



Accuracy of mammography and magnetic resonance imaging to diagnose underlying malignancy in Paget's disease of the nipple: a systematic review and meta-analysis

Karen-Lise Madsen^{1#}, Anne D. H. Mosebo^{1#}, Sören Möller², Birthe H. Pedersen³, Camilla Bille¹

¹Department of Plastic Surgery, Odense University Hospital, Odense, Denmark; ²Open Patient data Explorative Network, Odense University Hospital and Department of Clinical Research, University of Southern Denmark, Odense, Denmark; ³Department of Radiology, Odense University Hospital, Odense, Denmark

Contributions: (I) Conception and design: KL Madsen, ADH Mosebo, C Bille; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: KL Madsen, ADH Mosebo; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Anne D. H. Mosebo, Department of Plastic Surgery, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense C, Denmark. Email: annemosebo94@hotmail.com.

Background: Paget's disease (PD) of the nipple is a less common type of breast cancer. Often it is accompanied by underlying malignancy and identification of this prior to surgery plays an important role in the surgical assessment. Until now no agreement about a standard diagnostic program exists. A few smaller studies of the accuracy of mammography (MMG) and magnetic resonance imaging (MRI) to determine presence and extent of underlying malignancy in patients with PD have been published. The aim of this study was to pool the published evidence with a systematic review and meta-analysis.

Methods: EMBASE, Medline, Cochrane Library, Google Scholar, ClinicalTrials.gov and WHO International Clinical Trial Registry Platform were searched, and 719 published articles identified. Only six fulfilled the inclusion criteria; were prospective/retrospective, included patients with biopsy-proven PD and had preoperative MMG and/or MRI performed. Findings from diagnostic imaging were correlated to postoperative histopathology to determine sensitivity and specificity. Risk of bias and applicability were assessed independently by the two authors using QUADAS-2 which is divided into patient selection, index test, reference standard and flow and timing.

Results: The six studies included 125 patients combined. A total of 73 and 47 patients remained for pooled calculations for MMG and MRI, respectively. The total sensitivity [95% confidence interval (CI)] of MMG and MRI was 39% (95% CI: 27–51%) and 68% (95% CI: 52–81%), respectively. The difference was significant ($P=0.0025$). The total specificity (95% CI) of MMG and MRI was 100% (95% CI: 54–100%) and 100% (95% CI: 29–100%), respectively. Risk of bias assessment indicated an intermediate risk of bias in the patient selection, and intermediate or high in the index test. Both the reference standard and flow and timing had a low risk of bias.

Discussion: Although a small sample and wide 95% CIs, this systematic review and meta-analysis propose that MRI is significantly superior in diagnosing underlying malignancy in patients with PD compared to MMG alone. It therefore seems to improve decision making in surgical assessment of PD.

Keywords: Magnetic resonance imaging of the breast; mammography; Paget's disease of the nipple; underlying malignancy in the breast

Received: 10 August 2021; Accepted: 30 November 2021; Published online: 06 December 2021.

doi: 10.21037/abs-21-95

View this article at: <https://dx.doi.org/10.21037/abs-21-95>

Introduction

Paget's disease (PD) of the nipple is less common and accounts for 1–3% of all mamma cancers (1–3). PD is primarily seen in postmenopausal women with a peak incidence in the sixties and seventies (4). Symptoms such as eczema, erythema, ulceration, bleeding, or scaling are the most common and can be the only clinical finding. These symptoms are similar to dermatitis or other benign conditions that often result in a delayed diagnosis (3,5).

Multiple studies suggest that up to 90% of the cases of PD are associated with underlying malignancy in the mamma (2,6–8). Three factors have been proven to affect the prognosis of PD: (I) the axillary lymph nodes status, (II) if the histopathology identifies ductal carcinoma in situ (DCIS) or invasive ductal carcinoma (IDC) and (III) if the clinical examination identifies a palpable mass. Patients presenting with a palpable mass have IDC in up to 90% of the cases (7).

Since the majority of PD cases are associated with an underlying malignancy, the epidermotropic theory is the most widely accepted theory regarding pathogenesis. This theory proposes a migration of tumour cells originating from DCIS or IDC in the mamma through the lactiferous ducts to the epidermis on the nipple-areolar complex (NAC) (2,3,6). Histologically, the cells are seen as characteristic big, round so-called Paget cells (2). The second theory, the transformation theory, proposes that PD cells are transformed in situ keratinocytes of the nipple epidermis (5). This theory is supported by the fact that in a few cases no underlying malignancy is associated (2).

PD is generally diagnosed by a skin biopsy from the NAC and provides information on the presence of disease on the NAC (3,5). Mammography (MMG) and magnetic resonance imaging (MRI) are used to evaluate underlying malignancy and the extent of the disease. The extent of the disease is decisive for the surgical strategy. Patients are offered either breast conserving surgery (BCS) with radiotherapy, or mastectomy depending on the extent of the disease. BCS should not be offered if diagnostic imaging suggests widely spread disease and therefore the accuracy of the diagnostic imaging plays an important role (8). It has previously been indicated that MMG has a lower sensitivity compared to MRI when investigating pathological changes in association with PD (3). Most studies are based on a small sample size and lack statistical power. The aim of this study was to systematically review the literature analysing the accuracy of MMG and MRI to determine

underlying malignancy in patients with PD. We present the following article in accordance with the PRISMA reporting checklist (available at <https://abs.amegroups.com/article/view/10.21037/abs-21-95/rc>) (9).

