

BI-RADS에 의거하여 영상의학과 의사가 평가한 유방 밀도와 Volpara 유방 밀도의 비교

최지수¹ · 이찬화¹ · 서희정¹ · 김민주¹ · 이주희¹ · 이창윤¹ · 주정남² · 고경란¹

¹국립암센터 영상의학과

²국립암센터 바이오테크 연구과

목적: 영상의학과 의사가 육안적으로 판정한 유방 밀도 측정의 재현성을 평가하고, 영상의학과 의사가 판정한 유방 밀도와 Volpara 소프트웨어가 측정한 유방 밀도의 일치도를 비교하고자 한다.

대상 및 방법: 2011년 9월부터 11월 사이에 본원에서 유방촬영술을 시행한 환자 중 Volpara 유방 밀도 결과가 있는 913명의 환자를 대상으로 하였다. 유방 촬영 영상의 유방 밀도를 7명의 영상의학과 의사가 미국 방사선의학회의 유방영상 판독 및 데이터체계 (Breast Imaging Reporting and Data System: BI-RADS)에 의거하여 측정하였고, 2주 이상의 간격을 두고 두 번에 걸쳐 판독하였다. 영상의학과 의사의 유방 밀도 결과와 Volpara 유방 밀도 결과를 비교하기 위해, 첫 번째 판독 세션에서 영상의학과 의사 다수가 보고한 측정 결과 (majority report)를 참조 표준 (reference standard)으로 삼고, 이와 두 방법의 일치도를 카파 분석 (kappa test)을 이용하여 평가하였다. 일치도는 네 개의 범주 (범주1/2/3/4)로 나누었을 때와 두 개의 범주로 나누었을 때 (범주1-2/범주3-4)로 각각 평가하였다.

결과: 영상의학과 의사의 유방 밀도 측정의 판독자내 변이 (intraobserver variability, 카파 값 범위 0.69-1.00)는 우수 혹은 아주 우수한 일치도를 보였으며, 판독자간 변이 (interobserver variability, 카파 값 범위 0.55-0.77)는 중등도 혹은 우수한 일치도를 보였다. 참조 표준 (majority report)과 영상의학과 의사의 유방 밀도 결과는 우수 혹은 아주 우수한 일치도를 보였으며 (범주1/2/3/4: 카파 값 범위 0.75-0.86, 범주1-2/3-4: 카파 값 범위 0.70-0.90), Volpara 유방 밀도 결과는 중등도의 일치도를 보였다 (범주1/2/3/4 카파 값 0.55, 범주1-2/3-4 카파 값 0.59).

결론: 영상의학과 의사가 BI-RADS에 의거하여 평가한 유방 조직 밀도는 우수한 재현성을 보였다. 영상의학과 의사가 평가한 유방 조직 밀도와 Volpara 소프트웨어가 측정한 유방 조직 밀도는 중등도의 일치도를 보였다.

Index words: Mammographic breast density; Volpara; Breast Imaging Reporting and Data System

서 론

유방 밀도는 유방암의 위험인자로 알려져 있으며 (1-

3), 유방 촬영 영상의 민감도를 결정하는 요인으로도 알려져 있다 (4-6). 범주 4의 치밀유방인 경우, 유방 촬영술의 민감도는 30%까지 감소하며 (7, 8), 이러한 이유로 고밀도 유방을 가진 환자에게 유방 촬영술 외에 초음파와 같은 추가적인 스크리닝 검사가 선택적으로 시행되기도 한다 (9, 10). 따라서 환자의 유방암 위험도와 유방 촬영술의 민감도 예측을 위해서 정확한 유방 밀도 판정이 필요하다.

