

High Risk (B3) Breast Lesions: What is the Incidence of Malignancy for Individual Lesion Subtypes? A Systematic Review and Meta-Analysis

Nerys Dawn Forester¹, Simon Lowes², Elizabeth Mitchell³, Maureen Twiddy⁴

¹Consultant Radiologist, Breast Screening and Assessment Unit, Royal Victoria Infirmary, Queen Victoria Road, Newcastle NE1 4LP

²Consultant Radiologist, Breast Screening and Assessment Unit, Queen Elizabeth Hospital, Gateshead, NE9 6SX

³Hull York Medical School, Institute of Clinical and Applied Health Research, The Allam Medical Building, University of Hull, Hull, HU6 7RX

⁴Hull York Medical School, Institute of Clinical and Applied Health Research, The Allam Medical Building, University of Hull, Hull, HU6 7RX

Correspondence to: Dr Nerys Forester

8 Shaftesbury Avenue, Roundhay, Leeds LS8 1DT UK

nerysforester@mac.com

Declaration of Interest: None

Word count 4457

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Abbreviations

ADH	atypical ductal hyperplasia
AIDP	atypical intraductal proliferation
ALH	atypical lobular hyperplasia
B3	Breast lesions of uncertain malignant potential
BSP	Breast Screening Program
CI	Confidence Interval
CNB	Core Needle Biopsy
DCIS	Ductal Carcinoma in Situ
FEA	flat epithelial atypia
G	Gauge
LCIS	lobular carcinoma in situ
LISN	lobular in situ neoplasia
LN	Lobular Neoplasia
NHS	National Health Service
NICE	National Institute for Clinical Excellence
PICO	Population, Interventions, Controls, Outcome
RS/CSL	radial scar/complex sclerosing lesion
UK	United Kingdom
US/USA	United States of America
VAE	Vacuum Assisted Excision
VANCB	Vacuum Assisted Needle Core Biopsy

Abstract

Introduction Provide evidence to support evolving management strategies for high-risk (B3) breast lesions by assessing risk of carcinoma in subgroups of B3 lesions using systematic review and meta-analysis.

Methods Databases identified observational studies between 1980 and 2015 that reported on underestimation of malignancy following B3 lesion diagnosis at core needle biopsy. Critical appraisal, quality assessment, data extraction and meta-analysis was undertaken to calculate rate of malignancy of the whole B3 group and individual lesions. Study heterogeneity and association between variables and underestimation rate was investigated using random effects logistic modelling.

Results Meta-analysis, using data from 129 studies, assessed 11 423 lesions of which 2160 were upgraded to malignancy after surgical excision biopsy (17% malignancy rate, 95% CI 15 to 19%). Malignancy rates varied from 6% in radial scars with no atypia (95% CI 2 to 13%, I² 72.8%), to 32% in papillomas with atypia (95% CI 23 to 41%, I² 57.4%). Differences in upgrade rates between atypical and non-atypical lesions were statistically significant ($p < 0.05$). Study heterogeneity could not be explained by differences in core biopsy size or year of publication.

Conclusions This comprehensive, inclusive assessment of all published literature, provides an accurate estimate of malignancy risk in subgroups of B3 lesions, to guide tailored management strategies. Some lesions have a high risk of malignancy, while others have a much lower risk, and could be safely managed with surveillance strategies rather than surgery.

Keywords: B3 breast lesions, lesions of uncertain malignant potential, breast cancer

1 Introduction

Breast cancer remains the most commonly diagnosed cancer in females in the United Kingdom, accounting for 15% of all new cancer cases in 2015 [1], with a 1 in 9 lifetime risk [2]. Diagnosis is usually made following targeted image-guided needle biopsy of the breast, with biopsy findings categorised to one of five categories: B1 (normal breast tissue/inadequate sample), B2 (benign tissue), B3 (benign but of uncertain malignant potential), B4 (suspicious for malignancy), and B5 (malignant) [3].

Lesions falling into the B3 category are the subject of a great deal of interest, precisely because of their uncertain malignant potential and the lack of a definitive consensus on how they should be managed once diagnosed. They represent a group of diverse histopathological entities, with a variable degree of associated malignancy, and include fibroepithelial lesions, papilloma, radial scars/complex sclerosing lesions, atypical ductal hyperplasia (ADH; also termed atypical intraductal proliferation, AIDP), flat epithelial atypia (FEA), atypical lobular hyperplasia (ALH), and lobular carcinoma in situ (LCIS). In isolation these pathological changes are not necessarily malignant, however, malignant change can coexist with the lesion, and the presence of the lesion itself confers an increased risk of subsequent breast malignancy over time [4-6].