Methods

Eligibility criteria

Eligible studies were included according to the PIRD criteria. This question format has been used as this systematic review would be characterised a Diagnostic Test Accuracy Review and the aim was to determine how well MMG, and MRI worked in terms of sensitivity and specificity for the diagnosis of PD (10). Both prospective and retrospective studies were included. The population was patients with biopsy-proven PD as skin-biopsy from the NAC is considered the golden diagnosis of malignancy. All included patients had surgery performed. In addition, the patients had MMG (reference test) and/or MRI (index test) done prior to the operation. The diagnostic imaging provided information on the extent of disease which was the diagnosis of interest. The studies excluded patients with a history of mamma cancer. There were no restrictions according to setting, time frame, language, publication status or definitions of outcomes in the included studies.

Search strategy and study selection

A search was performed using EMBASE, Medline and Cochrane Library. Furthermore, the first 100 articles on Google Scholar were screened using a simple search. The last search was performed on October 14, 2020. Reference lists of identified studies were also explored. ClinicalTrials.gov and WHO International Clinical Trial Registry Platform were searched to identify possible ongoing/unpublished studies regarding the subject of interest.

The search was based on both subject headings and free text words. The search consisted of the terms: Paget nipple disease and MRI, or variations of these terms depending on the database. The aim of the search was to secure a high sensitivity. Since PD is a less common disease and therefore a small amount of literature exists, the search was only based on two blocks ([Appendix 1](#)). No limiting filters were used.

The two screening authors were independently active in the full collection process. Any disagreements were

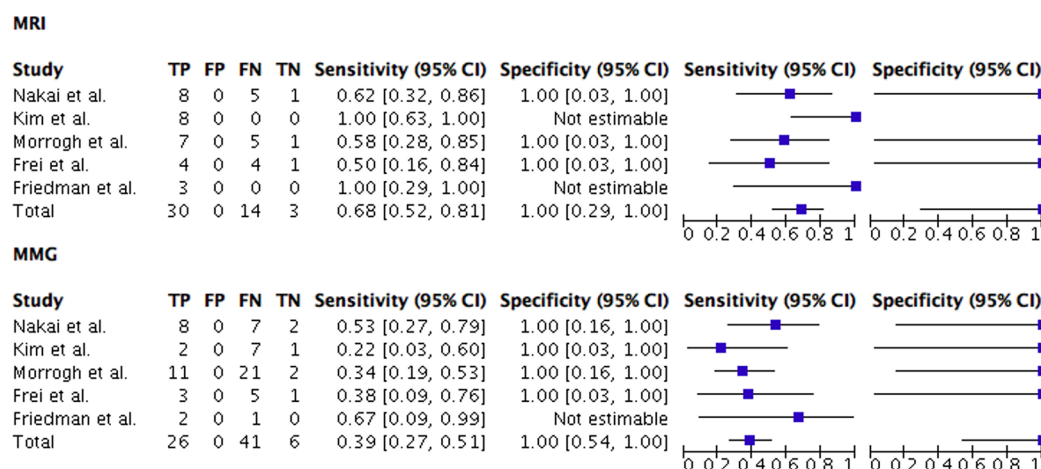


Figure 1 Results and forest plot depicting sensitivity and specificity for magnetic resonance imaging and mammography in a systematic review and meta-analysis investigating the two imaging methods as a diagnostic tool to examine underlying malignancy in patients with Paget's disease of the nipple. A total sensitivity and specificity were calculated for both magnetic resonance imaging and mammography. MRI, magnetic resonance imaging; TP, true positive; FP, false positive; FN, false negative; TN, true negative; 95% CI, 95% confidence interval; MMG, mammography.

discussed and resolved in consensus by the two authors. Covidence software was used for title and abstract screen, full-text screen and data extraction (11).

Risk of bias in individual studies

The two authors independently used QUADAS-2, a quality assessment tool for diagnostic accuracy studies, to evaluate the risk of bias and applicability of the six studies (12). The tool consists of four key domains (patient selection, index test, reference test and flow and timing). In this study MMG or MRI were defined as the index test, whereas histopathologic diagnosis was defined as the reference test. Signalling questions were added to adjust the tool to the included studies, which were finally evaluated with low risk of bias, intermediate risk of bias or high risk of bias (Appendix 2). Any disagreements were discussed and resolved in consensus by the two authors.

Synthesis of results

The following parameters were reviewed; year of publication, country of origin, period of inclusion, size of study population, age of the included patients, inclusion and exclusion criteria, pathological findings and number of patients undergoing diagnostic tests.

