

유방촬영검사 팬텀영상평가의 부적합 사례 연구

고은영¹ · 우옥희² · 최선형³ · 장정민⁴ · 김혜원⁵ · 김유미⁶ · 김명경¹

¹성균관대학교 의과대학 삼성서울병원 영상의학과, ²고려대학교 구로병원 영상의학과, ³퀸즈유의원,
⁴서울대학교병원 영상의학과, ⁵원광대학교병원 영상의학과, ⁶동군산병원 영상의학과

목적: 부적합 판정을 받았던 유방팬텀영상들을 분석하여 부적합 증례들의 특징과 원인이 되는 요소들을 알아보고자 한다.

대상 및 방법: 2013년 1월부터 2016년 3월까지 한국의료영상 품질관리원에 접수되었던 유방팬텀영상들 중 최종 부적합 판정을 받았던 총 207례의 유방팬텀영상을 후향적으로 분석하였다. 유방촬영영상 품질관리에 20년 이상의 경력을 가진 6명의 유방영상의학 전문의가 2명씩 한 조로 나누어 팬텀영상 채점 기준에 따라 재채점을 하였고, 두 명의 판정 결과가 서로 일치하지 않은 경우 6명의 판정단 중 또 다른 3명이 재검토한 후 다수결에 따라 최종 점수를 결정하였다. 팬텀영상의 배경 불균질이나 인공물의 정도를 판단하여, 없는 경우, 약간 있는 경우, 심한 경우에 따라 각각 0,1,2점으로 분류하였다. 부적합 판정을 초래했던 원인이 되는 모조병소들을 살펴보고, 각 병소별 비율을 조사하였다. 또한 평가자들 사이 적합-부적합 판정이 일치하지 않았던 평가 항목과 그 비율을 분석하였다. 영상의 배경 불균질이나 인공물의 정도가 각 병소의 점수나 적합-부적합 판정, 평가자 간의 일치도와 관련이 있는지를 Pearson's correlation을 이용하여 분석하였다.

결과: 가장 부적합율이 높았던 항목은 종괴였고(90.85%), 가장 낮은 항목은 미세석회화 그룹이었다(2.4%). 적합-부적합 판정에 대한 평가자 별 일치도를 보았을 때, 각 증례에 대한 적합-부적합 일치도는 100%였으나 모든 항목에 대해 결과가 일치했던 경우는 60.9%였다. 총 621개의 모조 병소들로 나누어 분석했을 때에 항목 별 불일치율은 15.5%였다. 불일치율은 종괴 항목에서 가장 많았고, 배경 불균질이나 인공물의 정도가 심할수록 불일치율이 증가하였다 ($p=0.027$).

결론: 부적합 판정이나 평가자 간의 불일치율은 모두 종괴 병소에서 가장 높아 종괴에 대한 채점기준의 보완이나 교육이 더 필요할 것으로 생각된다. 또한 배경 불균질이나 인공물에 대한 평가 방법을 추가하는 것이 판정 일치도를 높이고 더 객관적인 평가를 위해 필요할 것으로 보인다.

Index words: Breast phantom, Mammography quality control, Phantom image evaluation, Noncompliance cases

본 연구는 2024년도 식품의약품안전처 연구개발비(RS-2024-00332874)로 수행되었으며 이에 감사드립니다.

Correspondence to: Eun Young Ko, M.D., Ph.D.

Department of Radiology and Center for Imaging Science,

Samsung Medical Center, Sungkyunkwan University School of Medicine,

81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea

Tel. +82-2-3410-6418 Fax. +82-2-3410-0084

E-mail: claudel@skku.edu

서 론

국가암검진 사업 중 수검율이 가장 높은 유방암 검사는 유방촬영술이라는 검사 방법으로 시행한다 (1). 유방촬영술에 쓰이는 유방촬영장치는 의료법 제 38조 특수의료장비의 설치 및 운영에 관한 법률에 의거하여 특수의료장비로 분류되며, 이를 설치하고 운영하는 의료기관에서는 전속 또는 비전속의 품질관리 담당 전문의를 채용하여 품질관리 점검표에 따른 검사를 시행해야 한다. 또한 국가에서 인정하는 품질관리 검사기관에 매년 서류검사와 3년마다 정밀검사를 받아야 한다 (2, 3).