통신저자 : 고경란
(410-769)경기도 고양시 일산동구 일산로 323, 국립암센터 영상의학과
Tel. (031) 920-1186, Fax. (031) 920-1184
E-mail: kokyounglan@hanmail.net

유방 촬영 영상에서 유방 밀도를 측정하는 방법 중 가장 널리 쓰이고 있는 방법은 미국방사선의학회(American College of Radiology: ACR)의 유방영상 판독 및 데이터체계(Breast Imaging Reporting and Data System: BI-RADS, 이하 BI-RADS)에서 제시한 기준에 따라 유방 밀도를 정성적으로 평가하는 방법이다(11). 이 방법은 영상의학과 의사가 전체 유방에서 섬유유선 조직이 차지하는 비율을 육안으로 보고 4단계 범주로 분류하는 것인데, 육안 평가이기 때문에 주관적이며, 판독자간, 판독자내 변이가 많을 수 있다(12, 13). 이러한 정성적인 방법에 비해 정량적인 방법은 전체 유방 볼륨 중 섬유유선 조직이 차지하는 볼륨을 비율로 계산하여 나타내기 때문에 더 객관적이고 재현성이 높다는 장점이 있다(14). 유방 밀도의 정량적 측정을 위해 여러 소프트웨어가 개발되었으며, Volpara 소프트웨어는 볼륨 기반의 유방 밀도 측정법(fully-automated volumetric measure) 중 하나로, 총 유방 볼륨(total breast tissue volume, cm^3)과 섬유유선 조직의 볼륨(fibroglandular tissue volume, cm^3)의 백분율을 자동으로 계산하여 유방 밀도를 4단계의 범주로 분류한다.

본 연구는 ACR BI-RADS에 의거하여 영상의학과 의사가 육안적으로 분류한 유방 밀도 범주의 재현성을 평가하고, 또한 영상의학과 의사가 분류한 범주와 Volpara 소프트웨어가 측정한 유방 밀도 범주가 얼마나 일치하는지를 알아보고자 한다.

대상 및 방법

1. 환자

2011년 9월부터 11월까지 본원에서 자각증상, 스크리닝, 유방 수술 후 추적 검사 등의 이유로 유방 촬영술을 시행한 환자 중 Volpara 유방 밀도 측정 결과가 있는 913명의 환자의 유방 촬영 영상이 연구에 포함되었다. 대상 환자는 무작위로 선택되었으며, 대상 환자의 평균 나이는 53.6세(범위: 39-88세)였다.

2. 영상 검사 및 유방 밀도 측정

연구에 포함된 유방 촬영 영상은 1대의 디지털 유방촬영기를 이용하여 얻었다(Senographe 2000D digital mammography unit, General Electric Medical Systems, Milwaukee, WI, USA). 유방 촬영 영상에서 보이는 유방 밀도를 7명의 영상의학과 전문의가 ACR BI-RADS의 권고에 따라 4단계의 범주로 나누어 평가하였다(범주1: almost entirely fatty, <25%

glandular/ 2: scattered fibroglandular, 25-50% glandular/ 3: heterogeneously dense, 51-75% glandular/ 4: extremely dense, >75% glandular). 7명의 영상의학과 전문의 중 4명은 1년 이상의 유방 영상 판독 경험이 있었고, 3명의 의사는 1년 이내 경험이 있었다. 영상의학과 의사 별 판독의 일관성(intraobserver consistency)을 보기 위해 유방 밀도 판정은 2주 이상의 간격을 두고 두 번에 걸쳐 시행되었다. 영상의학과 의사들이 유방 밀도 판정을 할 때 유방 촬영 영상 이외의 다른 정보들은 제공되지 않았다(blind reading).

Volpara (Version 1.4, Matakina Technology, LTD, Wellington, New Zealand) 소프트웨어를 이용한 유방 밀도 측정은 유방 조직 내의 섬유유선 조직과 지방 조직의 방사선 감쇠(X-ray attenuation)의 차이를 바탕으로 유방 촬영 영상의 각 픽셀(pixel)당 포함되어 있는 섬유 유선 조직의 볼륨을 계산함으로써 이루어졌다. 각 픽셀 값을 바탕으로 좌우 유방 각각에서 섬유유선 조직 볼륨을 전체 유방 볼륨으로 나눈 유방밀도 값(volumetric breast density, %)이 측정되었다. 좌우 유방에서 얻은 값의 평균 값을 바탕으로 유방 밀도를 4단계로 분류하였다(범주 1: 0-4.5%/ 2: 4.5-7.5%/ 3: 7.5-15.5%/ 4: > 15.5%). 각각의 단계는 ACR BI-RADS의 각 범주에 해당한다고 보았다.