Siponen *et al.* only provided sensitivity, whereas all other studies also provided the supporting counts. Therefore, the

corresponding author of Siponen *et al.* was contacted. Since there was no response, it was not possible to include this data in the analysis (13).

Friedman *et al.* included three patients with PD and descriptions of clinical, MRI, histopathological and mammographic findings were presented in a table (14). Based on these descriptions, the authors of this review classified which patients had a true positive and true negative test. This was determined based on the degree of malignancy and whether the imaging findings would change the surgical assessment.

Altogether the data was used to estimate a pooled sensitivity and specificity, which is illustrated in a forest plot (Figure 1).

Statistical analysis

Data from each study was organised in a 2x2 contingency table and used to estimate sensitivity and specificity with associated exact binomial 95% CI of MMG and MRI. Pearson's chi-square test was used to compare sensitivity between MMG and MRI. While a paired McNemar test in principle would have achieved higher power, this was infeasible due to unavailability of individual paired data in the included studies (15). A two-sided P value <0.05 was considered statistically significant.

The meta-analysis was performed using the Review Manager 5.4 program developed and supported by the

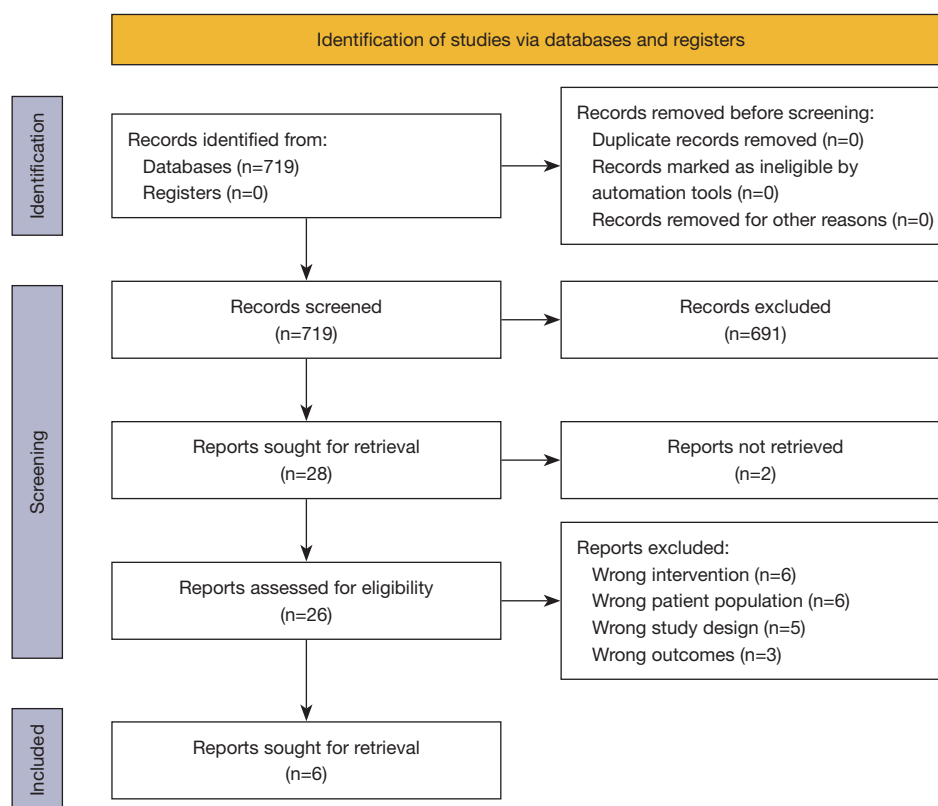


Figure 2 PRISMA flowchart showing the selection of articles in a systematic review and meta-analysis investigating mammography and magnetic resonance imaging as a diagnostic tool to examine underlying malignancy in patients with Paget's disease of the nipple.

Nordic Cochrane Centre (16).

Results

Study selection

A flowchart of the study selection is shown in *Figure 2*. A total of 719 citations were found from the databases. By removing duplicates and irrelevant articles on title screening, 28 articles remained for more detailed evaluation. One unpublished article was found, but it was not possible to contact the corresponding author (17). Furthermore, the corresponding author of a published but unavailable article was contacted without response (18). Twenty-six articles remained for full-text screening and six of these full-filled the inclusion criteria.

Study characteristics

Characteristics of the individual studies can be found in *Table 1*. The six studies originated from Switzerland, Korea,

Japan, USA, UK, and Finland. Only retrospective studies were found. The data collection was obtained between 1994 and 2018. A total of 125 distinct patients were included in this systematic review, with a minimum of three and a maximum of 52 patients included in the individual studies.

Most studies included patients retrospectively via databases from the hospital at which the patients were treated. Kim *et al.* found patients by reviewing pathology databases without further specification (19). Friedman *et al.* and Frei *et al.* did not specify how the patients were collected (14,20).

Most studies only included patients with biopsy-proven PD, though Friedman *et al.* included patients with primary mamma cancer, whereas three patients were diagnosed with PD (14).