유방촬영장치의 경우 CT, MRI와 비교했을 때에 의원급의 설치대수가 많다는 것이 특징이며, 개원시에 중고장비를 설치하는 경우가 많고, 이에 따라 노후된 장비가 많으며, 이는 유방촬영장치의 팬텀영상 부적합율이 다른 특수의료장비에 비해 높은 이유로 여겨진다. 또한 품질관리 담당 전문의가 비전속으로 있는 의료기관의 경우나 영상의학과 전문의가 아닌 다른 과 전문의가 관리자로 등록되어 있는 경우에, 영상의학과 상근 전문의가 있는 기관보다는 특수의료장비의 품질관리에 대한 개념이 부족하다. 따라서 평소에 지속적인 품질관리 업무를 하지 않고 정기검사 시기에 임박하여 일시적으로 외부업체에 장비 관리를 맡기는 곳들도 있고, 평상시 품질관리가 잘 이루어지지 않는 의료기관들도 많다 (4, 5).

유방촬영장치의 경우, 매년 인력검사, 정도관리기록검사와 팬텀영상 검사를 받아야 하는데, 이 때에 각 기관에서 6개월마다 스스로 시행하고 기록한 팬텀영상평가 결과와 함께 최근 1달 이내에 촬영한 유방팬텀영상을 제출해야 한다. 유방팬텀영상 검사는 실제 환자의 임상영상을 촬영하기 전에 해당 유방촬영장치가 적절한 영상이 얻어지는지를 확인하는 목적으로 시행하는 검사로, 4.2cm 두께의 아크릴로 만든 모조 유방 내부에 microcalcifications과 비슷한 specks와 fibrous linear structure와 비슷한 fibers, 유방의 nodule 혹은 mass와 비슷한 masses가 포함되어 있는 팬텀을 촬영하여 각 모조병소들이 기준 점수(섬유소 4점, 미세석회화 그룹 3점, 종괴 3점, 총점 10점) 이상으로 잘 보이는지를 확인하는 검사이다. 팬텀영상에서 보이는 각 그룹의 점수는 같은 사람이 같은 판독 조건에서 평가했을 때에 이전과 비교하여 0.5이상 감소하지 않아야 하며, 만일 점수에 많은 변화가 있거나 불합격 판정이 내려진다면 유방촬영 영상의 화질에 영향을 미칠 수 있는 원인에 대해 조사하여 교정하는 작업이 필요하다 (6, 7).

본 연구에서는 과거 유방팬텀영상 검사에서 부적합 판정을 받았던 영상들을 후향적으로 분석하여, 부적합 판정을 받았던 증례들의 특징과 원인이 되었던 요소들을 알아보고자 한다.

대상 및 방법

본 연구는 유방촬영장치의 품질검사를 위해 2013년 1월부터 2016년 3월 사이에 한국의료영상 품질관리원에 접수되었던 팬텀영상들 중 최종 부적합 판정을 받았던 총 207례의 유방팬텀영상을 대상으로 하였다. 팬텀영상들은 한국의료영상 품질관리원의 서버에 압축되어 저장되어 있는 자료들을 백업 받아 원래 크기의 영상으로 복원한 후에 재평가되었고, 평가 위원들이 영상을 재평가할 당시에 모든 영상들이 부적합 판정이 나간 증례들임은 알고 있으나 어떤 항목 또는 항목들에서 부적합 판정을 받았는지, 각 항목의 점수는 몇 점이었는지는 모르는 상태에서 평가하였다.

