European Journal of Radiology 83 (2014) 2019–2023

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Single reading with computer-aided detection performed by selected radiologists in a breast cancer screening program

Xavier Bargalló[a,*], Gorane Santamaría[a], Montse del Amo[a], Pedro Arguis[a], José Ríos[c], Jaume Grau[b], Marta Burrel[a], Enrique Cores[a], Martín Velasco[a]

[a] Department of Radiology (CDIC), Hospital Clínic de Barcelona, C/ Villarroel, 170, 08036 Barcelona, Spain
[b] Preventive Medicine and Epidemiology Unit, Hospital Clínic de Barcelona, C/ Villarroel, 170, 08036 Barcelona, Spain
[c] Biostatistics and Data Management Core Facility, IDIBAPS, (Hospital Clinic) C/ Mallorca, 183. Floor -1. Office #60. 08036 Barcelona, Spain

ARTICLE INFO

Article history:
Received 11 July 2014
Accepted 13 August 2014

Keywords:
Mammography
Breast screening
Breast neoplasms
Computer-assisted diagnosis
Breast

ABSTRACT

Objectives: To assess the impact of shifting from a standard double reading plus arbitration protocol to a single reading by experienced radiologists assisted by computer-aided detection (CAD) in a breast cancer screening program.
Methods: This was a prospective study approved by the ethics committee. Data from 21,321 consecutive screening mammograms in incident rounds (2010–2012) were read following a single reading plus CAD protocol and compared with data from 47,462 consecutive screening mammograms in incident rounds (2004–2010) that were interpreted following a double reading plus arbitration protocol. For the single reading, radiologists were selected on the basis of the appraisement of their previous performance.
Results: Period 2010–2012 vs. period 2004–2010: Cancer detection rate (CDR): 6.1‰ (95% confidence interval: 5.1–7.2) vs. 5.25‰; Recall rate (RR): 7.02% (95% confidence interval: 6.7–7.4) vs. 7.24% (selected readers before arbitration) and vs. 3.94 (all readers after arbitration); Predictive positive value of recall: 8.69% vs. 13.32%. Average size of invasive cancers: 14.6 ± 9.5 mm vs. 14.3 ± 9.5 mm. Stage: 0 (22.3/26.1%); I (59.2/50.8%); II (19.2/17.1%); III (3.1/3.3%); IV (0/1.9%). Specialized breast radiologists performed better than general radiologists.
Conclusions: The cancer detection rate of the screening program improved using a single reading protocol by experienced radiologists assisted by CAD, at the cost of a moderate increase of the recall rate mainly related to the lack of arbitration.

1. Introduction

Screening mammography reduces breast cancer mortality among invited women by 25% in incidence-based mortality studies and 31% in case-control studies (38% and 48%, respectively, among screened women) [1]. Although mammography is the most commonly used screening tool to detect breast cancer, its sensitivity is estimated at about 70–85% [2].

Double reading with arbitration is the standard procedure in most regional screening programs in Europe because it has been shown to increase the cancer detection rate (CDR) by 5–15% [3,4]. In general, an experienced breast radiologist is paired with a general diagnostic radiologist or even with a radiographer in some countries. The European guidelines for breast cancer screening programs recommend double reading, even though they also state that "double reading is recommended in centralized programs for the first screening round and until the performance of the radiologists can be fully assessed" [5].

On the other hand, computer-aided detection (CAD) systems are becoming widely used [6]. Several studies in the literature show that CAD increases the detection rates by up to 20% [7,8]. The reported sensitivity of CAD in full-field digital mammography (FFDM) varies from 78 to 96% [9–11]. Gilbert et al. [12] demonstrated in a large study of equivalence that single reading with CAD could be an alternative to double reading.

Taking into account that European guidelines allow a single reading once the radiologist's performance has been fully assessed, we undertook this prospective study, in which mammograms of incident rounds were read by only one experienced radiologist assisted by CAD. The purpose of the study was to assess the impact of shifting from a standard double reading plus arbitration protocol to a new reading protocol consisting of a single reading with CAD done by radiologists selected for their good historical performance.

* Corresponding author. Tel.: +34 932275400; fax: +34 932275454.
E-mail address: xbarga@clinic.cat (X. Bargalló).

http://dx.doi.org/10.1016/j.ejrad.2014.08.010
0720-048X/

2. Materials and methods

2.1. Background

Our population-based breast cancer screening program started in May 2002 and involved the areas of Sants-Montjuic, Les Corts, and Eixample Esquerre in the city of Barcelona. The target population included women between 50- and 69-years-old who were invited to participate every 2 years. From 8 May 2002 to 16 November 2010, 85,304 mammograms were performed.

Following the recommendations of European guidelines [5], we performed a double reading plus arbitration of discordant readings by a third radiologist who was the most experienced of the team.