Not all studies had the number of true positives and true negatives of the imaging tests as the primary outcome. When this was not the case, it was possible in all, but one, to extract this data.

All patients were women and the median age varied from 57 to 70 years.

Table 1 Characteristics of the six included studies in a systematic review and meta-analysis investigating mammography and MRI as a diagnostic tool to examine underlying malignancy in patients with Paget’s disease of the nipple

Reference	Year	Country	Period of inclusion	Population, n	Age, median [range], years	Inclusion criteria	Exclusion criteria	Pathological finding	Diagnostic test, N		Outcomes			
											True positive, n		Sensitivity, %	
									MRI	MMG	MRI	MMG	MRI	MMG
Nakai <i>et al.</i>	2019	Japan	2008–2018	17	64 [†] [48–79]	Patients diagnosed with PD either with skin or needle biopsy	Patients who presented with a palpable mass	DCIS, n=10; IDC, n=5; no underlying malignancy, n=2	14	17	8	8	61.5	53.3
Kim <i>et al.</i>	2010	Korea	2003–2005	10	59.5 [47–75]	Patients with biopsy-proven PD	None	DCIS, n=7; IDC, n=2; no underlying malignancy, n=1	8	10	8	2	100	22
Siponen <i>et al.</i>	2010	Finland	1995–2006	52 [58] [§]	64 [32–95]	Patients diagnosed with Paget’s disease of the breast treated with primary surgical therapy	Patients with distant metastasis	DCIS, n=22; IDC, n=31; microinvasive cancer, n=3; no underlying malignancy, n=2	14	52	–	–	44, 100 [‡]	39, 79 [‡]
Morrogh <i>et al.</i>	2008	USA	1995–2005	34	63 [33–83]	Patients with biopsy-proven PD; only physical finding was nipple changes suspicious for PD	Patients who presented with a palpable mass; patients with ipsilateral breast cancer; patients who have declined surgery	DCIS, n=19; IDC, n=7; DCIS with microinvasion, n=6; underlying intraductal papilloma with foci of atypical ductal hyperplasia, n=1; no underlying malignancy, n=1	13	34	7	11	58.3	34.38
Frei <i>et al.</i>	2005	Switzerland	1995–2004	9	57 [36–76]	Patients with biopsy-proven PD	None	DCIS in the underlying lactiferous ducts of the NAC and associated DCIS or IDC elsewhere in the breast, at least 20 mm away from the NAC, n=4; Paget disease of the nipple with associated DCIS in the underlying lactiferous ducts of the NAC, n=4; no underlying malignancy, n=1	9	9	4	3	50	37.5
Friedman <i>et al.</i>	1997	UK	1994–1996	3	70 [41–74]	Women with primary breast cancer diagnosed by triple assessment (clinical examination, FNAC and MMG) and surgically treated by mastectomy. Patients underwent MRI and MMG	None	DCIS, n=1; IDC, n=2	3	3	3	2	100	66.7

[†], mean age; [‡], DCIS, IDC; [§], 6 patients were excluded: 2 patients had only ultrasound performed, while information regarding breast imaging was missing for 4 patients. MRI, magnetic resonance imaging; MMG, mammography; PD, Paget’s disease of the nipple; FNAC, fine needle aspiration cytology; DCIS, ductal carcinoma in situ; IDC, invasive ductal carcinoma; NAC, nipple-areolar complex.

It is assumed that all women went through approximately the same assessment according to time between diagnostic imaging and histopathological diagnosis since data was extracted retrospectively from patient journals.

Classification of diagnostic imaging

The imaging findings were interpreted differently. In Kim *et al.* MMGs were retrospectively reviewed by two breast radiologists assessing the nipple and parenchymal (microcalcifications, mass/asymmetry, and ductal dilatation) abnormality (19). MMGs were reviewed and correlated with MRI findings. The MRIs were reviewed according to shape and enhancement of the involved nipple and compared with the opposite mamma.

Frei *et al.* used two radiologists to retrospectively review both MMG and MRI in consensus (20). The study did not specify in more detail how MMGs were assessed. According to MRI the enhancing lesions were recorded, and the degree were rated qualitatively and relative to the signal increase in adjacent blood vessels as absent, low, mild, moderate, or strong. A predefinition of a normal and abnormal nipple enhancement was also described.

In Morrogh *et al.* MMGs were reviewed according to BI-RADS, breast density, calcifications and retro-areolar thickening (21). MRIs were reviewed by radiologists according to previously pre-described protocols. The authors had predefined how patients would be divided according to presence and extent of disease.

Friedman *et al.* gave no information as to how the MMGs were performed or reviewed. MRIs were retrospectively reviewed by two radiologists together who were blinded to the clinical, mammographic, and histopathological findings. Friedman *et al.* had predefined how the nipple morphology, enhancement characteristics of NAC, retro-areolar tissue and presence of any underlying enhancement were assessed (14).

Siponen *et al.* and Nakai *et al.* poorly described the interpretations of diagnostic imaging (13,22). Siponen *et al.* mentioned that only negative MRIs were re-evaluated.