유방팬텀영상의 평가는 품질관리 검사기관에 접수된 영상을 평가하는 과정과 동일하게 진행하였다. 먼저 유방촬영영상 품질관리와 유방영상 판독에 20년 이상의 경력을 가진 6명의 영상의학과 전문의가 2명씩 한 조로 나뉘어 팬텀영상들을 후향적으로 재평가하였고, 두 명의 판정 결과가 서로 일치하지 않은 경우에는 6명의 판정단 중 첫번째 판정에 참여하지 않는 3명이 모여 영상을 재검토한 후 5명의 다수결에 따라 최종 점수를 결정하였다.

팬텀영상에 대한 점수는 섬유소 병변의 경우, 모든 길이가 다 보이고 위치와 방향이 올바르면 1점, 모든 길이가 다 보이지는 않지만 절반 이상이 보이고 위치와 방향이 올바르면 0.5점, 절반이 안되거나 보이지 않으면 0점으로 판정하였다. 미세석회화 그룹 병변의 경우, 하나의 그룹을 구성하는 6개의 점들 중 4개 이상이 제 위치에서 보이는 경우 1점, 2-3개가 제 위치에서 보이는 경우 0.5점, 1개만 보이거나 안 보이는 경우에는 0점으로 판정하였다. 종괴의 점수는 원으로 인정할 수 있으면 1점을 주었는데, 원으로 인정할 수 있는 기준으로 종괴의 등근 변연이 3/4이상 보이고 제 위치에 있는 경우 1점, 종괴가 제 위치에 있으나 등근 모양을 보이지 않는 경우 0.5점, 종괴가 보이지 않는 경우 0점으로 판정하였다.

부적합 판정을 받았던 증례들의 점수들을 후향적으로 분석하여, 부적합을 받았던 항목과 점수의 분포를 알아보고 부적합 판정에 가장 결정적인 역할을 했던 병소는 무엇인지를 분석해보았다. 또한 두 명의 평가자 사이에 적합-

부적합에 이견이 있었다가 다섯 명의 다수결에 의해 최종 부적합 판정을 받는 비율과 두 명의 평가자들이 처음부터 모두 부적합으로 판정했던 증례들의 비율을 살펴보고, 평가자들 간의 적합-부적합에 이견이 가장 많았던 평가 항목과 그 비율을 분석하였다.

현재 정도관리 기록검사의 팬텀영상 채점표에는 포함되어 있지 않지만, 팬텀영상의 전반적인 화질과 인공물을 평가하기 위해 배경의 불균질이나 인공물의 정도를 판단하여, 없으면 0점, 약간 있으나 모조병소의 판정에는 지장이 없는 정도라면 1점, 모조병소의 판정에 지장이 있을만큼 심한 경우에는 2점으로 분류하여, 두 명의 평가자가 각각 기록하였고, 이 결과가 팬텀영상점수의 총점이나 특정 모조병소의 점수와 관련이 있는지를 분석해보았다. 상관 관계는 Pearson's correlation을 이용하여 분석하였다.

결 과

해당 기간동안 품질관리 검사를 위해 접수되었던 유방 팬텀영상은 총 6,912례였고, 그 중 부적합은 317례로, 전체 부적합율은 4.6%였다. 부적합 판정을 받았던 영상들을 의료기관의 종류별로 분류했을 때에는 종합병원급이 26개(8.2%), 병원급이 89개(요양병원 7개 포함)(28.1%), 의원급이 202개(63.7%)였다. 총 317례의 부적합 판정을 받았던 영상들 중 110례 (34.7%)는 필름 유방촬영영상으로 영상이 보존되어 있지 않았고, 서버에 백업이 되어있었던 Computed Radiography (CR) 및 Digital Radiography (DR)는 207례 (65.3%)였다. 이러한 이유로, 부적합 판정을 받았던 유방팬텀영상에 대한 후향적인 재평가는 207례의 CR과 DR 영상에 대해서만 이루어졌다.