B3 breast lesions are usually identified on imaging as areas of calcification or small masses, and are mostly asymptomatic, identified on screening mammography in well women, though they may also be found in women presenting to symptomatic breast clinics. The age extension of the UK screening programme to include women between 47 and 50 years, and

recent improvements in digital technology have increased the number of B3 lesions identified in women of screening age; currently B3 lesions are diagnosed in 5-10% of core needle biopsies performed as part of the National Health Service Breast Screening Programme (NHSBSP) in the UK [7,8].

The management of B3 lesions has undergone significant change over the last few years. As these lesions are usually diagnosed with 11 or 14 Gauge (G) core biopsy (a relatively small tissue sample), it is recognised that under-sampling of the lesion can mask malignancy [3]. Studies show that between 16 and 20% of all B3 lesions will be upgraded to malignant, if larger gauge biopsy (e.g. 7 or 8 G) or a diagnostic surgical excision biopsy is performed (Bianchi et al., 2011) [9]. In practice, this means that if a woman is diagnosed with any B3 lesion, there is approximately a 1 in 5 chance that this is actually a malignant lesion, but the biopsy has not provided enough tissue to diagnose this. Historically, all B3 lesions were managed with surgical excision, regardless of the underlying lesion type, to be sure to identify any co-existing malignancy, which may have been missed on core biopsy. However, for 4 out of 5 women with a B3 lesion, this would mean unnecessary surgery.

NICE guidelines concluded that there is sufficient evidence regarding the safety and efficacy of image-guided large volume biopsy to use for the diagnostic sampling of B3 breast lesions rather than surgical excision biopsy [10]. Following large volume biopsy, any lesions upgraded to a malignancy are treated accordingly and the remaining B3 lesions are left within the breast. This has begun to reduce unnecessary surgery in this group of women. However, this risk of missing the area of malignancy with a B3 lesion varies depending on

the underlying B3 lesion subgroup, with some lesions having a lower risk of missed malignancy than others.

In view of this, it would be beneficial to develop more sophisticated management strategies for individual lesion types. However, the published literature reports a very variable degree of malignancy underestimation, both for B3 lesions as a group, and for individual lesion types. In part, this is due to statistical error from studies with small sample sizes, and a degree of variability between pathologists and institutions [5, 9]. Formulating a robust management strategy when the published risk of malignancy of a B3 lesion can range from 0% [11,12] to 79% [13] is challenging. The presence of conflicting literature makes it difficult to know how conservative or aggressive management pathways should be for this group of lesions.

There is currently no accepted international consensus regarding the definitive management of such lesions following their identification, though while conducting our review, suggested guidance for the management of screen-detected B3 lesions has been appended to the most recent NHSBSP breast cancer screening guidance document [14]. However, despite this, there has still, to date, been no registered attempt to perform a formal systematic review of the literature and combine data from all the available published studies. Whilst individual studies may reflect the rate of upgrade to malignancy for individual groups at a given time, a systematic review of all the relevant published studies allows a more accurate estimate of the risk of malignancy in the different subtypes of this complex group of lesions. In turn this will provide a more robust and generalisable evidence base from which to design tailored management strategies for individual lesions.

2 Methods

The results of this systematic review and meta-analysis are reported according to Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines [15].

2.1 Data sources and searches

A systematic search of Medline, Embase, HMIC, Scopus and ISI Web of Knowledge, between 1 January 1980 and 15 March 2015, was undertaken using a search strategy constructed with the support of an Information specialist using the search terms shown in Appendix I. Additional references were identified by interrogating reference lists of all identified articles and suggested article links whilst reading papers on-line. Searches were limited to English Language studies, as the majority of literature was expected to be from the UK, USA and Europe.