3. 통계 분석

영상의학과 의사가 BI-RADS에 의거하여 판정한 유방 밀도 결과의 재현성을 보기 위해 판독자간 변이(interobserver variability)와 판독자내 변이(intraobserver variability)를 평가하기 위해 카파 분석(kappa test)을 이용하였다. 카파 값(k value) 0.2 이하는 불량(minimal agreement), 0.21-0.40은 보통(fair agreement), 0.41-0.60은 중등도(moderate agreement), 0.60-0.80은 우수(substantial agreement), 0.81-1.00은 아주 우수(almost perfect agreement)로 분류하였다. 일치도는 4단계의 범주(범주1/2/3/4)로 나누었을 때와 2단계의 범주로 나누었을 때(범주1-2/범주3-4)로 각각 평가하였다.

영상의학과 의사의 유방 밀도 결과와 Volpara 유방 밀도 결과를 비교하기 첫 번째 판독 세션에서 영상의학과 의사 다수가 보고한 측정 결과(majority report)를 참조 표준(reference standard)로 삼았다. 참조 표준과 영상의학과 의사가 측정한 유방 밀도, Volpara 유방 밀도와 일치도를 카파 분석(kappa test)을 이용하여 평가하였다. 일치도는 4단계의 범주(범주1/2/3/4)로

나누었을 때와 2단계의 범주로 나누었을 때 (범주1-2/범주3-4)로 각각 평가하였다.

결 과

7명의 영상의학과 의사가 분류한 유방 밀도 범주 결과를 Table 1에 나타내었다. BI-RADS에 의거하여 각 영상의학과 의사가 판정한 유방 밀도 범주는 두 번의 판정에서 각 단계별로 비슷한 분포를 보였다. 첫 번째와 두 번째 판독에서 범주 3-4로 나타내는 고밀도유방의 평균 분포는 각각 66.2%, 65.0%였다. 영상의학과 의사의 유방 밀도 측정의 판독자내 변이 (intraobserver variability)는 4단계의 범주로 나누었을 때, 4명의 의사에서 아주 우수한 일치도를 (카파 값 범위 0.82-0.99), 3명에서 우수한 일치도를 (카파 값 범위 0.69-0.80) 보였다 (Table 2). 2단계로 나누었을 때는 5명의 의사에서 아주 우수한 일치도를 (카파 값 범위 0.82-1.00), 2명에서 우수한 일치도를 (카파 값 범위 0.70-0.78) 보였다

(Table 2). 7명의 판독자간 변이 (interobserver variability)는 4단계로 나누었을 때 (카파 값 범위 0.552-0.773)와 2단계로 나누었을 때 (카파 값 범위 0.585-0.825) 모두 중등도 혹은 우수한 일치도를 보였다.

Table 3는 영상의학과 의사의 유방 밀도 결과와 Volpara 유방 밀도 결과의 일치도와 참조 표준과 이 두 방법을 이용한 결과의 일치도를 같이 나타내었다. 각 영상의학과 의사의 유방 밀도 결과는 7명 모두 참조 표준과 우수 혹은 아주 우수한 일치도를 보였다 (4단계 범주: 카파 값 범위 0.668-0.867/ 2단계 범주: 카파 값 범위 0.704-0.905). 각 영상의학과 의사의 유방 밀도 결과와 Volpara 유방 밀도 결과는 4단계 범주로 나누었을 때 7명 모두 중등도의 일치도를 보였으며 (카파 값 범위 0.441-0.599), 2단계로 나누었을 때 6명이 중등도의 일치도를 (카파 값 범위 0.457-0.598) 1명이 우수한 일치도를 보였다 (카파 값 0.656). Volpara 유방 밀도 결과는 참조 표준과 중등도의 일치도를 보였다 (4단계 범주 카파 값 0.554/ 2단계 범주: 카파 값 0.598).

