Preventive Medicine 115 (2018) 19–25

Contents lists available at ScienceDirect

Preventive Medicine

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

Accuracy of mammography and clinical breast examination in the implementation of breast cancer screening programs in Colombia

Luz Helena Alba[a], Sandra Díaz[b], Oscar Gamboa[c], César Poveda[d,e], Andrés Henao[f], Fernando Perry[b], Catherine Duggan[g], Fabián Gil[h], Raúl Murillo[i,c,*]

[a] Department of Preventive and Social Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia
[b] Clínica de seno, Instituto Nacional de Cancerología, Bogotá, Colombia
[c] Subdirección de Investigaciones y Salud Pública, Instituto Nacional de Cancerología, Bogotá, Colombia
[d] Servicio de Imaginología, Instituto Nacional de Cancerología, Bogotá, Colombia
[e] Facultad de Medicina, Universidad Nacional, Bogotá, Colombia
[f] Radiología, Centro de Enfermedades Mamarias, Bogotá, Colombia
[g] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[h] Department of Clinical Epidemiology and Biostatistics, Pontificia Universidad Javeriana, Bogotá, Colombia
[i] Prevention and Implementation Group, International Agency for Research on Cancer, Lyon, France

ARTICLE INFO

Keywords:
Sensitivity
Specificity
Mammography
Clinical breast exam
Colombia
Massive screening
Breast neoplasms

ABSTRACT

Most evidence on breast cancer screening accuracy derives from high income countries. We evaluated screening accuracy and factors related to program implementation in Bogota, Colombia. Between 2008 and 2012 participants underwent clinical breast examination (CBE) and mammography. Positive results underwent histological verification. Adherence to screening protocols was analyzed. Sensitivity, specificity, and predictive values were estimated and adjusted by overdiagnosis. Impact of alternative screening algorithms on follow-up was explored, including combined screening tests and modified coding systems for mammography. In total, 7436 women aged 50–69 were enrolled; 400 discontinued and 1003 non-compliant with screening protocols. 23 cancer cases were diagnosed. Mammography sensitivity and specificity were 78.3% (95%CI 77.3–99.3) and 99.4% (95%CI 99.2–99.6). CBE sensitivity was 39.1% (95%CI 37.9–40.3) and specificity 83.4% (95%CI 82.6–84.3). Parallel mammography and CBE showed the highest sensitivity (95.6%) and combined as serial tests the lowest (positive CBE followed by mammography 13.0%). A simplified coding system for mammography (recall/no-recall) had 6.3% of positive results and a minor reduction in specificity compared with standard mammography, but reported the best balance between recall rates and screening protocol compliance. Call-backs had high rates of loss-to-follow-up; thus, alternative screening algorithms might help increase screening compliance and follow-up in low and middle income countries, particularly in populations with poor screening history and low access to health services.

1. Introduction

Breast cancer is a leading cause of cancer incidence and mortality among women in low and middle income countries (LMICs) (Ferlay et al., 2013). Improved treatment and early detection via mammography has shown to reduce breast cancer mortality by 20–30% in screened women over 50 years of age in high income countries (Massat et al., 2015; Nelson et al., 2016).

Some in-depth analyses on the benefits and harms of breast cancer screening highlight over-diagnosis as leading to reduced cost-effectiveness of mammographic screening programs (Myers et al., 2015; Nelson et al., 2016); these issues combined with a lack of resources have prevented the implementation of early detection programs in most LMICs. Thus, recommendations for breast screening in LMICs rely mostly on promoting breast awareness and clinical breast examination (CBE); however, no strong evidence currently supports these recommendations (Thomas et al., 2013).

While the efficacy of mortality reduction informed by randomized clinical trials (RCT) represents the gold standard when evaluating cancer screening (Nelson et al., 2016), diagnostic accuracy studies can provide reliable information on the performance of screening tests in different populations (Eusebi, 2013; Mallett et al., 2012). In addition,

* Corresponding author at: Diagonal 70B No. 2-14 este apto. 104, Bogotá, Colombia.
E-mail addresses: murillor@iarc.fr, rmurillo@husi.org.co (R. Murillo).

https://doi.org/10.1016/j.ypmed.2018.08.005
Received 25 January 2018; Received in revised form 30 May 2018; Accepted 4 August 2018
Available online 06 August 2018

screening sensitivity has been identified as one of the major factors influencing the heterogeneity of clinical trials results (Chen, 2017). Hence, data on the ability of tests to discriminate patients with and without the disease will indicate their role in the clinical pathway (Eusebi, 2013) and, consequently, this type of information should be considered for proper planning of screening programs.

Although mammography and CBE are observer-dependent tests and might be influenced by several factors including interobserver variability, quality assurance, breast density, and disease prevalence, their accuracy has been evaluated mainly by RCT in high income countries (Nelson et al., 2016). Few data from LMICs are available. Recently, a cluster randomized trial carried out by the National Cancer Institute of Colombia (NCIC) evaluated the implementation of breast cancer screening guidelines with biennial mammography and CBE for women 50–69 years old, showing a positive effect on disease downstaging (Murillo et al., 2016). Here, we present results from a diagnostic study nested within this trial, aimed at evaluating the accuracy of mammography and CBE and its potential implications for implementing screening programs in low and middle income settings. In particular, we analyzed patient compliance with screening protocols and explored alternative screening algorithms to overcome limitations in adherence, a common problem in LMICs (Unger-Saldaña, 2014).