2.2 Study selection criteria

All published studies which identified patients diagnosed with B3 breast lesions at core needle biopsy, who subsequently had a timely (within 6 months, without interval radiological investigations) surgical excision to assess for any co-existing malignancy, were identified. Studies must have reported the subgroups of B3 lesions analysed. This included patients with papilloma, radial scar/complex sclerosing lesions (RS/CSL), atypical ductal hyperplasia (ADH), atypical intraductal proliferation (AIDP), flat epithelial atypia (FEA), atypical lobular hyperplasia (ALH), lobular in situ neoplasia (LISN), and lobular carcinoma in situ (LCIS). Fibroepithelial lesions have not been included in this review, as following their

diagnosis on core biopsy, diagnostic excision biopsy is indicated, both for the potential risk of undersampling of invasive or in situ malignancy as for the rest of the B3 lesion group, but also because of the potential for this lesion to be a benign or malignant phyllodes tumour. These need to be removed completely for diagnosis and alternative management strategies with large volume biopsy and/or surveillance mammography would not be appropriate. Mucocoele-like lesions were also not considered as these lesions are extremely rare.

The primary outcome measure was rate of malignancy of B3 lesions following immediate surgical excision biopsy.

2.3 Exclusion criteria

Studies were excluded if the management following diagnosis of B3 a lesion was unclear, if patients did not have excision biopsy within 6 months, or had an excision biopsy after a period of follow up if the radiological appearances had altered. Studies were not included if the initial lesion diagnosis was by surgery or cytology rather than core needle biopsy. Studies were also excluded if the initial biopsy method was solely by Magnetic Resonance Imaging (MRI)-guided breast biopsy. This was to reduce heterogeneity between included studies as much as possible; MRI biopsy is generally performed with larger gauge needles (7 and 8G), whereas the vast majority of B3 lesions are identified using 14-10G biopsy needles. However, the decision was made to include a few studies that contained one or two MRI biopsies alongside ultrasound or stereotactic biopsies, as excluding these papers would have limited the pool of B3 lesions available for analysis. Studies were also excluded in patients with co-existing breast cancer prior to B3 lesion diagnosis.

As the indications for surgical excision biopsy have evolved recently, four more recent papers were found where the second line sampling was with a large volume biopsy, rather

than surgical excision biopsy [16-19]. Whilst it is likely that large volume biopsy and surgical excision are comparable, these papers were excluded to minimise heterogeneity.

2.4 Patient involvement

As this is a systematic review of published literature, no patients were involved in setting the research question, or in the design or implementation of the study. We did not evaluate whether the studies included in the review had any patient involvement. However, it is hoped that dissemination of the results following publication will allow national or international guidelines for the management of B3 lesions to be evidence based and developed further for the benefit of this patient group.

2.5 Data extraction and assessment of risk of bias

One author (NF) screened the titles of all studies identified by the searches. Two reviewers (NF and SL) independently reviewed all full text articles. Disagreements between reviewers were resolved by consensus. The study selection process is summarised in Figure 1.

Variables were extracted from each paper using a data extraction sheet in excel that both reviewers completed independently. Data extracted were: author names, publication year and country of origin of study. Variables that could be associated with heterogeneity between studies were also recorded such as mean/median age of subjects and core needle biopsy size.

Study size and the number of lesions included for each of the subgroups of B3 lesions was recorded along with the number of malignancies identified after surgical excision biopsy for

each lesion subtype. This enabled calculation of the proportion of patients with an initial B3 diagnosis at core needle biopsy, who subsequently had malignancy discovered at surgical excision biopsy. The proportion of malignant lesions was calculated by taking the number of malignant lesions found at surgical excision biopsy divided by the total number of indeterminate lesions at initial core needle biopsy included in the study.

2.6 Assessment of bias

The Newcastle Ottawa Scale [20] for cohort studies was adapted to determine inclusion in the review. Although not formally 'graded' to assess quality, a series of checks were developed from the Newcastle Ottawa scale with respect to case selection, case comparability and reporting of study outcomes (Appendix II checklist).

Although studies were only included where it was stated that cases were identified retrospectively from a pathology database, or prospectively as they were identified, it was impossible to fully exclude the presence of a case selection bias.

To assess a degree of comparability between patient cohorts, data were obtained from each study about mean patient age and biopsy gauge used.

Reporting of outcomes was assessed such that only studies where the majority of subjects underwent surgical excision were included. If fewer than 20% of patients were lost to surgical follow up the study was included, if patients without surgical excision could be excluded from the final analysis.