Table 1. Frequency of BI-RADS Mammographic Categories Reported at First or Second Reading by 7 Radiologists

Radiologist	A	B	C	D	E	F	G	Average (%)
First Reading								
D1	65	37	57	45	57	47	67	53.6 (5.9)
D2	362	254	298	206	138	281	249	255.4 (27.9)
D3	262	493	461	380	611	457	400	437.7 (47.8)
D4	226	131	99	284	107	129	199	167.9 (18.3)
D3-4	488	624	560	664	718	586	599	605.6 (66.2)
Second Reading								
D1	58	37	60	43	41	53	76	52.6 (5.7)
D2	462	256	293	200	130	279	248	266.9 (29.2)
D3	237	488	498	387	623	445	437	445.0 (48.6)
D4	158	134	63	285	120	137	152	149.9 (16.4)
D3-4	395	622	561	672	743	582	589	594.9 (65.0)

D1 < 25% glandular; D2 25-50% glandular; D3 50-75% glandular; D4 > 75% glandular

Table 2. Intraobserver Consistency (First vs. Second Reading within Each Reader) in Reporting BI-RADS Mammographic Density Categories

Radiologist	A	B	C	D	E	F	G	Average
Absolute Agreement (D1/D2/D3/D4, %)	78.90	99.45	86.21	73.45	87.73	95.18	84.99	86.56
Absolute Agreement (D1-D2 vs. D3-D4, %)	88.96	99.78	91.57	87.97	94.64	97.37	93.64	93.42
Kappa Value (D1/D2/D3/D4)	0.78 (0.75-0.81)	0.99 (0.99-1.00)	0.82 (0.79-0.85)	0.69 (0.66-0.73)	0.80 (0.77-0.84)	0.94 (0.92-0.96)	0.84 (0.81-0.86)	0.84
Kappa Value (D1-D2 vs. D3-D4)	0.78 (0.74-0.82)	1.00 (0.99-1.00)	0.82 (0.78-0.86)	0.70 (0.64-0.75)	0.83 (0.79-0.88)	0.94 (0.92-0.97)	0.86 (0.83-0.90)	0.85

95% confidence limits are reported in brackets

Table 3. Kappa Statistics of Agreement between BI-RADS Mammographic Density Categories and Volpara Mammographic Density Categories

Reader	D1 (%)	D2 (%)	D3 (%)	D4 (%)	Kappa vs Majority (D1/D2/D3/D4)	Kappa vs Majority (D1-2/D3-4)	Kappa vs Volpara (D1/D2/D3/D4)	Kappa vs Volpara (D1-2/D3-4)
A	65 (7.1)	362 (39.6)	262 (28.6)	226 (24.7)	0.750 (0.719-0.781)	0.718 (0.674-0.762)	0.549 (0.513-0.586)	0.457 (0.407-0.508)
B	37 (4.0)	254 (27.8)	493 (53.9)	131 (14.3)	0.850 (0.824-0.877)	0.890 (0.859-0.922)	0.561 (0.523-0.599)	0.656 (0.603-0.710)
C	57 (6.2)	298 (32.6)	461 (50.4)	99 (10.8)	0.803 (0.772-0.833)	0.811 (0.772-0.850)	0.441 (0.402-0.480)	0.507 (0.452-0.563)
D	45 (4.9)	206 (22.5)	380 (41.5)	284 (31.0)	0.668 (0.632-0.704)	0.737 (0.690-0.784)	0.538 (0.497-0.579)	0.528 (0.465-0.592)
E	57 (6.2)	138 (15.1)	611 (66.9)	107 (11.7)	0.751 (0.715-0.787)	0.704 (0.654-0.754)	0.475 (0.432-0.518)	0.566 (0.501-0.632)
F	47 (5.1)	281 (30.7)	457 (50.0)	129 (14.1)	0.862 (0.837-0.888)	0.886 (0.855-0.918)	0.485 (0.445-0.524)	0.537 (0.480-0.594)
G	67 (7.3)	249 (27.2)	400 (43.7)	199 (21.8)	0.867 (0.842-0.891)	0.905 (0.876-0.934)	0.599 (0.563-0.635)	0.592 (0.537-0.648)
Majority	57 (6.2)	247 (27.0)	462 (50.5)	149 (16.3)	Ref.	Ref.	0.554 (0.515-0.592)	0.598 (0.542-0.654)
Volpara	45 (4.9)	155 (16.9)	386 (42.2)	329 (36.0)			Ref.	Ref.

