Magnetic Resonance Imaging and Mammography in Women With a Hereditary Risk of Breast Cancer

Mark J. Stoutjesdijk, Carla Boetes, Gerrit J. Jager, Louk Beex, Peter Bult, Jan H. C. L. Hendriks, Robert J. F. Laheij, Leon Massuger, Lya E. van Die, Theo Wobbes, Jelle O. Barentsz

Background: Although breast cancer screening is recommended to start at a younger age for women with a hereditary risk of breast cancer, the sensitivity of mammography for these women is reduced. We compared magnetic resonance imaging (MRI) with mammography to determine which is more sensitive and whether MRI could play a role in the early detection of breast cancer for these women. Methods: We constructed a retrospective cohort of all breast MRI and mammography surveillance reports made in our department from November 1994 to February 2001. All of the 179 women in the cohort had received biannual palpation in addition to annual imaging by MRI, mammography, or both. The 258 MRI images and the 262 mammograms were classified with the use of the BI-RADS™ (i.e., Breast Imaging Reporting and Data System) scoring system, which has five categories to indicate the level of suspicion of a lesion. Receiver operator characteristic curves were generated for MRI and mammography, and the area under each curve (AUC) was assessed for the entire cohort of 179 women and for a subset of 75 women who had received both an MRI and a mammographic examination within a 4-month period. All statistical tests were two-sided. Results: In the cohort of 179 women, we detected 13 breast cancers. Seven cancers were not revealed by mammography, but all were detected by MRI. For the entire cohort, the AUC for mammography was 0.74 (95% confidence interval [CI] = 0.68 to 0.79), and the AUC for MRI was 0.99 (95% CI = 0.98 to 1.0). For the subset of women who had both examinations, the AUC for mammography was 0.70 (95% CI = 0.60 to 0.80), and the AUC for MRI was 0.98 (95% CI = 0.95 to 1.0). Conclusion: MRI was more accurate than mammography in annual breast cancer surveillance of women with a hereditary risk of breast cancer. Larger prospective studies to examine the role of MRI in screening programs are justified. [J Natl Cancer Inst 2001;93:1095–102]

The average lifetime risk of developing breast cancer for a woman in The Netherlands is 11.6% (1). Many factors increase a woman’s risk of breast cancer to values far greater than this general-population average. Such factors include germline mutations of several genes, most notably BRCA1 and BRCA2, that increase the lifetime risk of breast cancer for carriers of these mutations to between 60% and 85% (2,3). More than 50% of women with mutations in BRCA1 and BRCA2 are diagnosed with breast cancer before the age of 50 years (4–7). Therefore, a woman from a family at risk for breast cancer should be encouraged to participate in a surveillance program, starting at a young age.

For many years, the imaging modality of choice for breast cancer screening has been mammography, but its sensitivity is severely reduced when screening young women (8–11). Because of the reported high sensitivity of magnetic resonance imaging (MRI) (12,13), the diagnostic performance of this modality is expected to be higher than that of mammography when imaging young women. However, so far only very few reports have directly compared the two modalities (14,15).

In this report, we validated more than 6 years of MRI and mammography screening reports with histology or 2-year follow-up by imaging, to determine whether MRI has a substantially higher

Affiliations of authors: M. J. Stoutjesdijk, C. Boetes, G. J. Jager, J. H. C. L. Hendriks, L. E. van Die, J. O. Barentsz (Department of Radiology), L. Beex (Department of Oncology), P. Bult (Department of Pathology), R. J. F. Laheij (Department of Medical Technology Assessment), L. Massuger (Department of Gynecology), T. Wobbes (Department of Surgery), University Medical Center St Radboud, Nijmegen, The Netherlands.

Correspondence to: Mark J. Stoutjesdijk, M.D., M.Sc., Department of Radiology, University Medical Center St Radboud, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands (e-mail: m.stoutjesdijk@rdiag.azn.nl).

See “Notes” following “References.”

© Oxford University Press
overall diagnostic performance than mammography for the detection of breast cancer in women with a hereditary risk of breast cancer.