섬유소, 미세석회화 그룹, 종괴의 모조병소 별로 기준 점수인 4점, 3점, 3점을 넘지 못해 부적합으로 판정된 경우는 각각 섬유소 179례, 미세석회화 그룹 5례, 종괴는 188례였고, 이들 중 섬유소와 종괴 두 모조병소에서 부적합 판정을 받은 경우는 119례가 있었고, 섬유소, 미세석회화 그룹, 그리고 종괴 세 개의 모조병소 모두가 부적합인 경우는 5례였다. 부적합 판정을 받은 증례들 중 종괴의 점수가 미달이었던 경우가 90.8%로 가장 많았고, 반면에 미세석회화 그룹의 점수가 미달이었던 경우는 2.4%로 가장 적었으며 모두 세 가지 병소들 전체에서 부적합이었던 경우였다.

적합-부적합 판정에 대한 평가자 간 일치도를 분석해 보았을 때에는 모든 207 증례들에 대해 두 명의 평가자들이 모두 부적합으로 판정하였고, 두 명 중 한 명이라도 적합

판정을 내렸던 증례는 없었다. 다만 각 모조병소별로 점수를 나누어 분석해 보았을 때에는 두 명의 평가자들 사이에 각 모조 병소별 판정이 모두 일치했던 경우는 125례(60.4%)였으며, 두 명의 평가자들 사이에 한 개의 병소에 대한 판정이 달랐던 경우가 66례(31.9%), 두 개의 병소에 대한 판정이 달랐던 경우는 15례(7.2%)가 있었고, 3개 병소 모두의 판정이 달랐던 경우는 0례였다. 207개의 팬텀 영상에 있는 총 621개의 모조 병소들에 대한 판정 불일치율을 계산했을 때에는 96/621 (15.5%)였고, 판정 불일치가 있었던 모조병소들 중에는 섬유소 병변이 42례(20.3%), 작은 석회화 그룹이 7례(3.4%), 종괴가 47례(22.7%)로 종괴에서의 불일치가 가장 많았다 (표1).

팬텀영상의 배경 불균질이나 인공물의 정도를 평가하는 점수에서는 배경이 균질하고 인공물이 없는 영상은 20례(9.7%)였으며, 나머지 97례(46.9%)와 90례(43.5%)에서는 판독에 지장은 없을 정도로 약간 있거나 또는 판독에 지장이 있을 만큼 심한 경우였다. 배경 불균질이나 인공물을 평가하는 점수와 모조병소들의 총점이나 병소별 점수 사이에는 상관 관계가 없었고 ($p=0.25$, $p=0.16$, $p=0.09$, $p=0.24$), 평가자들 사이에 판정 불일치가 있었던 모조병소들의 수와는 의미 있는 상관 관계를 보였다 ($p=0.027$).

고 찰

같은 기간동안 등록되어 있던 총 2,963대의 유방촬영장치들 중 종합병원이 489개(16.5%), 병원급이 691(요양병원 47개 포함)(23.3%), 의원급이 1,783개(60.2%)였던 것과 비교했을 때에 (8), 부적합 판정을 받았던 유방촬영장치들 중 의원급과 병원급에 설치된 장비의 비율이 63.7%, 28.1%로 8.2%인 종합병원급에 비해 높았던 것을 알 수 있다. 이는 대학병원 등이 속해 있는 종합병원급에서 영상 의학과 상근 전문의가 있고 특히 유방영상을 전공으로 하는 전문의들이 유방촬영장치의 품질관리에 관여하고 있을 가능성이 큼을 시사한다 (9). 또한 의원급이나 병원급에 신규 설치되는 장비들에 중고 장비가 많다는 사실도 이와 관련이 있을 가능성이 있다 (10). 실제로 한국의료영상 품질관리원에서 발간했던 10년사의 내용 중에도 의료기관 종별에 따른 검사 결과에서 전체 검사 의뢰 영상들 중 부적합으로 판정을 받았던 영상의 비율에서 의원급의 부적합율이 병원급이나 종합병원급에 비해 월등히 높았었다 (3.9% vs 3.3%, 1.6%) (11). 또한 부적합 판정을 받았던 장비 317개 중에 필름 유방촬영장치가 110개(34.7%)를 차지하였는데, 이는 같은 기간 전체 유방촬영장치들 중