Row percentages and 95% confidence limits are reported in brackets

고 찰

유방 촬영 영상에서 고밀도 유방을 보이는 여성은 저 밀도 유방을 보이는 환자보다 유방암 발생 위험이 3-6배 높은 것으로 알려져 있다. 이는 기존에 유방암 위험인자로 알려져 있던 미분만 (nulliparity)이나 이른 초경 (early menarche) 의 위험도보다 높다 (15). 따라서 유방암 발생 위험 예측에 있어 유방 밀도와 유방암 발생 위험과의 관계를 규명하는 것이 임상적으로 중요하며, 이를 알기 위한 유방암 위험 예측 모델이나 역학 연구 (epidemiological study)에 유방 밀도가 이용된다면 유방 밀도의 정확한 측정이 무엇보다 중요하다. 유방암 발생 위험과 유방 밀도의 관계를 알아보았던 기존 연구에서는 유방 촬영 영상에서 보이는 섬유 유선 조직 픽셀 값을 기준으로 유방 밀도를 측정하는 반 자동 소프트웨어를 이용한 (semi-automated software) 정량적 방법이나 BI-RADS 등을 이용한 정성적 방법이 이용되어 왔다. 그러나 반 자동 소프트웨어를 이용할 경우, 매 영상마다 판독자가 기준이 되는 픽셀 값을 정해주어야 하고, BI-RADS 분류 또한 판독자의 육안 평가로만 이루어지기 때문에 이 두 방법 모두 재현성이 낮을 수 있다는 단점이 있다 (14, 15). 따라서 유방암 위험 인자 관련 연구에 쓰일 수 있는 객관적이고 재현성이 높은 정량적 분석

방법이 필요하였고, 최근 유방 조직의 볼륨을 측정하여 유방 밀도를 자동으로 측정하는 소프트웨어들이 개발되었다 (3, 14-17). 본 연구는 이중 하나인 Volpara 소프트웨어를 유방 밀도 측정에 이용하였다.

본 연구에서 영상의학과 의사의 BI-RADS 유방 밀도 측정 결과는 우수한 재현성 (reproducibility)을 보였으며, 이는 이전 연구 결과와도 일치하는 것이다 (16, 18). 이 연구는 이전 연구보다 많은 수의 유방 촬영 영상 (n=913)을 대상으로 하였고, 유방 밀도를 평가한 영상의학과 의사 중 유방 영상 판독의 경력이 1년 이내인 의사가 포함된 것을 감안할 때, 영상의학과 의사가 BI-RADS에 의거하여 측정한 유방 밀도는 임상적으로 이용되기에 적합하다고 볼 수 있겠다. 또한 이는 의사 다수가 보고한 유방 밀도 범주 (majority report)가 신뢰할만한 육안 밀도 평가의 참조 표준이 될 수 있음을 의미한다.

이 연구의 통계적 분석에서는 BI-RADS 유방 밀도 측정이 우수한 재현성을 보였으나, 유방 밀도가 임상적으로 개인 맞춤형 스크리닝 (personalized screening, 범주 3-4로 분류되는 고밀도 유방을 가진 여성에서 초음파를 스크리닝의 목적으로 추가하거나 매년 유방 촬영 영상을 시행 하는 예)의 기준으로 이용된다면 유방 밀도의 육안 평가는 제한이 있다 (14). 예를 들어 전체 유방의 49%가 섬유유선 조직인 환자와 51%인 환자의 유방 밀