Follow-up

Four out of six studies registered the disease-free follow-up time. In Siponen *et al.* the median follow-up period was 52 months (range, 1–158 months) (13). In Frei *et al.* the follow-up time was 60 months (range, 2–96 months) for all patients (20). In Morrogh *et al.* the median follow-up time was 37.2 months (4.4–100.4 months) for the patients, who

had BCS (21). The patients with no underlying malignancy had a median follow-up time of 83.5 months (70–97 months). Nakai *et al.* had a median 60-month observation period for all patients (22). Friedman *et al.* had no follow-up time since all patients had mastectomy performed (14). Kim *et al.* did not specify a follow-up period (19).

Risk of bias in individual studies

It was not possible to find the protocols of any of the included studies.

Not all patients had both MMG, and MRI performed. In some cases, MRI was not yet an available imaging tool. In other cases, the decision to use MRI was at the discretion of the surgeon.

This was the case in Siponen *et al.*, who included patients from 1995–2006. MRI was not introduced until 1999 for patients with negative findings on MMG and ultrasound in Finland. Furthermore, Siponen *et al.* only presented sensitivity without specifying the data of which the values were calculated (13).

Not all radiologists in the studies were blinded to histopathology and some studies only used one radiologist.

Risk of bias between studies

Assessment of risk of bias and concerns of applicability is presented in *Figures 3,4*, which is based on QUADAS-2 table ([Appendix 3](#)).

There was an intermediate risk of bias in the patient selection in all studies since some studies were unclear as to how patients were enrolled. Furthermore, four studies did not account for the reason why some patients only had one of the two imaging techniques performed.

Risk of bias of the index test was intermediate or high. Multiple reasons were considered to explain this trend. Some studies did not have a specified method to evaluate the index test. Also, some studies did not have two radiologists to independently evaluate the images. Lastly, not all radiologists were blinded. In general, the risk of bias of the reference standard and flow and timing was low.

There were some applicability concerns regarding patient selection and index test. This was due to lack of specification regarding the methods used in the studies. In general, there was little concern regarding applicability of the reference standard since the studies described histopathology well. Also, there seem to be more specific guidelines to assess histopathology in general.

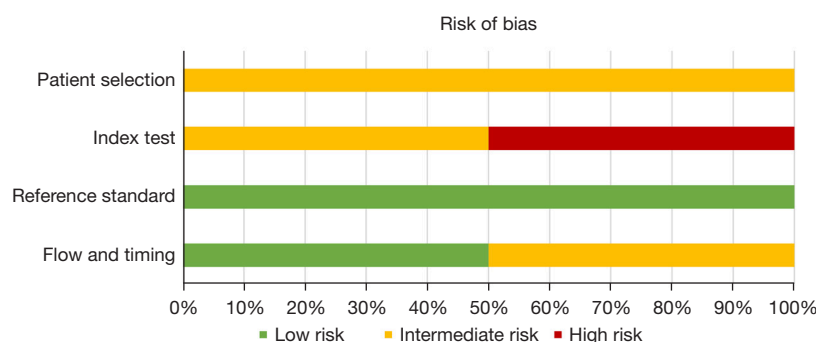


Figure 3 Figure illustrating risk of bias in the included articles in a systematic review and meta-analysis investigating mammography and magnetic resonance imaging as a diagnostic tool to examine underlying malignancy in patients with Paget's disease of the nipple. In accordance with QUADAS-2, four key domains (patient selection, index test, reference test and flow and timing) were assessed. Risk of bias was categorized as either; low risk of bias, intermediate risk of bias or high risk of bias.

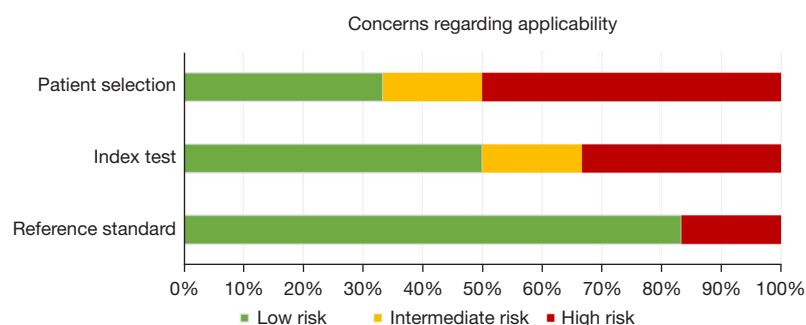


Figure 4 Figure illustrating evaluation of concerns regarding applicability in the included articles in a systematic review and meta-analysis investigating mammography and magnetic resonance imaging as a diagnostic tool to examine underlying malignancy in patients with Paget's disease of the nipple. In accordance with QUADAS-2, three key domains (patient selection, index test and reference test) were assessed. Risk of bias was categorized as either; low risk of bias, intermediate risk of bias or high risk of bias.

Synthesis of results

A total of 125 patients in the six included studies had MMG and/or MRI performed. Since Siponen *et al.* did not specify the underlying data, only 73 patients from the remaining studies were included in the forest plot (13). All 73 patients had MMG performed, whereas a total of 47 patients had an additional MRI performed.

