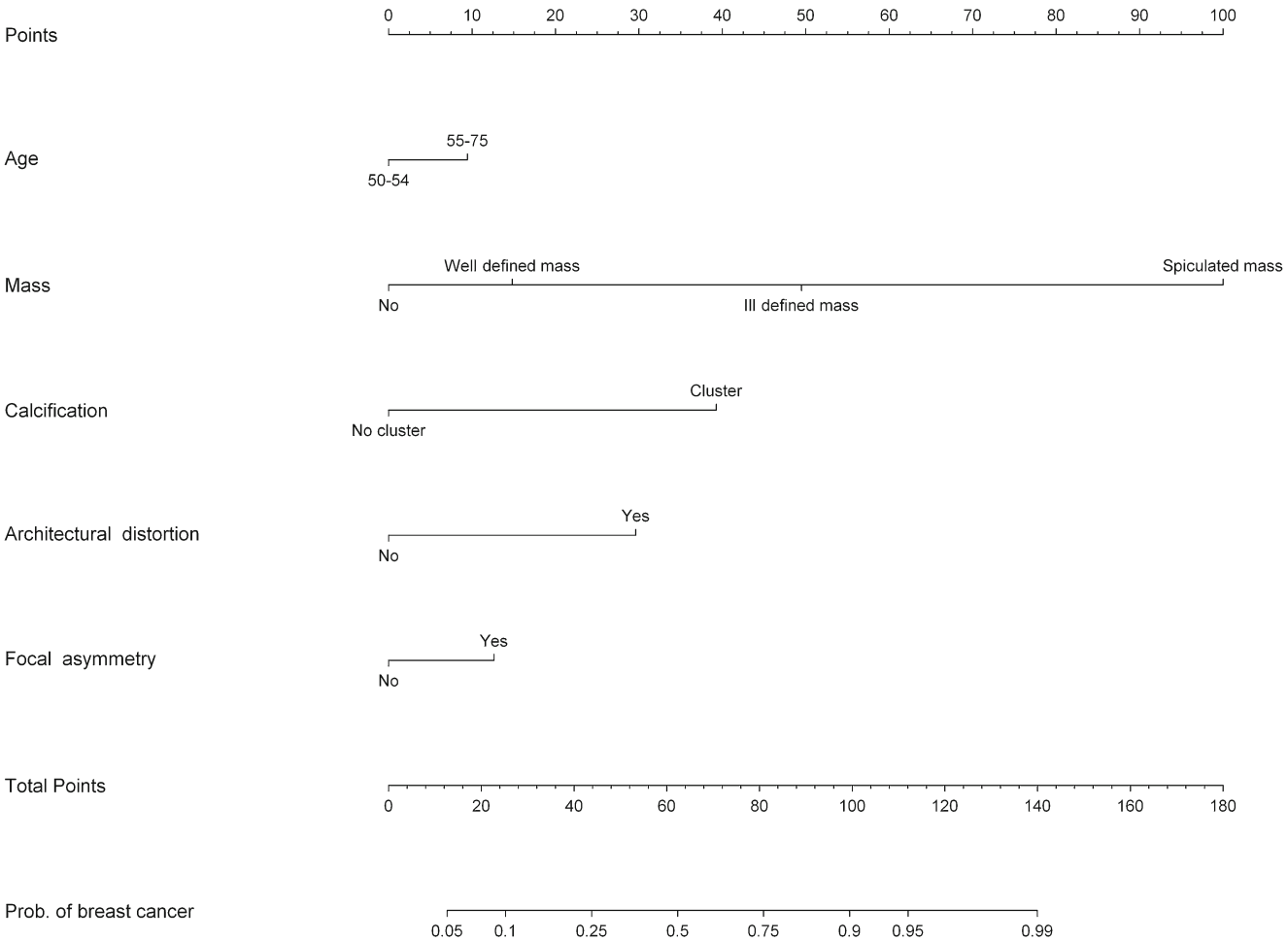


**Fig. 2** Nomogram representing the probability of breast cancer. This can be determined by drawing a vertical line connecting the value of each variable with the point score at the top of the diagram. The scores for each variable are then summed to give a total points score, which is plotted along the "total points" line at the bottom of the nomogram. This line is projected to the probability of breast cancer of the mammographic findings

accordingly, offers the ability to assign the correct BI-RADS to the screening examination.

