ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 942–950

Journal of Clinical Epidemiology

The use of additional imaging increased specificity and decreased sensitivity in screening mammography

Berta M. Geller[a,*], Pamela M. Vacek[b], Joan Skelly[c], Susan C. Harvey[d]

[a]*University of Vermont, Office of Health Promotion Research, Departments of Family Practice and Radiology, Vermont Cancer Center, 1 South Prospect Street, Burlington, VT 05401-3444, USA*

[b]*University of Vermont, Departments of Medical Biostatistics and Pathology, Vermont Cancer Center, 1 South Prospect Street, Burlington, VT 05401-3444, USA*

[c]*University of Vermont, Department of Medical Biostatistics, 1 South Prospect Street, Burlington, VT 05401-3444, USA*

[d]*University of Vermont, Fletcher Allen Health Care, Radiology Department, Vermont Cancer Center, 1 South Prospect Street, Burlington, VT 05401-3444, USA*

Accepted 17 February 2005

Abstract

Background and Objectives: To examine the use of additional imaging after standard mammographic screening views to better understand the value of these additional testing in improving accuracy.

Materials and Methods: Statewide population data on screening mammography were used to report accuracy measures at screening and after additional imaging tests. Pathology data from biopsies performed within 1 year following the screening mammogram were used to determine cancer outcome (489 *in situ* and invasive cancers). Pathologic and population characteristics of women receiving different types of imaging were assessed by chi-square or *t*-tests. Similar tests compared women with the same imaging and differing outcomes.

Results: Of 77,799 women with screening mammograms 9.9% had additional imaging. Additional imaging reduced false positives from 7,765 (100/1,000 mammograms) to 1,112 (14/1,000 mammograms). The majority of false negatives (82%) occurred in women receiving only screening views, and additional imaging increased the number of false negatives from 82 (1/1,000 mammograms) to 115 (1.5/1,000 mammograms).

Conclusion: Additional imaging can reduce unnecessary biopsy but at the cost of some additional false negatives. Additional imaging's potential for improving the sensitivity of screening is limited because most missed cancers occur in women who do not have additional imaging.

Keywords: Accuracy; Breast; Cancer; Mammography; Observational; Screening

1. Introduction

Mammographic screening to detect nonpalpable breast cancer has become a routine medical procedure for women in many parts of the world, and has been shown to result in a significant reduction in mortality from breast cancer [1]. In the United States, mammographic screening is not always a single event. It has the potential to be a series of imaging events that begins with the two standard screening views of each breast (medial lateral oblique and craniocaudal). Radiologists may obtain additional imaging studies to augment the standard screening views before completing the assessment and recommendation [2,3]. Mammographic compression and magnification views are often used to evaluate micro calcifications, asymmetric densities, masses, or areas of architectural distortion. Ultrasound is frequently used to differentiate solid from cystic masses or to further evaluate masses and asymmetrical densities that are seen on the mammogram.

The use of additional imaging to augment the standard screening views varies [4]. The American College of Radiology has suggested that less than 10% of women should be recalled for additional imaging [3]. Within the United States the recall rates among mammography facilities has varied by as much as 11% (range 1.9–13.4) [5], and has recently been reported to average 13.1% for initial mammograms and 8% for subsequent mammograms [6]. Many other countries have a much lower recall rate than the United States [4,6–8]. In The Netherlands the referral rate is under 2% for initial screens, and is lower for subsequent screening examinations [7], while in Finland [4] and in the United Kingdom [6] the recall rate is 3.2 and 7.4%, respectively, for initial mammograms. However, definitions of recall in the literature are not necessarily the same, making comparisons difficult [8].

* Corresponding author. Tel.: 802-656-4115; fax: 802-656-8825.
E-mail address: berta.geller@uvm.edu (B.M. Geller).

0895-4356/05/$ – see front matter
doi: 10.1016/j.jclinepi.2005.02.009

Many factors may influence the use of additional imaging. These include: facility and individual radiologist practice patterns, the radiologists' level of training and years of experience [9,10], availability of previous comparison films, the radiographic density of the breast tissue, patient use of hormone therapy (HT) [11], and family or personal history of breast cancer [12].