표 1. 각 모조병소별 평가자 두 명의 부적합 판정 결과 (207 증례)

판독자 2	판독자 1									모두 (n=6)
	섬유소 (n=138) &			미세석회화 그룹 (n=7) &			종괴 (n=167) &			
	섬유소 (138)	미세석회화 그룹 (7)	종괴 (98)	섬유소 (7)	미세석회화 그룹 (7)	종괴 (6)	섬유소 (98)	미세석회화 그룹 (6)	종괴 (167)	
섬유소 (n=178) &										
섬유소 (178)	135	7	95	7	7	6	95	6	138	1
미세석회화 그룹 (7)	5	3	0	3	3	3	4	3	6	2
종괴 (143)	111	6	83	6	6	6	83	5	116	2
미세석회화 그룹 (n=8) &										
섬유소 (7)	5	3	6	3	3	3	4	3	6	0
미세석회화 그룹 (8)	5	3	6	3	3	3	4	3	6	0
종괴 (8)	5	3	6	3	3	3	4	3	6	0
종괴 (n=173) &										
섬유소 (137)	111	6	83	6	6	6	83	5	116	2
미세석회화 그룹 (1)	5	3	4	3	3	3	4	3	7	0
종괴 (173)	114	6	86	6	6	5	86	5	145	0
모두 (n=7)	1	0	1	0	0	0	1	0	2	3

굵은 글씨 = 각 병소에 대한 부적합 판정이 정확히 일치했던 경우

필름 유방촬영장치가 차지하는 비율(평균 27.7%, 20.4%, 30.6%)보다 훨씬 높은 수치로, 필름 유방촬영장치에서의 부적합율이 CR이나 DR 유방촬영장치에서 보다 높다는 것을 알 수 있다 (12). 이는 오래된 중고 장비들 중 필름 유방촬영장치가 많았을 가능성도 있지만 또한 필름 유방촬영장치가 CR이나 DR 장비에 비해 품질관리에 더 어려움을 겪고 있을 가능성을 생각해 볼 수 있다.

부적합으로 판정이 나갔던 증례들에 대해 재재점을 했던 후향적 평가에서도 적합 판정이 단 한 건도 없었던 것을 고려하면, 하나의 영상을 두고 그 영상이 적합한지 부적합인지를 판정하는 판정에 대한 신뢰도나 재현성은 매우 높은 것으로 생각된다. 그러나 평가자 별로 부적합으로 판정한 근거가 되는 모조 병소들의 점수가 완전히 일치하는 경우는 60.4% (병소 기준으로는 84.5%)에 그쳐, 좀 더 객관적이고 일관성 있는 판정 결과가 요구되는 상황이다. 이는 부적합 판정을 받은 의료기관에서 부적합 결과에 대한 의문의 여지를 남길 수 있고, 또한 어떤 부분을 개선해야 적합한 영상을 얻을 수 있는 지 그 기준을 파악하기 어렵게 할 수 있다. 특히 종괴 병변에 대한 판독자 간의 일치도가 비교적 낮게 나온 부분은 ‘원으로 인정할 수 있는가 없는가’ 하는 종괴 병변에 대한 판정 기준이 다소 주관적일 수 있다는 부분에 기인한다고 볼 수 있다 (3, 13).