**Subjects and Methods**

**Subjects**

The data for this retrospective cohort study were collected from the radiologic report and pathology databases of the University Medical Center St Radboud, Nijmegen, The Netherlands, where annual breast MRI for women at risk for early-onset familial breast cancer has been practiced since 1994. We first selected all reports of breast cancer surveillance examinations that used MRI or mammography from November 1994 to February 2001. We next refined the group to be included in the study based on both indication and confirmation requirements. First, the patient’s lifetime risk of breast cancer had to exceed 15% based on family history of breast or ovarian cancer or the presence of a germline mutation in the BRCA1 or BRCA2 gene. Second, the patient had to have no personal history of breast cancer. Third, adequate follow-up data had to be available for the confirmation of the radiologic findings; i.e., a report was included in the study if follow-up MRI or mammography was available at least 2 years later. A report was also included in the study if a lesion found by MRI or by mammography was confirmed by histology. Because all of the patients’ data were evaluated anonymously and retrospectively, no institutional approval or written informed consent from cohort members was necessary.

The age at examination and the category of lifetime risk of breast cancer were recorded for all of the patients in the cohort. To estimate the lifetime risk of breast cancer, we used the method from an ongoing Dutch national multicenter breast cancer study (16), which is based on the cumulative risk at the age of 79 years according to the model of Claus et al. (17). The risk categories were mutation carrier (group 1: 50%–85% estimated lifetime risk), very high (group 2: 30%–50% estimated lifetime risk), and high (group 3: 15%–30% estimated lifetime risk). The presence of ovarian carcinoma in the family history was also incorporated in the risk stratification. If breast cancer was found in one first- or second-degree family member at the age of 55 years or younger and ovarian cancer was found in two first- or second-degree family members, then the lifetime risk for breast cancer for the patient was high (15%–30%). If three or more first- or second-degree family members were affected by breast or ovarian cancer, then the lifetime risk of breast cancer for the patient was high (15%–30%); if at least one of the three was a first-degree family member, then the lifetime risk for the patient was even higher (30%–50%). For some women in the cohort, the lifetime risk was between 15% and 50%, but the distinction between a risk category of 2 or 3 could not be made because not all of the data needed for the estimation had been recorded.

In addition to surveillance for breast cancer, women known to be carriers of germline BRCA1 or BRCA2 mutations received annual screening for ovarian carcinoma. This screening consisted of a gynecologic examination, a transvaginal ultrasound, and a serum CA-125 measurement. Preventive oophorectomy was discussed with patients who were 35 years or older. Suspected carriers were also screened for ovarian carcinoma, but oophorectomy was not discussed.

Because the cohort was constructed retrospectively, we had no control over the choice of modality for the examinations. For each woman, this choice had depended on a variety of factors, including whether the amount of glandular tissue on previous mammograms had affected the mammography interpretation, whether the woman had experienced physical distress during previous mammography examinations, and whether a family member of the woman had ever had a malignancy missed by mammography.

A total of 245 women with a familial risk of breast cancer received one or more surveillance MRI or mammography examinations during our study. Sixty-six of these women were excluded from the study because no history or adequate follow-up information was available. Of the remaining 179 women in the cohort (aged 21–71 years) (Table 1), 40 had received mammography only, 49 had received MRI only, 75 had received mammography and MRI within a 4-month period at least once, and 15 had received mammography and MRI but at a year or more apart. If only one imaging modality and physical examination were used initially, equivocal or suspicious findings often led to additional diagnostic imaging that included the other modality as part of the work-up.

All of the women in the cohort had received biannual physical examination and annual imaging by MRI, mammography, or both. Slightly more than 92% of the women had a first-degree relative with breast cancer who was diagnosed before the age of 55 years, and 5% of the women had a first- or second-degree relative with ovarian cancer. In the analysis of the cohort, we used all confirmed surveillance rounds of the 179 women. A total of 262 mammograms had been interpreted (from 130 women), and a total of 258 MRI examinations had been interpreted (from 139 women).

In addition, we analyzed a subgroup within the full cohort, which contained only the women who had received both MRI and mammography within a 4-month interval at least once during the study period. For the analysis of this subgroup, we used only the most recent surveillance round.