도를 판독자가 육안적으로 정확히 구분해내는 것은 쉽지 않을 것이다. 따라서 고밀도와 저밀도로 나누는 기준 값 (cut off value)과 가까운 유방 밀도를 가진 환자의 경우, 같은 유방 촬영 영상이라도 판독자에 따라 유방 밀도 범주가 다르게 평가될 가능성이 있으며 이러한 측정 결과의 차이는 스크리닝 방법의 차이를 가져올 수 있다. 반면 Volpara 소프트웨어는 자동으로 유방 밀도 결과를 측정하기 때문에 객관적이며, 재현성이 높다는 장점이 있다. 따라서 위에서 언급한 개인 맞춤형 스크리닝의 척도로 유방 밀도를 측정하는 경우나, 타목시펜이나 에스트로젠 투여 후 유방 밀도의 감소 혹은 증가를 평가하는 경우에서도 Volpara 소프트웨어가 유용하게 쓰일 수 있을 것이다. 그러나 유방암과 유방 밀도의 연관성을 보여준 기존 연구 결과들은 반 자동 소프트웨어를 이용한 (semi-automated software) 정량적 방법이나 BI-RADS등을 이용한 정성적 방법을 이용하였기 때문에 기존 연구 결과를 바탕으로 Volpara 소프트웨어를 이용하여 유방암 연구를 진행하기 위해서는 기존 방법과 Volpara 소프트웨어의 통계적 연관성이 먼저 입증되어야 한다.

본 연구에서 영상의학과 의사가 평가한 BI-RADS 유방 밀도의 참조표준과 Volpara 유방 밀도는 4단계 범주로 나누었을 때 카과 값 0.554, 2단계 범주로 나누었을 때 카과 값 0.598로 중등도의 일치도를 보였으며, 이는 이전 연구와 일치하는 결과이다 (19). 그러나 이 연구는 두 방법에서 측정한 유방 밀도 범주의 일치도만 분석하였고, Volpara 소프트웨어에서 계산한 유방 밀도 값 (volumetric breast density, %)과 BI-RADS 유방 밀도 범주와의 관계는 분석하지 않았다. Volpara 소프트웨어가 유방 밀도 측정에 널리 쓰이기 위해서 현재 범주 구분에 쓰이고 있는 Volpara 유방 밀도 기준 값이 적절한지, 혹은 기존에 쓰이던 측정 방법과 우수한 일치도를 보일 수 있는 Volpara 유방 밀도 기준 값에 대한 추가적인 연구가 필요할 것으로 보인다.

본 연구는 몇 가지 제한점이 있다. 첫째, 이 연구에서는 Volpara 소프트웨어 유방 밀도 측정의 오류가 발생한 경우는 제외되었다. 이전 연구에서 Volpara 소프트웨어는 전체 유방 볼륨이 작거나 유방 밀도가 높은 경우에는 기술적인 오류가 발생할 수 있다고 보고된 바 있다 (19). 둘째, 이 연구는 하나의 기관에서 후향적으로 진행되었기 때문에 연구 결과를 일반화하기에는 제한이 있다. 그러나 이 연구에 포함된 영상의학과 의사들이 유방 밀도 판정을 할 때 유방 촬영 영상 이외의 다른 정보들은 제공되지 않았으며 (blind reading), 각 의사들은 각기 다른 기관에서 교육 (training)을 받은 의사들이었기 때

문에 우리 결과는 편향적이지 않다고 할 수 있을 것이다.

결론적으로 영상의학과 의사가 BI-RADS에 의거하여 평가한 유방 조직 밀도는 우수한 재현성을 보였다. 영상의학과 의사가 평가한 유방 조직 밀도와 Volpara 소프트웨어가 측정한 유방 조직 밀도는 중등도의 일치도를 보였다.