Assessment of heterogeneity was initially performed with visual inspection of the Forest Plot and tested statistically using a Higgins' I^2 test [21]. An I^2 value of more than 50% was considered to indicate moderate statistical heterogeneity, with values over 75% indicating high heterogeneity [22].

Data were analysed using STATA version 10.0 (Stata Corp, College Station, TX, USA), using the meta-analysis of proportion commands with a random effects model and a Freeman-Tukey transformation [23].

The study specific underestimation of malignancy rates were calculated with 95% confidence intervals and were shown in Forest plots, ordered alphabetically by first author of each study. The homogeneity of effect size was tested using the I^2 statistic. To assess heterogeneity, a stratified meta-analysis was performed by examining data from subgroups of B3 lesions to assess potential contributing factors such as needle biopsy size, year of publication or the presence or absence of atypia.

3 Results

Initial searches returned 2289 citations. After removing duplicates, 1962 references remained. An additional 11 references were identified during web searching for full text articles, and by checking references. Overall, 1973 references were screened using title and abstract for further reading. Following this, 183 studies remained for full text review.

After full text review, 54 papers were excluded as they did not fulfil the inclusion criteria after detailed evaluation of the published article, leaving 129 references for data extraction and meta-analysis. The article selection process is summarised in Figure 1. Excluded papers are shown in Appendix III. The 129 included studies are shown in Appendix IV. Overall, there were 11,423 B3 lesions included in the study, of which 2160 were subsequently upgraded to malignant lesions after surgical excision biopsy. Included studies were from 15 countries (Appendix V). The studies were then divided into sub groups by lesion type (papilloma, ADH and AIDP, radial scar, lobular neoplasia and FEA) and the breakdown of studies and number of lesions analysed per group is shown in Table 1. As expected, the majority of the lesions identified were in women of screen-detected age (that is, over 50), although a minority of studies included women of pre-screening age.

The pooled estimate of the underestimation of malignancy for the whole B3 group was 17% (95% CI: 15 to 19%). When considering individual lesion subtypes this malignancy risk ranges from 6% for a radial scar without atypia, to 32% in a papilloma with atypia (Figure 2, Table 1). The Higgins I² for the group as a whole was 86.2%, showing that a high amount of heterogeneity between studies was present (Table 1) [22].

Table 1 Table of overall results for individual B3 lesion subtypes

	Number of included studies	Number malignant Lesions	Total Number Lesions	Rate of Malignancy (%)	95% confidence interval (%)	Higgins I ² (%)
Papilloma	42	351	2278	12	10-15	77.2

<i>No atypia</i>	14	90	1162	7	4-10	62.0
<i>Atypia</i>	11	91	298	32	23-41	57.4
ADH/AIDP	47	1114	4031	28	24-31	81.6
Radial Scar	15	88	934	8	6-11	44.3
<i>No atypia</i>	4	22	334	6	2-13	72.8
<i>Atypia</i>	2	8	43	18	8-32	90.9
Lobular Neoplasia	38	345	2014	17	13-21	80.9
<i>ALH</i>	16	54	463	12	5-21	72.9
<i>LCIS</i>	16	76	359	22	14-31	63.6
FEA	19	179	1413	11	8-14	42.6
All B3 lesions	129	2160	11423	17	15-19	86.2

3.1 Presence or absence of atypia

When considering both papillomas and radial scars, the presence of atypia had a significant effect on the malignancy underestimation rate. For papillomas, upgrade rate to malignancy was 7% without atypia compared to 32% with atypia ($p < 0.01$, Figure 3). In radial scars, upgrade rate to malignancy was 6% without atypia compared to 18% with atypia ($p < 0.034$, Figure 3).

Studies considering a diagnosis of lobular neoplasia on core needle biopsy demonstrated a 17% upgrade to malignancy. Sub-group analysis was performed as some studies subdivided lesions into atypical lobular hyperplasia and lobular carcinoma in situ. Upgrade rates were

12 and 22% respectively, however, there was no statistically significant difference between these groups ($p=0.11$).

Atypical ductal hyperplasia and atypical intraductal proliferation were analysed together and shown in Figure 4. The rate of underestimation of malignancy was 28%. The Forest plot for flat epithelial atypia shows the rate of malignancy underestimation was 11% (Figure 4).