**Imaging Protocol**

The MRI examinations were carried out on a 1.5T system (Magnetom Vision; Siemens, Erlangen, Germany) with a standard bilateral dedicated breast coil (CP Breast Array; Siemens). MRI examinations for all of the premenopausal patients were done in the second week of the menstrual cycle to minimize glandular tissue enhancement. We used a dynamic contrast-enhanced FLASH-3D sequence, with a repetition time of 8.1 milliseconds, an echo time of 4 milliseconds, and a flip angle of 20 degrees. Both breasts were fully imaged at all times. After a precontrast series of images, 0.2 mmol/kg body weight of gadopentetate dimeglumine contrast agent (Magnevist®; Schering, Berlin, Germany) was given by bolus intravenous injection with the use of a power injector (Powerinjector®; Medrad, Pittsburgh, PA). A series of five postcontrast images were obtained. For all series in the examination, the scanning orientation was axial, and the acquisition time for each series was 80 seconds. In November 1999, the imaging protocol changed to a coronal scanning orientation to reduce artifacts in the axillary region due to the motion of the heart. The new acquisition time for each series was 87 seconds.

All of the images were displayed in the transverse plane for interpretation. Subtraction images were
used to visualize the early and late contrast enhancement of possible tumors (18,19). Since 1995, curves of contrast enhancement versus time (20,21) have been used to aid in the analysis of suspicious-looking breast lesions. Maximum intensity projections were introduced early in 1999 to improve the visualization of suspicious lesions. Throughout the study, all breast MRI examinations were interpreted with the use of state-of-the-art criteria for the detection and diagnosis of breast lesions (21–23).

Mammography examinations were carried out on a Mammatom 3000 (Siemens) or a Senographe 2000D (GE Medical Systems, Milwaukee, WI). In addition to the standard mediolateral oblique and craniocaudal projections, magnification views in both projections were obtained, if needed. If possible, mammograms were made during the second week of the menstrual cycle, initially because the examination would be less painful. The choice of screening in this menstrual phase has recently received additional support; it has been suggested that mammographic sensitivity is probably decreased during the luteal phase of the menstrual cycle (24), in which the breast density is substantially increased (25) and the risk of carcinogenic mutation from the radiation exposure is high (26).

For some women, high-frequency ultrasound examination was performed by an experienced radiologist (C. Boetes or J. H. C. L. Hendriks) using a Toshiba SSA 380 ultrasound unit (Toshiba America Medical Systems, Tustin, CA) that was equipped with a 10-MHz transducer.

Reporting and Confirmation

The MRI examinations were prospectively interpreted by one of two radiologists (C. Boetes and L. E. van Die), who were blinded to the mammography results. The mammograms were interpreted by one of three radiologists (C. Boetes, J. H. C. L. Hendriks, or L. E. van Die), who were blinded to the mammographic interpretation throughout the United States. The images were classified as 1) negative, 2) benign, 3) probably benign, 4) suspicious abnormality, and 5) highly suggestive of malignancy. BI-RADS also includes a classification of 0, which indicates that additional imaging is required. If a lesion was classified as probably benign, a follow-up examination 3 months later was recommended. If a lesion was classified as suspicious or highly suggestive of malignancy, a biopsy specimen was taken or the lesion was excised to obtain a histologic diagnosis. Examinations that were done before the introduction of BI-RADS™ in our center were retrospectively given a BI-RADS™ score of 1 if they were reported as “negative” and 2 if they were reported as “benign.”

To minimize the effect of differences between surveillance and diagnostic (i.e., work-up) examinations, a radiologist (C. Boetes) re-read all of the “positive” examinations, i.e., those that were not initially reported as negative or benign. We added an equal number of randomly chosen negative or benign examinations from each modality to the reading set. The radiologist was blinded to patient data, the number of “positive” examinations, and the number of added negative or benign examinations.