미국 영상의학회 등 다른 나라의 기준에는 종괴 등의 모조병소와 비슷하게 보이거나 모조 병소의 판정에 영향을

줄 수 있는 주변 영상의 아주 미세한 불균질성이나 인공물이 있는 경우 팬텀영상평가의 점수에서 감점을 하는 요소가 있으나, 유감스럽게도 우리나라의 평가 시스템은 처음 유방촬영영상에 대한 품질관리가 도입되던 시절에 품질관리 자체에 대한 의료기관의 참여를 높이고 좀 더 간단한 준비만으로 시작할 수 있도록 이러한 감점 사항까지는 포함하지 않았다 (2, 13). 정도관리 항목들에 대한 이행을 잘 하지 않거나 또는 너무 오래되어 교체가 필요한 이미징 플레이트를 사용하는 CR 장비의 경우에 배경 농도의 불균질이나 인공물의 발생을 유발하게 되므로, 정도관리항목 점검표가 잘 기록되어 있더라도 팬텀영상이나 임상영상에서 배경이 불균질하고 인공물이 보인다면 정도관리항목 점검표의 기재 내용에 대한 신뢰도는 떨어질 수 있다. 현재 유방촬영영상의 임상영상평가표에는 잡영과 인공물이라는 항목을 통해 전체 100점 중 14점의 점수를 할애하고 있으며, 판독에 지장을 줄 정도로 심한 경우에는 추가로 5점을 감점하여, 총 19점의 불이익을 받도록 하고 있다. 유방촬영영상에 대한 품질관리가 시작되고 처음 사용하던 임상영상평가표에서 잡영과 인공물에 16점을 할애했던 것보다 늘어난 비율로, 이는 100점 중 60점 이상이 적합한 기준을 고려하면 인공물과 배경 잡영만으로도 그 중 절반에 해당하는 점수를 잃을 수 있도록 하고 있는 셈이다 (14). 팬텀영상평가의 부적합 사례들을 분석한 본 연구의 결과에서 배경이 불균질하고 인공물이 많은 영상일수록

관독자 간의 적합, 부적합에 대한 판정의 일치도가 의미 있게 낮았던 것을 고려하면, 배경 불균질과 인공물의 여부 또는 정도는 추후 팬텀영상평가 항목에 추가되어야 할 것으로 판단된다.

본 연구에는 몇 가지 제한점이 있다. 첫째로 본 연구는 3년 3개월의 기간동안 접수되었던 유방팬텀영상들 중 부적합으로 최종 판정되었던 증례들의 CR, DR 영상들의 병변들만을 분석한 것이다. 모조병소별 부적합율이나 인공물, 평가자 간의 일치도 등을 적합으로 판정되었던 영상들과 비교 분석한 것이 아니기 때문에, 그들이 부적합에 영향을 주었다는 결론을 내릴 수는 없다. 두번째, 본 연구에는 필름 유방촬영장치에서 얻은 영상은 포함되지 않았고, CR과 DR 장치의 팬텀영상만을 분석한 것이다. 필름 영상들은 보존이 어렵고, 현재 필름을 볼 수 있는 판독대가 남아있는 의료기관이 없었다. 본 연구의 결과는 CR과 DR 유방촬영영상에 한하여 적용될 수 있겠다. 세번째로 후향적으로 유방팬텀영상을 평가했던 평가자들은 영상들이 과거에 부적합으로 판정이 나간 사실을 인지하고 있었다. 물론 어느 병소의 점수에서 부적합을 받았는지, 병소 각각의 점수가 몇 점이었는지는 알지 못했고, 팬텀영상평가 기준을 엄격하게 적용하여 기준대로만 평가할 것을 요구받았으나 이러한 환경은 평가자들의 점수에 영향을 미쳤을 가능성이 있다.

결론적으로, 부적합 판정이 가장 많았고 적합-부적합 판정에 대한 평가자별 불일치 또한 가장 많았던 종괴 병소를 중심으로 한 영상 품질관리나 검사자들의 눈높이 교육, 또는 채점 기준의 보완이 더 필요할 것으로 보인다. 또한 배경 불균질이나 인공물이 많은 영상에서 평가자들 간의 불일치도가 높아지는 경향을 보여, 추후 이를 평가할 방법을 추가하는 것이 판정 일치도를 높이고 더 객관적이고 공정한 평가를 가능하게 할 것으로 생각된다.