참 고 문 헌

1. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* 1976;37:2486-2492
2. Vachon CM, Kuni CC, Anderson K, Anderson VE, Sellers TA. Association of mammographically defined percent breast density with epidemiologic risk factors for breast cancer (United States). *Cancer Causes Control* 2000;11:653-662
3. Ding J, Warren R, Warsi I, et al. Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study. *Cancer Epidemiol Biomarkers Prev* 2008;17:1074-1081
4. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-236
5. Ciatto S, Visioli C, Paci E, Zappa M. Breast density as a determinant of interval cancer at mammographic screening. *Br J Cancer* 2004;90:393-396
6. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. *J Natl Cancer Inst* 2004;96:1432-1440
7. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 2002;225:165-175
8. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 2000;92:1081-1087
9. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA* 2008;299:2151-2163
10. Corsetti V, Houssami N, Ferrari A, et al. Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost. *Eur J Cancer* 2008;44:539-544
11. ACR Bi-Rads : breast imaging system : breast imaging atlas : mammography, breast ultrasound, magnetic resonance imaging. 4th ed. Reston, Va.: American College of Radiology, 2003
12. Boyd NF, Wolfson C, Moskowitz M, et al. Observer variation in the classification of mammographic parenchymal patterns. *J Chronic Dis* 1986;39:465-472
13. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR Am J Roentgenol* 2000;174:1769-1777
14. Assi V, Warwick J, Cuzick J, Duffy SW. Clinical and epidemiological issues in mammographic density. *Nat Rev Clin Oncol* 2012;9:33-40
15. Boyd NF, Martin LJ, Bronskill M, Yaffe MJ, Duric N, Minkin S. Breast tissue composition and susceptibility to breast cancer. *J Natl Cancer Inst* 2010;102:1224-1237

16. Tagliafico A, Tagliafico G, Tosto S, et al. Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. *Breast* 2009;18:35-40
17. Ciatto S, Bernardi D, Calabrese M, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. *Breast* 2012;21:503-506
18. Ciatto S, Houssami N, Apruzzese A, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast* 2005;14:269-275
19. Su Yeon Ko E-KK, Min Jung Kim, Hee Jung Moon. Mammographic Density Estimation: Comparison between Radiologist's Visual Assessment and Volpara Breast Density. *J Korean Soc Breast Screening* 2012;9:11-17

J Korean Soc Breast Screening 2012; 9 : 127-132

Mammographic density assessment: Comparison of Volpara™ Software and Visual BI-RADS Classification

Ji Soo Choi, MD¹, Chan Wha Lee, MD¹, Hee Jung Seo, MD¹, Min Ju Kim, MD¹, Ju Hee Lee, MD¹,
Chang Yoon Lee, MD¹, Jungnam Joo, MD², Kyoung Lan Ko, MD¹

¹Department of Radiology, National Cancer Center

²Biometric Research Branch, National Cancer Center

Purpose: To evaluate the reproducibility of Breast Imaging Reporting and Data System (BI-RADS) classification by radiologists and to compare Volpara estimation with BI-RADS classification in mammographic density assessment.

Materials and Methods: From September to November 2011, 913 patients in whom mammography was performed at a referral center and mammographic density was graded by Volpara software were included in this study. Mammographic density was assessed by seven radiologists according to BI-RADS. Majority report of radiologists was used as standard reference to compare BI-RADS to Volpara classification for mammographic density assessment. Agreement between the two methods and the reference standard was assessed using kappa test according to four categories (D1/2/3/4) and two categories (D1-2/D3-4).

Results: Intraobserver (k range 0.69–1.00) agreement of the seven radiologists was substantial or almost perfect, and interobserver (k range 0.55–0.77) agreement of the radiologists was moderate or substantial. The agreement with the reference standard was substantial or almost perfect for BI-RADS classifications made by radiologists (D1/2/3/4: k range 0.75-0.86, D1-2/D3-4: k range 0.70–0.90), and was moderate for Volpara classification (D1/2/3/4: k = 0.55, D1-2/D3-4: k = 0.59).

Conclusion: Mammographic density classification by radiologists according to BI-RADS was highly reproducible. The agreement between BI-RADS and Volpara classifications was moderate in mammographic density estimation.

Index words: Mammographic breast density; Volpara; Breast Imaging Reporting and Data System

Corresponding author: Kyoung Lan Ko, M.D.