Confirmation of the imaging results from either mammography or MRI was obtained by histology within 2 months or by subsequent annual examinations. Material for histologic examination was obtained by core needle biopsy or by excisional biopsy. If ductal carcinoma in situ (DCIS) or invasive carcinoma was found, a simple or modified radical mastectomy was performed. If histologic examination identified a malignant lesion, then all of the imaging examinations from the preceding 2 years were evaluated retrospectively by an expert radiologist (C. Boetes) to identify any possible false-negative reports. Negative or benign reports were considered to be a true-negative result if all imaging results from at least the next 2 years were still negative. Negative or benign reports were also considered to be a true-negative result if a malignancy had been found within the 2-year follow-up period, but only if the lesion could not be identified retrospectively on the original images. In our study, a false-negative result was defined as the result of an examination that was originally reported as BI-RADS™ 1 or 2, but where a malignancy was actually visible in retrospect or a malignancy was visible on the other modality. The malignancy had to be detected (by either imaging or physical examination) within the 2-year follow-up period.

All of the lesions were classified at pathologic examination. For in situ and invasive carcinomas, tumor size, histologic type, and differentiation grade were determined. DCIS was graded I, II, or III, according to the classification of Holland et al. (28). For invasive carcinomas, the presence of metastases in the axillary lymph nodes was determined by examination of conventional hematoxylin–eosin-stained slides of the lymph nodes. Invasive carcinomas were graded according to the Elston method (29).

Estrogen receptor (ER) status and progesterone receptor (PgR) status of the invasive carcinomas were determined by immunohistochemistry. From the selected paraffin blocks of embedded tumors, 4-μm sections were cut and dried overnight at room temperature. After the section was dewaxed, endogenous peroxidase activity was inhibited with 3% hydrogen peroxide in phosphate-buffered saline. Antigen retrieval was performed twice in a microwave oven (800 W) for 5 minutes in 10 mM citrate buffer (pH 6) at a temperature of 100 °C. Blocking kit (Vector Laboratories Inc., Burlingame, CA) was used to inhibit endogenous biotin according to the manufacturer’s recommendations. The sections were incubated with 20% normal horse serum (NHS) in phosphate-buffered saline for 15 minutes at room temperature before they were incubated overnight at 4 °C with primary antibody (ER [1D5; M7047; Dako, Glostrup, Denmark] 1:200 dilution in 20% NHS or PgR [Ab-8 MS-298P, Neo Markers, Fremont, CA] 1:400 dilution in 20% NHS). The sections were subsequently incubated with a biotinylated horse anti-mouse antibody (Vector Laboratories Inc.) at a dilution of 1:200 in 20% NHS, ABC Elite 1:50 (Vector Laboratories Inc.), and 3,3'-diaminobenzidine (Sigma Chemical Co., St. Louis, MO). The sections were counterstained with Mayer’s hematoxylin and mounted in Permount. The sections immunohistochemically stained for ER and PgR were scored for the intensity of the staining (intensity score) and for the proportion of positively stained tumor cells (proportion score), similar to the method described by Barnes et al. (30). Both scores were determined by assigning a mean subjective value over the whole cross-section of the invasive tumor present in the tissue section. The intensity score ranged from 0 to 3 and was defined as follows: 0 = no staining even at high magnification; 1 = mean staining only at high magnification; 2 = mean staining obvious, but not intense, at low magnification; and 3 = mean intense staining even at low magnification. The proportion score ranged from 0 to 5 and was defined as the percentage of tumor cells stained: 0 = none, 1 = less than 1%, 2 = 1%–25%, 3 = 25%–50%, 4 = 50%–75%, and 5 = 75% or more. The “quick score” was defined as the sum of the intensity and proportion scores (31). The hormone receptor status was considered to be negative if the quick score was 0 or 2, to be slightly positive if the quick score was 3 or 4, to be moderately positive if the quick score was 5, 6, or 7, and to be strongly positive if the quick score was 8.