The total sensitivity (95% CI) of MMG and MRI was 39% (95% CI: 27–51%) and 68% (95% CI: 52–81%), respectively. The difference was significant ($P=0.0025$). The total specificity (95% CI) of MMG and MRI was 100% (95% CI: 54–100%) and 100% (95% CI: 29–100%), respectively.

Discussion

This study demonstrated that MRI is a significantly more

accurate tool to diagnose underlying malignancy in patients with PD compared to MMG.

In general, the included studies had wide 95% confidence intervals of the sensitivities which is to be expected as the studies include small sample sizes. Despite this uncertainty, the confidence intervals of the total sensitivities do not overlap.

Some studies excluded patients with a palpable mass whereas other studies included all patients diagnosed with biopsy-proven PD. This affects the sensitivity of breast imaging, since the sensitivity improves when there is a palpable mass by examination (8).

The calculations for specificity of MMG and MRI only included six and three patients, respectively. A specificity of 100% was to be expected since the reference test has a high diagnostic accuracy, and it would be highly unlikely that

imaging would find any underlying malignancy not found in histopathology.

None of the included studies had the ideal setup for evaluating the diagnostic images. A higher sensitivity might have been achieved if the included studies had included two radiologists to review independently and blinded to histopathology diagnosis.

Since the imaging findings were interpreted differently among the included studies there is a risk of bias. Had all studies systematically used predefined data systems, such as BI-RADS, to review the images, the external validity of the studies would have been higher. Predefined data systems are generally essential since the analysis of diagnostic images depends on the experience of the individual radiologist.

Kim *et al.* correlated MMG with MRI but not with histopathology diagnose, which requires that the MRI diagnosis was correct (19). This might not be a problem since no cases of the six studies found that MMG diagnosed correctly in a case where MRI did not.

Not all studies specified to what extent MMG, and MRI were able to show the correct size and extent, but only if the imaging found the underlying malignancy or not. This is especially important in cases where the extent of disease would change the treatment strategy. In general, BCS is offered when the underlying malignancy is located in the central retro-areolar area. Mastectomy is offered in cases where underlying malignancies are found to be multifocal.

Siponen *et al.* only re-evaluated MRIs with negative findings by breast radiologists (13). Also, the assessment was done by general radiologists until 2006 and breast radiologists thereafter. This reduces the risk of false negative results but does not manage the risk of false positives.

The registration of the follow-up period is important to verify whether the initial diagnosis was correct, and no underlying malignancies were overlooked. The studies had a follow-up time of approximately three to seven years. Since the risk of recurrence is most prevalent within the first five years after the end of treatment, the studies had an acceptable follow-up.

Since some BCS-patients were not offered postoperative radiotherapy the risk of recurrence could be slightly more likely in this group of patients.

The risk of publication bias was difficult to evaluate since it was not possible to find any protocols of the included studies. To avoid publication bias different databases were searched to find ongoing/unpublished trials. Two studies were found; one unpublished and one unavailable, though

both abstracts indicated the same trend as our results (17,18). The two studies included 22 and 25 patients with PD, respectively. Due to our focus on sensitivity and specificity, contrary to classical hypothesis testing, more formal assessment of publication bias, e.g., funnel plots, was infeasible.

All included studies were retrospective. This implies a risk of information bias and misclassification bias. There is also a risk of selection bias since only a selected group of patients had MRI performed as part of the medical examination and at the discretion of the surgeon. An ideal design would have included a randomization of the diagnostic imaging assessment. However, this approach would be unethical and hence not a feasible study design in cancer patients.

The period of data collection varied in length of time and year of start. Furthermore, all studies originated from different countries and hospitals. These factors increase the likelihood of varying availability of breast MRI, outdated equipment, and inadequate standardized guidelines. As an example of inadequate standardized guidelines, some countries use ultrasound and clinical examination in addition with MMG in the initial phase of diagnosing mamma cancer while others use MMG alone. The meta-analysis of Helme *et al.* indicated that ultrasound in addition with MMG was able to improve the overall detection of underlying cancer compared to MMG alone (8). Therefore, it is important to distinguish between these different types of diagnostic investigation. In general, these factors affect risk of bias in individual studies as well as the applicability, as they indicate a possible high heterogeneity among the included studies. However, this can also be an advantage since it increases generalizability of results to a heterogeneous population.

A general limitation when investigating PD is the small sample size. This is difficult to avoid since the disease is less common. Also, the two-sided P value was calculated under the assumption that the studies included two independent groups. However, the included population in this study had in some cases both types of diagnostic imaging performed. This is a limitation of this study.

This study found a varying risk of bias. The high risk of bias was primarily seen in the applicability concerns. Only a small amount of literature exists on this topic and therefore these included studies were the only ones fitting the review question. This is to our knowledge the first systematic review and meta-analysis on this topic.