3.2 Factors contributing to heterogeneity

No difference could be found within the lobular neoplasia group when analysing sub-groups, and the number of lesions included in this group was large ($n=2014$). Therefore, the lobular neoplasia group was used to explore other factors that could contribute to the heterogeneity. It would not have been possible to do this analysis using the dataset as a whole, as the differences present due to atypia in radial scars and papillomas may have masked any differences present due to other factors, such as year of publication, or core needle biopsy size.

The lobular neoplasia dataset was explored using year of publication, grouped by decade (Figure 5). No difference between groups was observed. Similarly, analysis by core needle size did not show any differences between the groups (Figure 5). However, as most of the studies included used a variable size of needle core biopsy within each study, it was difficult to subdivide the studies robustly, and a degree of overlap of core needle sizes between groups was present. When the lobular neoplasia group was analysed with respect to country of publication (US versus non-US literature, Figure 5), the risk of upgrade to

malignancy was significantly higher in the non-US literature (31%) than in the US studies (11%, $p < 0.01$).

4 Discussion

This systematic review and meta-analysis shows that the pooled estimate for the risk of upgrade to malignancy after a B3 diagnosis on core needle biopsy is 17%. This is similar to that reported by some studies [9], but is much lower than that described by Lieske et al [6], who described a 34% malignancy rate in B3 lesions identified in the NHS BSP between 1999 and 2005. It is known that the presence or absence of atypia, pathological features that increase suspicion for the presence of malignancy, greatly affects the risk of carcinoma in B3 lesions. To consider these differences, some groups have begun to divide B3 lesions into “pre-malignant lesions” and “controversial lesions” depending on the presence of atypia, and use this to guide management [24]. Premalignant lesions include lesions containing atypia, such as atypical epithelial proliferation (AIDP) and lobular neoplasia (LN) and controversial lesions include lesions which may be associated with malignancy, such as papillary lesions and radial scars/complex sclerosing lesions without atypia present, mucocoele-like lesions, flat epithelial atypia, and phyllodes tumours [7]. This review shows that there would be justification for regarding lesions in this way, as all lesions containing atypia were shown to have a higher risk of malignancy than those without.

This review demonstrates that atypical intraductal proliferation (AIDP)/atypical ductal hyperplasia (ADH) has the greatest risk of associated malignancy within the B3 lesions (28%). Historically studies have reported the risk of co-existing malignancy in AIDP/ADH to vary as widely as between 18 and 87% [3, 25]. More recently, the First International Consensus conference on lesions of uncertain malignant potential in the breast (B3 lesions) [26] acknowledged that stereotactic vacuum assisted biopsy underestimation rates for ADH

vary between 9 and 58%. In contrast, ADH lesions undergoing a diagnostic ultrasound guided 8G vacuum assisted biopsy have been reported to have no underestimation rate [27]. Whilst this review included literature reporting ADH/AIDP upgrade rates between 7 and 62%, if the whole scope of available literature is searched, part of this variation is due to the presence of studies with small sample sizes in the literature. Combining these studies in a meta-analysis identifies an overall risk of upgrade to malignancy of 28%, and makes ADH/AIDP a B3 lesion at a high risk of having co-existing malignancy present at the time of initial core biopsy. However, this risk can be managed by offering a second line large gauge (7 or 8G) vacuum assisted biopsy if possible (taking into account patient factors, radiological size and pathological concern). This strategy reduces the risk of identifying as yet undiagnosed malignancy during diagnostic surgery. 7 or 8G biopsy should upgrade the lesion to invasive or in situ malignancy with a low risk of further underestimation as described by Lourenco [27]. However, following a second line 7 or 8 G VAB, annual surveillance is still recommended for these women. This review supports the notion that lesions containing AIDP/ADH should be excised, either by image-guided vacuum assisted excision (VAE), or surgical excision if VAE is not available or appropriate, and is consistent with the current NHSBSP guidance [14].