Statistical Analysis

Statistically significant differences in age and risk-category distribution between the two groups (MRI and mammography) of the entire cohort were compared with the use of the χ² test. The sensitivity, specificity, and positive and negative predictive values were each calculated for MRI and mammography. Because these parameters vary with the chosen operating point (i.e., the minimum BI-RADS™ score at which a report is considered to be “positive”), receiver operator characteristic (ROC) curves were determined with the use of the data from the entire cohort. To obtain a parameter for diagnostic performance, we assessed the area under the ROC curve (AUC). Because bias by the clinician’s selection of the modality (selection or referral bias) could affect the data from the entire cohort, we made an additional set of ROC curves, using the data from the subset of patients who had received both MRI and mammography within a 4-month interval. Despite the relatively small number of malignancies detected in the cohort, multivariate regression analysis was used on this subset to obtain ROC curves that were adjusted for age and risk category. Patients for whom the risk category could not be specified beyond “2 or 3” were excluded from the multivariate regression analysis. A difference in the AUCs between mammography and MRI was assessed for the entire cohort with the use of a z-score test and for the subset of women who had received both mammography and MRI within 4 months with the use of the method described by Hanley and McNeil (32).

For all statistical analyses, we used the SAS software package (SAS Institute, Inc., Cary, NC). All P values reported are from two-sided tests, with α = 0.05.

RESULTS

To determine whether young women would benefit from MRI in addition to mammography for the early detection of breast cancer, we generated a retrospective cohort to compare the two modalities. There was no statistically significant difference in age (P = .43) or in risk category (P = .61) distribution between

Journal of the National Cancer Institute, Vol. 93, No. 14, July 18, 2001 REPORTS 1097
the women in the MRI and mammography groups (Table 1). Among the 179 women in the entire cohort, we detected 13 malignant tumors, all of which were imaged with MRI and 12 of which were imaged with mammography (Table 2). One cancer was a low-grade non-Hodgkin’s lymphoma, which was not visible on mammography. Of the remaining 12 breast cancers, three (25%) were DCIS. None of these in situ carcinomas were found in confirmed carriers of BRCA1 or BRCA2 mutations. Nine cancers had a predominantly invasive component; five (56%) of these cancers were considered to be lymph node negative (i.e., were negative for cancer in the axillary lymph nodes), and four (44%) were considered to be lymph node positive. No interval carcinomas, i.e., symptomatic cases following a negative examination, were found in this study.

From the 258 MRI examinations, 30 lesions were assigned BI-RADS™ scores of 3–5 (Table 3). Thirteen of the 30 lesions proved to be malignant by core or excisional biopsy. There were no false-negative MRI reports, and there were 228 true-negative or benign MRI examinations. From the 262 mammograms, 15 lesions were assigned BI-RADS™ scores of 3–5. Five of the 15 lesions proved to be malignant by core or excisional biopsy. There were seven false-negative reports, in five of which the radiologist mentioned the presence of dysplasia. There were 240 true-negative mammographic reports.

We next determined the sensitivity, specificity, and predictive values, with the operating point of the ROC curve set at a BI-RADS™ score of 3. The sensitivity, specificity, predictive value of a positive report, and predictive value of a negative report were 42% (95% confidence interval [CI] = 36% to 48%), 96% (95% CI = 94% to 98%), 33% (95% CI = 27% to 39%), and 97% (95% CI = 95% to 99%), respectively, for mammography and 100% (95% CI = 98% to 100%), 93% (95% CI = 90% to 96%), 43% (95% CI = 37% to 49%), and 100% (95% CI = 98% to 100%), respectively, for MRI. If the operating point was set at a BI-RADS™ score of 4, the values were 42% (95% CI = 36% to 48%), 99% (95% CI = 98% to 100%), 63% (95% CI = 57% to 69%), and 97% (95% CI = 95% to 99%), respectively, for mammography and 92% (95% CI = 89% to 95%), 98% (95% CI = 96% to 100%), 71% (95% CI = 65% to 77%), and 99.6% (95% CI = 98.8% to 100%), respectively, for MRI.

From the sensitivity and the specificity data, we generated ROC curves, using the entire BI-RADS™ range (Fig. 1, A). The AUC was 0.74 (95% CI = 0.68 to 0.79) for mammography and 0.99 (95% CI = 0.98 to 1.0) for MRI. The difference in the AUCs (0.28 [95% CI = 0.17 to 0.39]) was statistically significant ($P = .02$).