참 고 문 헌

1. 2022 유방암백서. 한국유방암학회.
2. 특수의료장비의 설치 및 운영에 관한 규칙, 보건복지부. 2003.01.
3. 고은영. Mammography Quality Standard Act in Korea. 대한유방검진학회지 2005;2:196-204.
4. 문우경, 김태정, 차주희 등. 유방촬영술의 임상영상평가: 전국조사. 대한방사선의학회지 2003;49:507-511.
5. 이은혜, 정승은. 2011년 국가암검진기관평가: 유방촬영 임상영상 화질평가에서 나타난 영상의학과 의사의 역할. 대한유방검진학회지 2011;8:1-5.
6. 특수의료장비 품질관리 길라잡이 2013; 81-83.
7. 고은영. 유방영상 품질관리. 제 7차 대한유방검진학회 춘계학술대회 초록집 90-95.
8. 의료자원(인력/시설/장비) 현황(2018년 이전). HIRA 빅데이터 개방 포털. <https://opendata.hira.or.kr/op/opc/olapMdclRcStatsInfoTab9.do?docNo=02-009>
9. 이은혜, 정승은. 2011년 국가암검진기관평가: 유방촬영 임상영상화질평가에서 나타난 영상의학과 의사의 역할. 대한유방검진학회지 2011;1:99-108.
10. 백상현, 김승훈, 이은혜, 황지영, 전민숙, 김광균. 특수 의료장비 영상품질 평가 및 수가보상 연계를 위한 연구 2012.
11. 한국의료영상품질관리원 10년사 2015; 36.
12. 송수연, 홍세리, 전재관. 국가암검진사업에서 디지털 유방촬영술의 현황과 과제. 대한영상의학회 2021;82(1):2-11.
13. ACR. Mammography Quality Control Manual 1999;181-190.
14. 백문희, 김학희, 손은주 등. 유방촬영술 임상영상검사 평가: 기존 임상영상평가표와 신 임상영상평가표의 비교. 대한유방검진학회지 2006;3:111-120.

A Study of Noncompliance Cases in Mammography Phantom Image Evaluation

Eun Young Ko¹, Ok Hee Woo², Seon Hyeong Choi³, Jung Min Chang⁴, Hye Won Kim⁵, You Me Kim⁶,
Myoung Kyoung Kim¹

¹Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine,

²Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine,

³Queens' U Clinic, ⁴Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, ⁵Department of Radiology, Wonkwang University Hospital, Wonkwang University School of Medicine,

⁶Department of Radiology, Donggunsan Hospital

Purpose: This study aims to analyze breast phantom images that were deemed noncompliant and to identify the characteristics and contributing factors of such cases.

Materials and Methods: A total of 207 noncompliant breast phantom images submitted to the Korean Institute of Medical Imaging Quality Assurance between January 2013 and March 2016 were retrospectively analyzed. Six breast radiologists with over 20 years of experience in mammography quality control were grouped into three pairs to re-score the phantom images based on scoring criteria. In cases of disagreement between two evaluators, three other experts from the panel reviewed the images, and the final score was determined by majority vote. The level of background heterogeneity or artifacts in the images was classified as 0 (none), 1 (mild), or 2 (severe). The phantom lesions that caused noncompliance were identified, and the proportions of each lesion type were investigated. Additionally, the evaluation categories with inconsistent compliance decisions among evaluators and their rates of disagreement were analyzed. The relationships between background heterogeneity or artifacts and lesion scores, compliance decisions, and inter-evaluator agreement were assessed using Pearson's correlation analysis.

Results: The lesion with the highest rate of noncompliance was masses (90.85%), while microcalcifications had the lowest rate (2.4%). The overall agreement on compliance decisions for each case was 100%, but complete agreement across all evaluation items was achieved in only 60.9% of cases. Among 621 phantom lesions analyzed, the overall disagreement rate across evaluation items was 15.5%, with masses showing the highest disagreement rate. The severity of background heterogeneity or artifacts was positively correlated with higher disagreement rates ($p=0.027$).

Conclusion: Noncompliance rates and inter-evaluator disagreement were highest for masses, indicating a need for improved scoring criteria and education for good imaging. Incorporating evaluation methods for background heterogeneity and artifacts may enhance consistency and contribute to more objective assessments.

Index words: Breast phantom, Mammography quality control, Phantom image evaluation, Noncompliance cases

Corresponding author: Eun Young Ko, M.D., Ph.D.