Lobular neoplasia is often an incidental finding, as it is not usually associated with any changes on the mammogram [28], but the risk of DCIS on surgical excision is reported as ranging from 0 to 67% [29, 30]. This systematic review found that 17% had co-existing malignancy (when considering both invasive and in situ disease) at diagnostic excision biopsy. Such a wide variation in risks seen in the literature is likely to reflect small studies and variations in pathological reporting of lobular neoplasia. Whilst 17% is not quite as high

a risk of malignancy as for lesions such as AIDP/ADH, the lobular neoplasia group of lesions should still be considered to be of a moderate risk of concurrent malignancy (using a risk of malignancy of approximately 4 times higher than normal as a moderate degree of risk in the screening setting [31]), and management pathways should consider VAE and enhanced surveillance in this group. Lobular neoplasia can be an extensive lesion, and as such surveillance strategies are appropriate in this group of patients if managed by second line VAB rather than diagnostic excision biopsy which could otherwise require large areas of the breast to be excised.

Radiologically, FEA is usually associated with calcifications, with changes seen within secretions in the luminal spaces on pathology. Although this is a lesion that, by definition, contains a degree of epithelial atypia, upgrade rate to malignancy was found to be only 11%. This may be due to the relatively small numbers of studies identified that reported FEA in isolation, and may reflect different nomenclature used to describe such lesions. However, as this group showed the lowest amount of heterogeneity this would suggest that this is an accurate estimate of the risk of this lesion. Current management strategies include diagnostic surgical excision or enhanced mammographic surveillance [32]. As the risk of co-existing cancer is towards the lower end of the spectrum of risk for these lesions, patients with FEA could be safely managed with VAE alone.

As in the literature reviewed, we have considered radial scars and papillomas to be 'with atypia' if there is coexisting epithelial atypia present (ADH, AIDP, FEA, LN), but this atypia falls short of a diagnosis of DCIS. In both papillary lesions and radial scars/complex sclerosing lesions, the presence of atypia has been shown to be the strongest predictor of

an upgrade to malignancy [33]. Because of this, papillary lesions and radial scars without atypia should be relatively safe to leave in the breast, while those lesions associated with atypia should be removed surgically. This management strategy is supported by the findings in this review, and by the draft guidelines produced by the United Kingdom B3 lesion working group (currently unpublished). In particular, for papillomas with atypia, where the risk of malignancy underestimation in this review is 32%, surgical excision should be considered, if malignancy is not identified following second line VAE. In this review the data set for radial scars with atypia was very small (only 2 studies with 43 lesions), making it difficult to draw reliable conclusions on which to guide management.

4.1 Strengths and weaknesses

This is the first systematic review to consider the key subgroups of B3 lesions and to calculate the risk of malignancy for individual groups. The inclusive nature of the review allowed assessment of a large number of B3 lesions (11,423 in total). This provides a more accurate estimate of the risk of malignancy underestimation, and made it possible to explore explanatory factors associated with the rate of malignancy underestimation in B3 lesions, in particular, the presence or absence of atypia and the effects of core needle biopsy size.

There is no evidence that mode of presentation (screening versus symptomatic), affects the final outcome following B3 lesion diagnosis, so studies were not limited to include only screen detected lesions [34]. The data were analysed to calculate the most conservative estimate of malignancy risk for individual B3 lesions. This was achieved by using a random effects model in the meta-analysis and trying to keep as broad a range of published

literature in the study as possible. As such, the data presented can be considered the worst-case scenario on which to base safe management strategies.

This review only considered published articles available in English, however, it is possible that this excluded literature that would have fulfilled the inclusion criteria of the review and expanded the dataset. Similarly, some studies had to be excluded as they reported results as an overall group of all lesions, rather than individual subgroups. Although study selection attempted to remove any lesion selection bias by excluding studies that explicitly reported a period of follow-up prior to surgical excision biopsy, it cannot be determined whether some studies proceeded straight to excision biopsy for more suspicious lesions, thus selecting out those higher risk lesions from the core biopsy group. However, the difference between upgrade rates to malignancy between the US and non-US literature would support the conclusion that a degree of selection bias is present in some of the published studies.

Intuitively, and from published studies [35], it was expected that size of core needle biopsy size should have influenced malignancy upgrade rates. The effect of core biopsy size was explored using the lobular neoplasia dataset. No difference between biopsy size and risk of malignancy underestimation was shown. However, the biopsy sizes in the groups compared were a range of sizes, as very few studies used a single core needle size to diagnose all B3 lesions reported within that study. However, it is possible that biopsy size differences are not demonstrable between 10, 11 and 14G core sizes until significantly increases in needle size, such as 7 or 8G cores are used.