Finally, we used multivariate regression analysis to adjust for age and risk category and determined new AUCs. For the entire cohort, the AUC was 0.82 (95% CI = 0.76 to 0.87) for mammography and 0.99 (95% CI = 0.98 to 1.0) for MRI. The difference in AUCs (0.17 [95% CI = 0.12 to 0.26]) was statistically significant ($P = .02$). For the subgroup who received both mammography and MRI within a

<table>
<thead>
<tr>
<th>Malignant tumor No.</th>
<th>BI-RADS™</th>
<th>Risk category</th>
<th>Age, y</th>
<th>Dysplasia</th>
<th>Primary tumor type</th>
<th>Grade</th>
<th>TNM§</th>
<th>Receptor status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>5</td>
<td>1 (BRCA1)</td>
<td>No</td>
<td>Medullary carc</td>
<td>III</td>
<td>T2N0M0</td>
<td>e−, p− (QS 0.2)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1 (BRCA1)</td>
<td>No</td>
<td>MALT lymphoma</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>44</td>
<td>No</td>
<td>IDC</td>
<td>III</td>
<td>T1N0M0 e−, p− (QS 0.0)</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>47</td>
<td>No</td>
<td>IDC</td>
<td>II</td>
<td>T2N0M0 e+, p+ (QS 7.8)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>50</td>
<td>Yes</td>
<td>IDC</td>
<td>II</td>
<td>T1N0M0 e+</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>42</td>
<td>Yes</td>
<td>DCIS</td>
<td>II</td>
<td>TisN0M0 —</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>30</td>
<td>Yes</td>
<td>DCIS</td>
<td>III</td>
<td>TisN0M0 —</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>46</td>
<td>Yes</td>
<td>ILC</td>
<td>I</td>
<td>T2N1M0 e+, p+ (QS 8.7)</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>49</td>
<td>Yes</td>
<td>IDC/ILC</td>
<td>III</td>
<td>T2N1M0 e+, p+ (QS 7.6)</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>40</td>
<td>No</td>
<td>IDC</td>
<td>II</td>
<td>T2N1M0 e+, p+ (QS 7.4)</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>44</td>
<td>Yes</td>
<td>ILC/IDC</td>
<td>II</td>
<td>T2N0M0 e−, p+</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>50</td>
<td>Yes</td>
<td>IDC/ILC</td>
<td>III</td>
<td>T2N1M0 e−, p− (QS 0.0)</td>
</tr>
<tr>
<td>13</td>
<td>—</td>
<td>4</td>
<td>2/3</td>
<td>30</td>
<td>—</td>
<td>DCIS</td>
<td>III</td>
<td>TisN0M0 —</td>
</tr>
</tbody>
</table>

* $M =$ mammography; MRI = magnetic resonance imaging; TNM = tumor–node–metastasis staging (55); medullary carc = medullary carcinoma; e− = estrogen receptor negative; e+ = estrogen receptor positive; p− = progesterone receptor negative; p+ = progesterone receptor positive; QS = quick score; MALT = mucosa-associated lymphoid tissue; IDC = invasive ductal carcinoma; DCIS = ductal carcinoma in situ; ILC = invasive lobular carcinoma.

†BI-RADS™ (i.e., Breast Imaging Reporting and Data System) indicates the grade assigned to the examination as detected by mammography and MRI. Grading 3–5 initiates short-term follow-up or biopsy.

§Lifetime risk of breast cancer according to guidelines stated in an ongoing Dutch national multicenter breast cancer study (16) and based on the model of Claus et al. (17).

The estrogen and progesterone receptor status are shown only if determined. The QS is calculated as the sum of the intensity and proportion scores. A QS of 0–2 = negative, of 3–4 = slightly positive, of 5–7 = moderately positive, and of 8 = strongly positive.
4-month period, the AUCs were 0.80 (95% CI = 0.70 to 0.90) for mammography and 0.99 (95% CI = 0.96 to 1.0) for MRI. The difference in AUCs (0.19 [95% CI = 0.09 to 0.29]) was statistically significant ($P = .05$).

