The Breast 24 (2015) 680–686

Contents lists available at ScienceDirect

The Breast

journal homepage: www.elsevier.com/brst

Original article

Factors affecting recall rate and false positive fraction in breast cancer screening with breast tomosynthesis – A statistical approach

Aldana Rosso [a, *], Kristina Lång [c], Ingemar F. Petersson [a, b], Sophia Zackrisson [c]

[a] Epidemiology and Register Centre South, Skåne University Hospital, Lund, Sweden
[b] Orthopaedics, Department of Clinical Sciences Lund, Lund University, Lund, Sweden
[c] Department of Diagnostic Radiology, Translational Medicine Malmö, Lund University, Malmö, Sweden

ARTICLE INFO

Article history:
Received 24 July 2015
Received in revised form
19 August 2015
Accepted 21 August 2015
Available online 11 September 2015

Keywords:
Digital mammography
Breast cancer screening
Digital breast tomosynthesis
Binary marginal generalized linear models
Classification and regression trees
Generalized estimating equations

ABSTRACT

In this study, we investigate which factors affect the false positive fraction (FPF) for digital breast tomosynthesis (DBT) compared to digital mammography (DM) in a screening population by using classification and regression trees (C&RT) and binary marginal generalized linear models.

The data was obtained from the Malmö Breast Tomosynthesis Screening Trial, which aimed to compare the performance of DBT to DM in breast cancer screening. By using data from the first half of the study population (7500 women), a tree with the recall probability for different groups was calculated. The effect of age and breast density on the FPF was estimated using a binary marginal generalized linear model.

Our results show that breast density and breast cancer were the main factors influencing recall. The FPF is mainly affected by breast density and increases with breast density for DBT and DM.

In conclusion, the results obtained with C&RT are easy to interpret and similar to those obtained using binary marginal generalized linear models. The FPF is approximately 40% higher for DBT compared to DM for all breast density categories.

Introduction

Breast cancer screening programs are believed to improve the early detection of breast cancer and thus they may help to reduce breast cancer mortality [1]. However there are negative aspects associated with screening, such as overdiagnosis and false positive cases [1]. Digital mammography (DM) is the standard technique for breast cancer screening. However, it has limitations due to the fact that DM is a two dimensional technique that depicts a three dimensional organ. Hence, cancer detection can be hampered due to overlapping tissue in the images. Laming et al. [2] has estimated that around 15%–30% of cancer cases may not be detected when screening with DM. Digital breast tomosynthesis (DBT) is a three-dimensional imaging technique that may address some of the limitations that DM has, in particular problems related to overlapping tissue. Several recent studies have shown that the combination of DBT and DM improves the cancer detection rate [3–9].

The Malmö Breast Tomosynthesis Screening Trial (MBTST) was designed to compare the performance of one-view DBT as a single screening modality to two-view DM. The study population consisted of a random sample of 15 000 women invited to participate in the breast cancer screening program in the city of Malmö, Sweden. Women accepting to participate in the study were offered a DBT examination in addition to the DM examination at the screening visit. The first results of the screening trial, obtained after half of the study population was enrolled, were recently presented by Lång et al. [10]. The cancer detection rate for DBT was superior to that for DM, and that the overall recall rate for DBT was higher than that for DM [10].

One of the main concerns of breast cancer screening programs is the significant amount of healthy women that are recalled for further examination and then found free of breast cancer (false positive screening) [1,6,11]. It has been calculated that the cumulative risk of a false-positive screening result in women aged 50–69 undergoing 10 biennial screening tests is around 20% [6]. The purpose of this article is to quantify the probability of a false positive screening using the MBTST data for the first half of the study population using different statistical methods. The probability of a false positive screening is also called false positive

* Corresponding author. Klinikgatan 22, Skåne University Hospital, SE-221 85 Lund, Sweden. Tel.: +46 70 21 27 140.
E-mail address: aldana.rosso@skane.se (A. Rosso).

http://dx.doi.org/10.1016/j.breast.2015.08.007
0960-9776/

fraction. In some context it is also referred as false positive rate. We will use interchangeably the terms false positive fraction, probability of a false positive screening as well as recall probability and recall rate.

Binary marginal generalized *l*inear *m*odels (GLM) can be used to estimate how different factors would affect the recall probability for groups of women that share similar characteristics such as breast density and age. A more recently developed non-parametric tool suitable for this type of problems is called Classification and Regression Tree (C&RT) [12]. This technique is employed in clinical research with the aim to obtain a simple pattern to classify subjects between ill and healthy, and to get information about which groups of individuals could benefit more from targeted interventions [13–17]. One of the main advantages of C&RT is that the result of the analysis is a classification tree, which is easier to interpret in clinical practice [13]. However, due to the hierarchical nature of C&RT, it is not possible to estimate the effect of a single variable on the probability of recall. Therefore, we complemented the results obtained with C&RT with regression analysis. We applied C&RT to study which characteristics the recalled women have in common for both imaging methods and to present this information with a classification tree. In order to further analyse how these factors affect the probability of false positive screening we used a binary marginal GLM [18].