2. Methods

2.1. Participants

A diagnostic test accuracy study was conducted in the intervention arm of a cluster randomized trial (CRT) carried out in Bogotá between 2008 and 2012 (Murillo et al., 2016). The protocol was approved by the NCIC Ethical Committee and registered at Clinicaltrials.gov (NCT02337582). Women, 50 to 69 years old, attending ambulatory services for reasons not related to breast health, were consecutively recruited at 13 primary care centers without consideration of women's risk. Exclusion criteria comprised previous breast cancer diagnosis and history of screening mammography during the preceding two years.

2.2. Screening procedures

Upon providing informed consent, patients underwent CBE and were prescribed with two-view film mammography (cranio-caudal and mediolateral-oblique) (Fig. 1). Mammography results were interpreted in a single read by general radiologists according to the Breast Imaging Reporting and Data System (BIRADS) (Balleyguier et al., 2007); CBE was considered abnormal if any of the following were observed: skin changes (redness, orange skin, swelling), nipple retraction, discharge, asymmetric nodularity, hard or indefinite mass or lump, and axillary swellings (Barton et al., 1999).

General practitioners performing CBE received theoretical and practical training on breast cancer screening (BIRADS and for CBE the Barton technique). Radiographers and radiologists from mammography centers were trained on quality assurance procedures at the NCIC.

A quality control program was implemented. Mammography equipment was assessed prior to study outset and corrections implemented accordingly. NCIC breast surgeons periodically supervised quality of CBE by GPs in clinical practice. Two independent breast radiologists reviewed all BIRADS-4–5 mammograms and 10% of remaining results randomly selected. If readers disagreed (BIRADS-4–5 vs. lower), initial results were consensually re-categorized.

BIRADS-0 mammograms were prescribed with additional imaging evaluation (spot compression, magnified or additional views, ultrasound) to give a final BIRADS classification. Additional imaging was done in a radiology center other than the screening center. BIRADS-1–2 were recommended with a new screening round in two years; BIRADS-3 diagnosis had mammographic follow-up at 6 months; selected positive CBE cases based on medical criteria and all BIRADS-4–5 underwent biopsy (Fig. 1). Pathologists and investigators were blinded to screening test results and all breast cancer diagnoses were referred to treatment.

2.3. Follow-up

All women had up to two years of follow-up. The follow-up protocol has been described elsewhere (Murillo et al., 2016). Briefly, we used databases from health insurance companies to verify any diagnosis related to benign or malignant breast disease and possible metastatic breast cancer (bone, lung, and liver) (ICD-10). Additionally, we searched any report of procedures on the breast in accordance with Colombian health procedure codes (CUPS). Participants with publicly subsidized insurance, and women who withdrew from insurance companies during the study had no database entries; thus, data from personal interviews were used. All database findings and positive interview responses were verified by review of clinical records.

The gold-standard was the histological report if a biopsy was obtained; otherwise confirmation of disease status at the two-year follow-up was used. Each cancer case in the study underwent histological verification.

2.4. Statistical analysis

The sample size was estimated for the CRT (Murillo et al., 2016). With the number of women recruited in the intervention arm, a 10%–15% precision was expected for a mammography sensitivity about 85% (International Agency for Research on Cancer, 2016), and an anticipated breast cancer prevalence in Bogotá around 1497 cases (225.5 per 100,000 for women 50–69 years old) (Pardo & Cendales, 2015).

Sensitivity, specificity, positive predictive values (PPV), and likelihood ratios were estimated for mammography and CBE with corresponding confidence intervals (Mallett et al., 2012). Since not all patients with initial BIRADS-0 diagnosis returned to the screening center to be given a final BIRADS diagnosis after the additional imaging assessment, we used ultrasound BIRADS results (when available) as the final BIRADS classification for these patients.

We used data from the control arm in the CRT (Murillo et al., 2016) to estimate overdiagnosis as the cancer excess rate (Baines et al., 2016) and to adjust sensitivity correspondingly, as described by Prorok et al. (Prorok et al., 2015). Overdiagnosis was estimated by using two different denominators in the intervention arm: number of detected cases (clinical perspective) and total number of cancer cases (public health perspective) (Independent UK Panel on Breast Cancer Screening, 2012). Since no overdiagnosis was observed for CBE, we report adjusted sensitivity for this technique by excluding ductal carcinoma in situ (DCIS). In addition, we investigated the accuracy of alternatives to existing NCIC recommendations (Murillo et al., 2016), including: 1) combining CBE and mammography either in parallel assuming “any test positive”, or as serial tests assuming “both tests positive” (mammography after positive CBE); and 2) modifying the coding system for mammography.

Alternative coding systems for mammography screening were explored based on recommendations by the American College of Radiologists for screening mammography interpretation with BIRADS-0 as positive screening result (American College of Radiology, 2013) and by recreating the coding system used by the United Kingdom National Health Service (UK-NHS) screening program by assuming BIRADS-0, 3, 4, and 5 as “recall to diagnostic assessment” (NHS Cancer Screening Programmes, 2003). These are described as coding systems A and B, respectively.

Positivity rates for every screening algorithm were used as indicator of referrals for further examination (recall). We report adherence to clinical recommendations as attendance to screening and follow-up of positive or non-definitive screening results (BIRADS-0 and BIRADS-3). To analyze the effect of screening algorithms on protocol compliance we hypothetically assumed a two-visit approach (screening and diagnostic work-up). We assumed attendance to screening as attendance for