2.2. Study design

This was a prospectively designed study approved by the ethics committee. Informed consent was not required as the ethics committee asked for a surveillance protocol to be set up to ensure that the CDR did not drop to unacceptable levels and that acceptable recall rates (RR) were maintained.

From 17 November 2010 to 28 February 2013, 21,321 consecutive screening mammograms from women of the incident rounds were interpreted. All women from incident rounds had been screened at least once before in our program. The performance of the screening program was evaluated using a reading scheme based on a single reading plus CAD performed by radiologists selected according to their previous performance. There was a short period of 1 month to allow the readers to become familiar with CAD. Data from the prospective study were compared with those from 16 June 2004 (the date at which second round began) to 17 November 2010. In this period, 47,462 consecutive screening mammograms from incident rounds were performed. These mammograms followed a process of double reading with arbitration.

The following variables were collected: CDR; RR; biopsy rate; positive predictive value (PPV) of the recall; histological type; and stage of cancer detected.

2.3. Radiologist selection

Annual clinical audits were routinely performed. The performance of radiologists participating in the program was determined by collecting their historical results. The best radiologists were selected according to the CDR and RR.

The CDR in Catalonia in 2008–2009 was 3.91% (3.81% in incident rounds), with an RR of 4.5% [13]. In the city of Barcelona, the CDR in 2009 was 4.9% (4.4% in incident rounds) [14]. Based on these demographic data, we decided to set the inferior limit of CDR to 4.4% and the RR between 5 and 10%. Although European guidelines recommend that the RR be less than 5% in incident rounds, we decided to increase this limit to prioritize the detection as, in general, radiologists with a higher CDR tend to have a higher RR [15].

2.4. Mammograms and CAD system

All screening mammograms were performed using FFDM equipment (Senographe 2000 D, General Electric®). Standard two-view craniocaudal and mediolateral oblique mammograms of each breast were done.

CAD (SecondLook, Premier, iCAD®) was used. The CAD operating point was set at "H", which means maximum sensitivity with poor specificity. CAD used two types of marks: ovals for masses (including true masses, asymmetries, and architectural distortions) and rectangles for calcifications.

The sensitivity of the CAD system was calculated as the number of cancers correctly marked divided by the total number of cancers. The specificity of the CAD system was calculated as the number of negative studies (mammograms without CAD marks) in 6674 consecutive mammograms. We also collected the American College of Radiology (ACR) type of the breast. All mammograms were read on a dedicated IDI MammoWorkstation (General Electric®).

2.5. Reading protocol

2.5.1. Prospective study (2010–2012)

Firstly, the selected radiologists read the mammograms in a standard way and made a decision as to whether to recall the patient. At this point, the CAD marks were engaged. The radiologists revised the marks and made a final decision according to the marks and their own opinion.

If the final decision was that the mammographic findings were BI-RADS (Breast Image Report and Data System) 1 or 2, it meant that CAD marks were not considered. Otherwise, if the finding was considered to be BI-RADS 0, 4, or 5, the patient was recalled and the radiologist wrote down the type of recall according to the following procedure:

- Type A recall: the radiologist found some mammographic finding that CAD did not mark
- Type B recall: the radiologist found some mammographic finding that CAD also marked
- Type C recall: the radiologist found the mammogram to be BI-RADS 1 or 2 but after considering the CAD marks, determined that a lesion could indeed be present and decided to recall. Therefore, CAD changed the radiologist's mind.

2.5.2. Historical cohort (2004–2010)

All screening mammograms were double read. The pairs were established by trying to match an experienced radiologist with a less experienced one. Both radiologists read blinded to the other's readings (independent double reading). When a major disagreement occurred between the readers (one reader recalled and the other did not), a third radiologist, the most experienced radiologist in the team, acted as a tiebreaker by assessing the case and making a final decision.

2.6. Monitoring of prospective study

As the CDR may vary slightly from 1 month to another, we decided to collect the CDR quarterly to obtain homogenous data. The 95% confidence interval (CI) was determined for each quarter and compared with the average historical value of that quarter. Feedback was provided to the readers. The results are shown in Fig. 1.

If the historical CDR for one quarter was above the upper CI limit of the same quarter in the prospective research, the study was stopped because it would mean that the radiologists were detecting fewer cancers compared with the historical period, which would have been unacceptable.

2.7. Statistical analysis

Evaluation of data was undertaken using Excel (Microsoft, 2010). The results are shown as absolute frequencies and rates by 100 or 1000 mammograms (CDR). For the estimations of 95% CI in the prospective study, the link between the Poisson and Chi-square distributions was used [16]. This method is based on the Poisson distribution and is more accurate and exact than methods based on the normal distribution. Uncommon events in populations, such as the occurrence of breast cancer are usefully modeled using the