Materials and methods

Study population and image reading

The MBTST was a clinical trial performed at the Mammographic Clinic at the Skåne University Hospital, in the city of Malmö (Clinical Trial number NCT01091545). The study was approved by the Regional Ethical Review Board at Lund University (Dnr 2009/770) and the local Radiation Safety Board at the Skåne University Hospital in Malmö. Participating women gave written informed consent. The main characteristics of the study are discussed here. A thorough description of the study and the evaluation of the results from the analysis of the first half of the study population are presented elsewhere [10].

The Swedish Board of Health and Welfare recommends breast cancer screening with DM for women aged 40–74 at 18–24 month intervals [19]. The participants of the MBTST were randomly selected from the screening population in Malmö. The women accepting to take part in the study were offered a DBT examination in addition to the DM examination at the screening visit.

Six readers with at least 8 years of breast imaging experience participated in the study. The readers had experience of DBT reading from previous studies [20–22]. Two blinded readers evaluated the DBT reading sequence independently from the two blinded readers of the DM reading sequence. The DBT sequence of images consisted of an initial presentation of a one-view DBT alone, followed by the addition of a one-view DM and finally previous two-view DM was shown if available. The DM sequence consisted of a two-view DM and then an addition of a prior two-view DM if available. The images were evaluated and scored at each step before moving to the next step according to a 5-point scale: 1. normal, 2. benign findings, 3. non-specific finding with low probability of malignancy, 4. findings suspicious of malignancy, 5. findings highly suspicious of malignancy.

If any of the readers at any step of a sequence scored at least 3 points for the case, it was discussed at an arbitration meeting, where at least two readers re-evaluated the images and decided whether to recall the woman or not [10]. Furthermore, a woman could be recalled if she reported symptoms from the breasts at the examination in spite of negative image findings.

Recalled women were assessed in accordance with ordinary screening routine [10]. The cancer cases were verified with record linkage with the South Swedish Cancer Register. For all women in the study there was at least one-year follow-up.

The breast density was also evaluated at the final step of the DM reading sequence using the 4th edition of the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) scale for breast composition [23]: 1. The breast is almost entirely fat, 2. There are scattered fibroglandular densities, 3. The breast tissue is heterogeneously dense, 4. The breast tissue is extremely dense.

The first 7500 women participating in the trial were examined in January 2010–December 2012. In this population, 352 women were recalled for further examination (282 recalled in the DBT sequence and 197 in the DM sequence) [10]. The total number of screening detected cancer cases was 68 (67 cases detected in the BT sequence and 47 in the DM sequence) [10]. In this sample, 6640 women had a density evaluation. Those without density evaluation were not included in this analysis. The group of women without density evaluation had similar age distribution to the studied population and were neither recalled nor had cancer. The population characteristics were discussed in a previous publication [10]. The most important parameters of the sample for the analysis are listed in Table 1.

Classification and regression tree

Classification and Regression Tree is a non-parametric technique that splits the data into different groups by searching which variables separate the data the most with respect to the response variable [13]. The separations performed in C&RT are binary. A brief introduction to this method is presented in the Appendix A. The aim of the analysis was to provide a clear visualization of which groups of women were recalled in the DM and DBT reading sequences. Furthermore, the tree also provided an estimate of the predicted probability of recall for the different groups.

The analysis was performed using The Salford Predictive Modeller Software Suite, version 7. The Gini impurity index was used as splitting criteria and the obtained trees were validated using 10-fold cross validation (see Appendix A for further discussion). The response variable was whether the woman was recalled or not. We calculated separate trees for the DM and DBT reading sequences. The variables included in the models were breast density, cancer status (whether the women had breast cancer or not) and age at examination. The variables breast density and cancer status were included as categorical variables and the variable age at

Table 1
Main characteristics of the study population [10]. All women in the study had at least one-year follow-up.

Total number of women		6640
Number of recalled women	Total	352
	Recalled in DBT reading sequence	282
	Recalled in DM reading sequence	197
Cancer cases	Total	68
	Detected in DM reading sequence	47
	Detected in DBT reading sequence	67
Age	Median	54.3
	Min	39.7
	Max	75.9
	38–49	35.0%
	50–59	28.2%
	60–76	36.8%
Breast Density (BI-RADS)	1	19.8%
	2	37.8%
	3	34.0%
	4	8.5%