A high level of sensitivity combined with a high degree of specificity is desirable for a screening program. The goal is to correctly diagnose breast cancer and limit the number of benign biopsies women undergo because of false positive assessments. It is generally believed that additional imaging will decrease the number of women who are sent for benign biopsy [13]. Although additional imaging may improve the accuracy of mammography, there are also negative consequences of doing additional imaging including an increase in cost [14], as well as psychologic consequences such as increased anxiety for women who have been asked to return for additional tests [15–17]. Achieving a balance between the number of women recalled for additional testing and the accurate diagnosis of cancer is a challenge, and has not been carefully researched [5].

Many studies have recently looked at the recall rates in the United States and have examined the real and induced costs of false positive mammography [5,10,11,14–18]. Most studies that assess the efficacy of additional imaging have used test sets and have focused on the detection of breast cancers that were not definitively identified using only screening views [19]. The purpose of this study is to describe the use of additional imaging following standard screening views in routine clinical practice, and to examine its effect on the accurate diagnosis of breast cancer in the larger context of all women undergoing mammographic screening. Observational data from a statewide mammography registry, the Vermont Breast Cancer Surveillance System [20] are used to: (1) characterize the women receiving additional imaging, (2) estimate screening accuracy before and after additional imaging, and (3) identify women-specific characteristics associated with accuracy among women receiving different imaging procedures. The unique aspects of this study are that it focuses on screening studies for asymptomatic women, and includes mammographic interpretations from all community radiology practices within a defined geographic region.

2. Methods

Data from the Vermont Breast Cancer Surveillance System (VBCSS), a member of the National Cancer Institute's Breast Cancer Surveillance Consortium, was used for this study [21]. Since 1994, the VBCSS has collected information on all mammography and breast ultrasound examinations performed in Vermont. Data collected include the reason for the breast imaging visit, the types of views taken, and the final assessment. In this study data are from 57 radiologists who work in 17 Food and Drug Administration (United States) certified breast-imaging facilities. The VBCSS also collects breast pathology data from all 12 pathology laboratories in Vermont and breast cancer information from the Vermont Cancer Registry, as well as from the New Hampshire Cancer Registry because some women from eastern Vermont receive medical care in New Hampshire. In addition, risk factor information is obtained by asking the patient to complete a health history questionnaire at the time of her mammogram. Data confidentiality is carefully protected, and has been described in detail elsewhere [22,23]. The Institutional Review Board at the University of Vermont approved the protocol for this project with an alteration of informed consent. A paragraph on the data collection form indicated the multiple uses for the data, including research. If women did not want their data used for research they checked a box indicating this choice.

Women who had a screening mammogram between February 1998 and December 1999 were included in this study. Until February 1998, the VBCSS did not collect the American College of Radiology's Breast Imaging Reporting and Data Systems (BIRADS)® [3] assessment categories following ultrasound examinations. Screening mammograms performed prior to 2000 were used to allow sufficient time following the mammogram to identify all of the cancer cases from the reporting cancer registries. Mammograms were included if the reason for mammography, as indicated by the radiologist or mammography technologist, was an asymptomatic screening examination. If a woman had more than one screening mammogram during this period, data from the first one was used. Because there is only one mammogram per woman we use the terms woman and mammogram interchangeably in this report. Final assessments were based on the BIRADS®, which uses six assessment categories: (0) "Needs additional imaging evaluation"; (1) "Negative"; (2) "Benign finding"; (3) "Probably benign finding"; (4) "Suspicious abnormality"; and (5) "Highly suggestive of malignancy." A negative assessment was defined as BI-RADS® categories 1–3. As per BI-RADS®, we considered mammograms in categories 0, 4, and 5 positive mammograms. As additional imaging was completed and assessment categories 1–5 were assigned, mammograms that were originally a category 0 became either a positive or negative assessment as defined above. Women who had category 0 assessments after all imaging were excluded from the study ($N = 230$) because we could not be sure that they did not have additional imaging at a facility out of state. In addition, because their assessments did not change after additional imaging, their exclusion does not alter the results regarding the effects of additional imaging. Women who did not give permission to use their data for research were also excluded ($N = 2{,}773$). After these exclusions there were a total of 77,799 women in the study, with 489 having a positive biopsy outcome.

Pathology data from biopsies performed within 1 year following the screening mammogram were used to determine cancer outcome. Biopsies included surgical excision,