**DISCUSSION**

The identification and management of women with a hereditary risk of breast cancer have received much attention during the last few years. Several studies have addressed the assessment of lifetime risk of breast cancer from the patient’s history (33,34) and the efficacy of early diagnosis and treatment (35,36). The options of prophylactic mastectomy and oophorectomy for women with a hereditary risk of breast cancer have been investigated (37–40), and intensive surveillance programs beginning at a younger age have been proposed (35). The imaging modality of choice for radiologic screening has, so far, been mammography. However, because young women have dense breast tissue, the sensitivity of mammography is severely reduced. This reduced sensitivity is reflected in the results of studies that compared mammographic sensitivity between young and older women (41,42). If surveillance is to be compared with options such as prophylactic mastectomy, it is necessary to use optimal radiologic screening techniques. Therefore, consideration of surveillance by means of breast MRI is essential.

We investigated whether MRI had a substantially higher overall diagnostic performance than mammography in a retrospective cohort of women with a hereditary risk of breast cancer. The most important result of our study was that the AUC was statistically significantly higher for MRI than for mammography. The AUC approximates the chance that the radiologist correctly described the presence or absence of a malignancy (43). However, there are several limitations to our study design.

First, one should be aware that the results are biased by the fact that the examined population consisted of women with a hereditary risk of breast cancer who opted for annual surveillance. In this group of women, the results may be different from those obtained in a group of women with a hereditary risk of breast cancer who all received an invitation to participate in a screening program. It is also not possible to fully disentangle the mix of surveillance and diagnostic examinations. This could account for some of our results because the radiologists may have known that a suspicious lesion was found with the other modality. In an attempt to minimize the effects of this mix, a radiologist re-read all reports from either modality that were assigned a BI-RADS™ grade 3 or higher. To create a “surveillance mind-set” in the radiologist, we added an equal number of randomly chosen reports with BI-RADS™ grades 1 or 2. The radiologist was blinded to all patient data and was told only that a large fraction of the reports were negative.

Second, only one radiologist was used for the identification of possible false-negative results. A panel of radiologists may have provided a higher degree of certainty.

Third, although a large number of examinations reduces the width of the CIs, repeated examinations on the same women (as done in the analysis of the full cohort) increase their width. It is, therefore, possible that the CIs of the full cohort are wider than those cited; however, given the large difference in diagnostic accuracy (i.e., in the AUCs), it is unlikely that the outcome would change.

![Fig. 1. Comparison of receiver operator characteristic (ROC) curves for magnetic resonance imaging (MRI) and mammography for the detection of breast cancer lesions in a retrospective cohort of young women with a hereditary risk of breast cancer. A) Data from the entire cohort. B) Data from a subset of women who received both MRI and mammography examinations within a 4-month interval. Numbered data points indicate the sensitivity (sens, y-axis) and 1 minus specificity (1-spec, x-axis) obtained if the minimum BI-RADS™ (i.e. Breast Imaging Reporting and Data System) grade for a positive report is varied from 1 to 5.](image-url)
Fourth, the number of breast cancers was low, which limited the analysis of contributing factors other than the examination modality to the results.

Fifth, the cohort was constructed retrospectively; i.e., there was no control over the type(s) of examination that each patient received. Selection bias may, therefore, have contributed to the outcome and should be considered to be a study design limitation. There are methods for correcting for selection bias, but it is difficult to identify precisely which factors contributed to any selective referral. Although important characteristics of those who received MRI and those who received mammography were similar (see Table 1), there may still have been some selection bias because of covariates that could not be analyzed.

Sixth, the examination results from different years on the same woman are correlated. To account for the correlated nature of the data, we separately analyzed the reports in which both mammography and MRI were done within a 4-month interval. In this subset of women, only the last screening round was used, which eliminated the contribution of selection bias to the difference in the AUC. There is a considerable similarity between the ROC curves and AUCs of the entire cohort and those of the subset who received both examinations, suggesting that the results from the two groups are comparable. This similarity is also present in the AUCs generated from data that were adjusted for age and risk category.