This meta-analysis shows that the pooled estimate for the risk of upgrade to malignancy after a B3 diagnosis on core needle biopsy is 17%. However, when considering individual lesion subtypes this risk ranges from 6% for a radial scar without atypia, to 32% in a papilloma with atypia. These results reinforce that lesions with atypia have a much higher rate of underestimation than those without atypia (statistically significant difference), which is concordant with the review of papillary lesions by Wen and Cheng [36].

Arguably the most important factor contributing to inter-study heterogeneity is variations in clinical practice between pathologists. Clinical variation exists as there is more scope for differences in opinion between pathologists when reporting on B3 breast lesions. A recent study of histopathologists' levels of agreement between breast biopsies showed that 53.6% of samples labelled as having 'atypia' were over-interpreted [37]. If a pathologist has a low threshold for calling a lesion B3, then the rate of upgrade to malignancy will be lower than someone who only reports a B3 lesion with higher degrees of atypia present. Unfortunately, without cross pathologist and inter-institution review, this is a factor that it is not possible to control for in this review, but it would explain why the rate of upgrade to malignancy varies so much between studies. However, this systematic review accepts that such variability exists and aims to calculate a conservative, but accurate, estimate of risk, taking the natural variability of reporting between pathologists into account.

Geographically, differences in the management of B3 lesions exist across the UK.

Internationally, management with surgical excision biopsy is usual practice, although emerging evidence suggests that this may be over-treatment for subgroups of B3 lesions. To date, NICE has not published any guidance regarding their management, leaving individual

breast units to produce their own care pathways, often based on their own local audit of small groups of lesions. The inherent small sample sizes within such studies makes the data available for individual lesions unreliable. This systematic review collates the available international data and provides a more robust estimate of the overall risk for individual B3 lesions.

4.2 Recommendations

The aim of this work was to provide high level evidence on which to base a framework for safe management of these heterogeneous lesions, without having to resort to a one size fits all approach. This review would recommend that higher risk lesions, such as papillary lesions with atypia, should undergo surgical excision biopsy. Lesions such as ADH/AIDP and lobular neoplasia should be removed by vacuum-assisted excision where possible and then followed up with surveillance mammography. Lower risk lesions such as radial scars and papillomas without atypia have a risk of malignancy of under 10%, which could be accepted as a low enough risk to leave the lesion within the breast.

Clearly, a careful, multidisciplinary approach to these lesions is required, and where there are radiological or pathological concerns about lesions where imaging or pathological features 'don't quite fit', these patients should not be managed within the confines of a treatment pathway. However, for the vast majority of lesions, this review provides a generalizable estimate of the risk of malignancy underestimation for specific groups of lesions, and we should be able to confidently apply this to management guidelines, and begin to have a more conservative management approach to certain groups of B3 lesions.

Although the majority of women with B3 lesions are eventually proven to have benign disease, it does generate a group of women who have potentially 'at risk' lesions in their breasts, who may benefit from more frequent screening or chemoprevention. However, this is outside the scope of this review, and will be addressed by the Sloane review, a national audit database of B3 lesions, which began in 2014, that will collect data concerning future malignancy risk in patients with subgroups of B3 lesions (ADH, LISN, FEA and pleomorphic LCIS). It is unlikely that this will report for some time.

4.3 Conclusion

In conclusion, this study presents pooled estimates of the underestimation of malignancy in subgroups of B3 lesions initially diagnosed on core needle biopsy. Systematic reviews are important tools, which can be used to guide clinical management in areas where individual studies report variability in outcomes. Due to the spectrum of risks of the lesions within the B3 group, these estimates should be used to guide further management strategies of these lesions on an individual lesion basis, rather than using management strategies for the group as a whole.

Acknowledgments: Assistance with the searches was provided by Erika Gavillet, Librarian, Newcastle University and Mark Chambers, Librarian, Newcastle Teaching Hospitals NHS Trust. Statistical advice was provided by Professor Robert West at the University of Leeds.