It is broadly accepted that MRI has a high rate of false

positives, which makes it less useful in screening for mamma cancer, but useful in examining high risk patients. PD patients should be considered high risk patients, since up to 90% are found to have an underlying malignancy (2,6,7). Regarding the treatment strategy the meta-analysis of Helme *et al.*, including 43 studies, concluded that BCS and radiotherapy is equally efficient compared to mastectomy when examining local control and survival in early mamma cancer (8). However, there are several factors to consider when deciding the surgical assessment such as breast size and the cosmetic outcome. Therefore, BCS might not always be the optimal choice.

To identify the group of patients with underlying malignancy, an accurate diagnostic tool plays an important role. The data of our study demonstrated that if MMG had a positive finding, MRI was unlikely to change the strategy of treatment and if MMG had a negative finding, MRI proved to be a useful additional tool. This demonstrates the role of both MMG and MRI in finding underlying malignancy in patients with PD.

Altogether the usefulness of MRI should be based on different considerations; a significant improvement in the diagnostic investigation, a more adequate surgical strategy, a reduced rate of recurrence and an improved quality of life for the patients. Another factor is the additional cost of MRI compared to MMG, which is important, but not within the scope of this review.

Conclusions

This systematic review and meta-analysis analysed the existing evidence regarding the diagnostic accuracy of MMG and MRI to find underlying malignancy in patients with PD. Identification of underlying malignancy was found to be significantly more accurate when MRI was used in addition to MMG. Adding MRI seems therefore to improve surgical treatment of patients with PD and a negative MMG.

However, this is based on retrospective data and only few numbers. Further studies using a prospective study design and evaluating a larger sample size are warranted.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the

PRISMA reporting checklist. Available at <https://abs.amegroups.com/article/view/10.21037/abs-21-95/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://abs.amegroups.com/article/view/10.21037/abs-21-95/coif>). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Dalberg K, Hellborg H, Wärnberg F. Paget's disease of the nipple in a population based cohort. *Breast Cancer Res Treat* 2008;111:313-9.
2. Lim HS, Jeong SJ, Lee JS, et al. Paget disease of the breast: mammographic, US, and MR imaging findings with pathologic correlation. *Radiographics* 2011;31:1973-87.
3. Dominici LS, Lester S, Liao GS, et al. Current surgical approach to Paget's disease. *Am J Surg* 2012;204:18-22.
4. Sakorafas GH, Blanchard DK, Sarr MG, et al. Paget's disease of the breast: a clinical perspective. *Langenbecks Arch Surg* 2001;386:444-50.
5. Caliskan M, Gatti G, Sosnovskikh I, et al. Paget's disease of the breast: the experience of the European Institute of Oncology and review of the literature. *Breast Cancer Res Treat* 2008;112:513-21.
6. Dubar S, Boukrid M, Bouquet de Joliniere J, et al. Paget's Breast Disease: A Case Report and Review of the Literature. *Front Surg* 2017;4:51.
7. Ling H, Xu XL, Liu ZB, et al. Patients with nipple-areola Paget's disease and underlying invasive breast carcinoma had a very poor survival: A matched cohort study. *Cancer Research Conference: 35th Annual CTRC AACR San*

- Antonio Breast Cancer Symposium San Antonio, TX United States Conference Publication: 2012:72.
8. Helme S, Harvey K, Agrawal A. Breast-conserving surgery in patients with Paget's disease. *Br J Surg* 2015;102:1167-74.
 9. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
 10. Munn Z, Stern C, Aromataris E, et al. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol* 2018;18:5.
 11. Covidence systematic review software Veritas Health Innovation, Melbourne, Australia; 2020.
 12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
 13. Siponen E, Hukkinen K, Heikkilä P, et al. Surgical treatment in Paget's disease of the breast. *Am J Surg* 2010;200:241-6.
 14. Friedman EP, Hall-Craggs MA, Mumtaz H, et al. Breast MR and the appearance of the normal and abnormal nipple. *Clin Radiol* 1997;52:854-61.
 15. Kim S, Lee W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? *Stat Methods Med Res* 2017;26:142-54.
 16. Review Manager (RevMan). 5.4 ed: The Cochrane Collaboration; 2020.
 17. Chang SL, Gray K. Imaging evaluation of Paget's disease of nipple: Are we doing enough? *Breast Cancer Res Treat* 2020;180:566.
 18. Zhu T, Shi J, Tang Z, et al. Comparison on imaging and pathology findings of mammary Paget disease. *Chinese Journal of Medical Imaging Technology* 2018;34:1216-9.
 19. Kim HS, Seok JH, Cha ES, et al. Significance of nipple enhancement of Paget's disease in contrast enhanced breast MRI. *Arch Gynecol Obstet* 2010;282:157-62.
 20. Frei KA, Bonel HM, Pelte MF, et al. Paget disease of the breast: findings at magnetic resonance imaging and histopathologic correlation. *Invest Radiol* 2005;40:363-7.
 21. Morrogh M, Morris EA, Liberman L, et al. MRI identifies otherwise occult disease in select patients with Paget disease of the nipple. *J Am Coll Surg* 2008;206:316-21.
 22. Nakai K, Horimoto Y, Semba R, et al. What is the most appropriate surgical procedure for Paget's disease? *Breast* 2019;44:S104.

doi: 10.21037/abs-21-95

Cite this article as: Madsen KL, Mosebo ADH, Möller S, Pedersen BH, Bille C. Accuracy of mammography and magnetic resonance imaging to diagnose underlying malignancy in Paget's disease of the nipple: a systematic review and meta-analysis. *Ann Breast Surg* 2023;7:13.