Breast cancer detection and diagnosis are dependent on the presentation of clinical features of malignancy that can be identified by the imaging technique. If there are no such features but a malignancy is present, there will be an avoidable delay in the time to diagnosis. We defined a false-negative result as one that contributed to the difference in the AUC. There is a considerable similarity between the ROC curves and AUCs of the entire cohort and those of the subset who received both examinations, suggesting that the results from the two groups are comparable.

A potential problem with MRI is the relatively low predictive value of a positive result. To minimize unnecessary invasive procedures, such as biopsy and lumpectomy, it is important to consider the options if a suspicious lesion is found. Follow-up MRI after 3–6 months may be feasible if the lesion is small (3–5 mm), especially if mammography and ultrasound are both negative. If an ultrasound-guided biopsy is not possible, MRI-guided localization or biopsy could be attempted before lumpectomy is done. This may increase the chance of obtaining material that is satisfactory for diagnosis. However, we should note that, in this study, the predictive value for a positive test for mammography was even lower than that for MRI. There are also additional drawbacks associated with breast MRI. For example, MRI uses an intravenous contrast agent, is more expensive and time consuming than mammography, is affected considerably by the patient’s menstrual cycle, and is problematic for claustrophobic patients because of the narrow magnet bore. However, the gadolinium-based contrast agent is not iodinated, and adverse reactions are extremely rare (46). Also, if the MRI examination is done around the second week of the menstrual cycle, hormone-induced glandular enhancement is minimal (47–49). Despite the drawbacks of MRI, one important advantage of MRI, as shown by this study, is that it has increased sensitivity compared with mammography. This observation agrees with reports that virtually all malignant tumors show sufficient contrast enhancement for detection of MRI (13); the exceptions are well-differentiated DCIS and, to a much lesser extent, invasive lobular carcinoma, which sometimes do not enhance very well (50,51). Another advantage of MRI over mammography is the absence of ionizing radiation with MRI. Because women with a hereditary risk of breast cancer are also at an increased risk of radiation-induced DNA damage (52), MRI may decrease unnecessary exposure to ionizing radiation in these women.

The objective of breast cancer surveillance is to detect malignant tumors that are in as early a stage as possible, so that the patient has the highest chance of survival after treatment. Such a strategy works for women in the general population aged 50 years and over (53) and seems to benefit younger women with a family history of breast cancer (35). However, these observations were made in studies that used mammography as the only imaging modality. We show that the overall sensitivity of breast MRI is higher than that of mammography in a population of young women with a hereditary risk of breast cancer. Our result confirms the recent work of Kuhl et al. (14) and Tilanus-Linthorst et al. (15). Therefore, patient survival may improve if these women were imaged annually with breast MRI.

With the data from this study, it is not possible to predict which women will benefit most from the higher sensitivity of MRI before a first mammographic examination is done. Furthermore, a minimum lifetime risk of breast cancer or an age at which mammography suffices cannot be given. However, if a large amount of mammographically dense tissue is present, e.g., due to dysplasia, MRI in place of mammography seems indicated.

Because findings such as a palpable mass or increased tissue density on the mammogram tend to initiate further imaging, the subset of women who had both MRI and mammography had a disproportionate number of tumors. Observations that are made from this select population clearly cannot simply be generalized to the entire cohort of women with a hereditary risk of breast cancer. However, the similarity of the ROC curves and AUCs between those of the subset and those of the entire cohort strongly supports the validity of the results. Nonetheless, prospective studies are needed to confirm these results, and large multicenter studies comparing MRI and mammography are presently under way in The Netherlands (16) and in the U.K. (54).

In conclusion, this retrospective study shows that annual screening with breast MRI is more accurate than mammography in the early detection of malignant breast tumors in women with a hereditary risk of breast cancer. This encouraging result needs to be confirmed by larger
considered for young women with a hereditary risk of breast cancer who opt for annual surveillance.

REFERENCES


(45) Spratt JS, Meyer JS, Spratt JA. Rates of growth


NOTES

We thank J. J. Füttterer, W. J. H. Veldkamp, and R. Heesakkers of the University Medical Center St Radboud, Nijmegen, The Netherlands, for their contributions to this study.

Manuscript received November 29, 2000; revised April 13, 2001; accepted May 24, 2001.