Currently, the BI-RADS score is often assigned based on intuition rather than on strictly defined guidelines, which results in substantial reader variability, both inter- and intraobserver variation [21, 22]. Our nomogram aims to reduce this variability. The BI-RADS lexicon is a recurring topic at the continuous medical education of screening radiologists provided by the NETCB. The NETCB proposes guiding principles in assigning BI-RADS scores during these sessions (for instance, a BI-RADS 4 should be assigned to calcifications). The results of our model are concordant with the proposed guiding principles. Our nomogram is therefore a valuable addition to these principles. BI-RADS after screening mammography in the Netherlands is important as diagnostic workup occurs in hospitals and is not necessarily carried out by the screening radiologists. Based on our findings, we propose the following cutoff values in our nomogram for assigning BI-RADS 0, 4 and 5: 0–20 %, 20–80 % and 80–100 % respectively (Fig. 3). As expected, there is some overlap between the BI-RADS categories. However, the cutoff values are in line with our experience of using BI-RADS in our centrally organised screening programme [10].

Nomograms have been developed and used by clinicians and patients in the field of breast cancer before, as they provide easily understood outcome probabilities [23, 24]. Mazouni et al. [8] developed a nomogram based on epidemiological and clinical information to predict the probability of breast cancer in radiological lesions classified as BI-RADS category 4. Their aim was to provide a diagnostic tool that could guide the decision-making regarding the need for a biopsy. However, they used features not available to screening radiologists in the screening setting, for example the use of hormone replacement therapy. Their nomogram serves a different purpose and cannot be used in the screening situation.

Our results are comparable to those of other studies that also found masses and calcifications (in 66 % and 25 % of all cases) to be the most predominant features with the

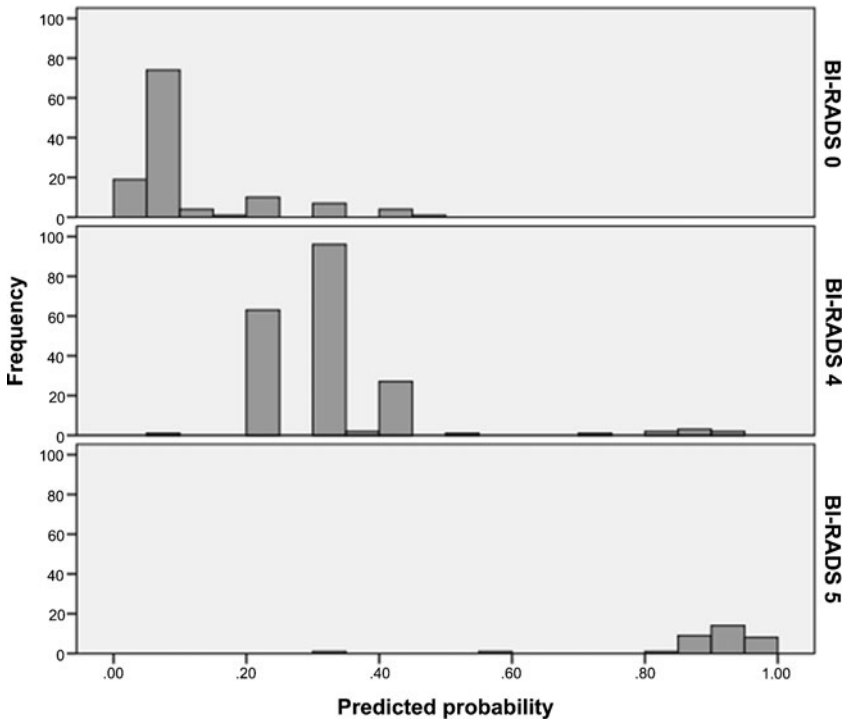


**Fig. 3** Predicted probability for breast cancer from the prediction model connected to the BI-RADS categories as assigned by our expert panel

highest PPVs (9 % for masses and 12 % for calcifications) [6, 17]. The PPVs of BIRADS (9 %, 32 % and 97 % for BI-RADS 0, 4 and 5 respectively) are in line with other Dutch studies [10]. The PPV we found for well-defined masses (10 %) is higher compared to other studies [6, 7]. This can be explained by the fact that multiple well-defined masses often seen in fibro-cystic disease as well as oval structures resembling lymph nodes are not recalled in our programme with a low recall rate of 2.1 % [25]. Our subset of well-defined masses therefore consists of (mostly new) real round solitary masses with a higher density.

We have noted a number of limitations in this study. First, the data set of recalled women is from only one screening practice in the Netherlands. However, the recalled abnormalities were comparable to those found in other screening practices [17]. Second, we aimed to set a quality standard that in this study is set by an expert panel of Dutch radiologists and thus represents a typical Dutch screening setting with a low recall rate. Sixty-four cases were excluded from the analysis because the expert panel would not have recalled these women after the initial screening. After diagnostic workup, no cancers were found in these women. However, if these cases had been included in the database, they would have contributed to the recall rate, making it comparable to those of other European programmes and daily practice. This did not influence the nomogram as no mammographic findings were assigned to these cases by the expert panel.