Supplementary appendix

Table of contents

Appendix 1: Search strategy	2
Appendix 2: QUADAS-2 questions	4
Appendix 3: QUADAS-2 table.....	6

Appendix 1: Search strategy

Embase (Ovid): 1947 - 2020 October 14: 560 hits

1. Paget nipple disease/
2. Paget's disease of the nipple.mp.
3. Paget nipple disease.mp.
4. Paget's disease, mammary.mp.
5. Paget's disease of the breast.mp.
6. mammary Paget's disease.mp.
7. Paget disease of the nipple.mp.
8. Paget disease of the breast.mp.
9. mammary Paget disease.mp.
10. Paget breast disease.mp.
11. breast, Paget disease.mp.
12. nipple, Paget disease.mp.
13. Paget's nipple disease.mp.
14. breast, Paget's disease.mp.
15. nipple, Paget's disease.mp.
16. Paget's disease, nipple.mp.
17. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18. nuclear magnetic resonance imaging/
19. nuclear magnetic resonance imaging.mp.
20. MRI.mp.
21. MR.mp.
22. breast MRI.mp.
23. breast MR.mp.
24. mamma MRI.mp.
25. mamma MR.mp.
26. 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27. 17 and 26

Medline (Ovid): 1946 - October 13, 2020: 48 hits

1. Paget's Disease, Mammary/
2. Paget's disease of the nipple.mp.
3. Paget's disease, mammary.mp.
4. Paget's disease of the breast.mp.
5. mammary Paget's disease.mp.
6. Paget disease of the nipple.mp.
7. Paget disease of the breast.mp.

8. mammary Paget disease.mp.
9. Paget breast disease.mp.
10. breast, Paget disease.mp.
11. nipple, Paget disease.mp.
12. breast, Paget's disease.mp.
13. nipple, Paget's disease.mp.
14. Paget's disease, nipple.mp.
15. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16. Magnetic Resonance Imaging/
17. magnetic resonance imaging.mp.
18. MRI.mp.
19. breast MRI.mp.
20. breast MR.mp.
21. MR.mp.
22. mamma MR.mp.
23. Mamma MRI.mp.
24. 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23
25. 15 and 24

Cochrane Library - October 15, 2020: 11 hits

ID	Search
#1	Paget's disease of the nipple
#2	Paget's disease, mammary
#3	Paget's disease of the breast
#4	Paget Nipple Disease
#5	Paget breast disease
#6	Paget's carcinoma
#7	#1 or #2 or #3 or #4 or #5 or #6
#8	MeSH descriptor: [Magnetic Resonance Imaging] explode all trees
#9	MRI
#10	MR
#11	Magnetic resonance imaging
#12	mamma MRI
#13	breast MRI
#14	breast MR
#15	mamma MR
#16	#8 or #9 or #10 or #11 or #12 or #13 or #14 or #15
#17	#7 and #16

Appendix 2: QUADAS-2 questions

The questions marked as **bold** are added to the tool by the authors.

Patient selection

Signalling question 1: Was a consecutive or random sample of patients enrolled?

Signalling question 2: Was a case-control design avoided?

Signalling question 3: Did the study avoid inappropriate exclusions?

Signalling question 4: Did all patients receive both MRI and MMG?

Applicability: Are there concerns that the included patients and setting do not match the review question?

Index test

Signalling question 1: Were the index test results interpreted without knowledge of the results of the reference standard?

Signalling question 2: If a threshold was used, was it pre-specified?

Signalling question 3: Did at least two radiologists independently evaluate the preoperative images?

Applicability: Are there concerns that the index test, its conduct, or interpretation differ from the review question?

Reference standard

Signalling question 1: Is the reference standard likely to correctly classify the target condition?

Signalling question 2: Were the reference standard results interpreted without knowledge of the results of the index test?

Applicability: Are there concerns that the target condition as defined by the reference standard does not match the question?

Flow and timing

Signalling question 1: Was there an appropriate interval between index test and reference standard?

Signalling question 2: Did all patients receive the same reference standard?

Signalling question 3: Were all patients included in the analysis?

Appendix 3: QUADAS-2 table

<u>Studies</u>	<u>Risk of bias</u>				<u>Applicability concerns</u>		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Nakai et al	O	O	+	+	+	-	+
Kim et al	O	-	+	O	-	+	+
Siponen et al	O	-	+	O	+	-	-
Morrogh et al	O	O	+	+	O	+	+
Frei et al	O	O	+	+	-	+	+
Friedman et al	O	O	+	O	-	O	+