The cutoff points for the proposed BI-RADS categories can easily be adapted for use in other screening practices. Our model can be used by other screening programmes regardless of the differences between these programmes in, for instance, recall rate. We have recently implemented our tool in the education and quality assurance programme of the Dutch screening programme. Whether it will significantly reduce observer variability needs to be further investigated. We are currently investigating the options for the external validation of our nomogram.

**Acknowledgments** We would like to thank the screening radiologists in our expert panel for taking the time to review all women in our database. We also wish to acknowledge Paul van de Looi for his valuable assistance in the data collection.

## References

1. American College of Radiology (2008) ACR practice guideline for the performance of screening and diagnostic mammography (2008); American College of Radiology, editor. Preston (Va)
2. Zonderland HM, Pope TL Jr, Nieborg AJ (2004) The positive predictive value of the breast imaging reporting and data system (BI-RADS) as a method of quality assessment in breast imaging in a hospital population. EurRadiology 14:1743–1750
3. Orel SG, Kay N, Reynolds C, Sullivan DC (1999) BI-RADS categorization as a predictor of malignancy. Radiology 211:845–850
4. Liberman L, Abramson AF, Squires FB, Glassman JR et al (1998) The breast imaging reporting and data system: positive predictive value of mammographic features and final assessment categories. AJR AmJRoentgenol 171:35–40
5. Shin HJ, Kim HH, Ko MS, Kim HJ et al (2010) BI-RADS descriptors for mammographically detected microcalcifications verified by histopathology after needle-localized open breast biopsy. AJR Am J Roentgenol 195:1466–1471

6. Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith-Bindman R (2009) Positive predictive value of specific mammographic findings according to reader and patient variables. Radiology 250:648–657
7. Burrell HC, Pinder SE, Wilson AR, Evans AJ et al (1996) The positive predictive value of mammographic signs: a review of 425 non-palpable breast lesions. Clin Radiol 51:277–281
8. Mazouni C, Sneige N, Rouzier R, Balleyguier C et al (2010) A nomogram to predict for malignant diagnosis of BI-RADS Category 4 breast lesions. J Surg Oncol 102:220–224
9. Holland R, Rijken HJ, Hendriks JH (2007) The Dutch population-based mammography screening: 30-years experience. Breast Care 2:12–18
10. Timmers JM, van Doorne-Nagtegaal HJ, Zonderland HM, van Tinteren H et al (2012) The Breast Imaging Reporting and Data System (BI-RADS) in the Dutch breast cancer screening programme: its role as an assessment and stratification tool. Eur Radiol 22:5
11. Nationaal Borstkanker Overleg Nederland (2008) National Guideline "Breast Cancer". Amsterdam
12. American College of Radiology (2003) ACR BI-RADS®-Mammography: Breast Imaging Reporting and Data System, Breast Imaging Atlas; American College of Radiology, editor. Preston (Va)
13. National Evaluation Team for Breast cancer screening (2009) National evaluation of breast cancer screening in the Netherlands, 1990–2007. NETB, Rotterdam
14. R. Foundation for Statistical Computing (2008) R: A language and environment for statistical computing
15. Kleinbaum D (2010) Logistic Regression: a self-learning text. Springer, New York
16. Sardanelli F, de Leo G (2009) Biostatistics for radiologists. 1st edn. Springer Milan, Berlin, Heidelberg, New York
17. James JJ, Gilbert FJ, Wallis MG, Gillan MG, Astley SM et al (2010) Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicenter prospective trial of computer-aided detection: CADET II. Radiology 256:379–386
18. Khan KS, Chien PF, Dwarakanath LS (1999) Logistic regression models in obstetrics and gynecology literature. Obstet Gynecol 93:1014–1020
19. Onega T, Smith M, Miglioretti DL, Carney PA, Geller BA et al (2012) Radiologist agreement for mammographic recall by case difficulty and finding type. J Am Coll Radiol 9:788–794
20. Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723
21. Redondo A, Comas M, Macia F, Ferrer F, Murta-Nascimento C, et al. (2012) Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol. 2012 Sep 19
22. Timmers JM, van Doorne-Nagtegaal HJ, Verbeek AL, den Heeten GJ, Broeders MJ (2012) A dedicated BI-RADS training programme: effect on the inter-observer variation among screening radiologists. Eur J Radiol 81:2184–2188
23. Marko NF, Xu Z, Gao T, Kattan MW, Weil RJ (2011) Predicting survival in women with breast cancer and brain metastasis: a nomogram outperforms current survival prediction models. Cancer. 2011 Dec 16
24. Shin HC, Han W, Moon HG, Cho N, Moon WK, et al. (2012) Nomogram for predicting positive resection margins after breast-conserving surgery. Breast Cancer Res Treat. 2012 Jun 13
25. National Evaluation Team for Breast Cancer screening (2013) Preliminary results breast cancer creening programme in the Netherlands 2011. NETB, Rotterdam