Competing Interests: None

Funding: This work was generously supported by the Symposium Mammographicum Research Award, which covered the university fees to submit the work as an MSc thesis

Legends for Tables and Figures

Figures

Figure 1 – Study selection flow chart

Figure 2 -Incremental malignancy rates of individual B3 lesion subtypes

Figure 3 - Forest plots for papilloma and radial scars with and without atypia

Figure 4 - Forest plots for Atypical Intraductal Hyperplasia/Atypical Ductal Hyperplasia and Flat Epithelial Atypia

Figure 5 - Forest plots exploring heterogeneity within the Lobular Neoplasia dataset (decade of publication, core needle biopsy size and country of publication)

Tables

Table 1 – Table of Overall Results for individual lesions

Appendix

I Search Terms

II Checklist

III Excluded Papers

IV Included Papers

V Country of origin of studies

References

1. Cancer Research UK <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer>
2. Office for National Statistics. 2018. Cancer statistics registrations: registrations of cancer diagnosed in 2016, England. Series MB1 no. 37. London: Office for National Statistics; 2018 <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2016>
3. Ellis, I.O., Humphries, S., Michell, M., Pinder, S.E., Wells, C.A., Zakhour, H.D.. 2001. Guidelines for nonoperative diagnostic procedures and reporting in breast cancer screening (Publication no 50), NHSBSP, Editor. June 2001, NHS Cancer Screening Programmes: Sheffield ISBN 1 871997 44 5 <http://webarchive.nationalarchives.gov.uk/20150506150512/http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp50.pdf>
4. Houssami, N., Ciatto, S., Bilous, M., Vezzosi, V. and Bianchi, S. 2007. Borderline breast core needle histology: predictive values for malignancy in lesions of uncertain malignant potential (B3). *British journal of cancer*. 96(8):1253–1257.
5. El-Sayed, M.E., Rakha, E.A., Reed, J., Lee, A.H.S., Evans, A.J. and Ellis, I.O. 2008. Predictive value of needle core biopsy diagnoses of lesions of uncertain malignant potential (B3) in abnormalities detected by mammographic screening. *Histopathology*. 53(6):650–657.
6. Lieske, B., Ravichandran, D., Alvi, A., Lawrence, D.A.S. and Wright, D.J. 2008. Screen-detected breast lesions with an indeterminate (B3) core needle biopsy should be excised. *European*

journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 34(12):1293–1298.

7. Andreu, F.J., Sáez, A., Sentís, M., Rey, M., Fernández, S., Dinarès, C., Tortajada, L., Ganau, S. and Palomar, G. 2007. Breast core biopsy reporting categories--An internal validation in a series of 3054 consecutive lesions. *The Breast*. 16(1):94–101.
8. Deb R, Ellis I, Jenkins J, Murphy A, Pinder SE. 2015 National breast screening pathology audit 2015. Performance for the period 2011–14. Sheffield: NHS Cancer Screening Programmes.
9. Bianchi, S., Caini, S., Renne, G., Cassano, E., Ambrogetti, D., Cattani, M.G., Saguatti, G., Chiaramondia, M., Bellotti, E., Bottiglieri, R., Ancona, A., Piubello, Q., Montemezzi, S., Ficarra, G., Mauri, C., Zito, F.A., Ventrella, V., Baccini, P., Calabrese, M., Palli, D. VANCB Study Group. 2011. Positive predictive value for malignancy on surgical excision of breast lesions of uncertain malignant potential (B3) diagnosed by stereotactic vacuum-assisted needle core biopsy (VANCB): a large multi-institutional study in Italy. *Breast*. 20(3):264–270.
10. National Institute for Health and Care Excellence 2006. IPG156: Image-guided vacuum-assisted excision biopsy of benign breast lesions. London: NICE; 2006
<https://www.nice.org.uk/guidance/ipg156>
11. Brem, R.F., Tran, K., Rapelyea, J., Michener, K.H., Zisman, G., Mohtashemi, K. and Berezowski, K. 2005. Percutaneous biopsy of papillary lesions of the breast: Accuracy of pathologic diagnosis. *Journal of Women's Imaging*. 7(4):157–162.
12. Carder, P.J., Garvican, J., Haigh, I. and Liston, J.C. 2005. Needle core biopsy can reliably distinguish between benign and malignant papillary lesions of the breast. *Histopathology*.

46(3):320–327.

13. Carder, P.J., Khan, T., Burrows, P. and Sharma, N. 2008. Large volume 'mammotome' biopsy may reduce the need for diagnostic surgery in papillary lesions of the breast. *Journal of Clinical Pathology*. 61(8):928–933.
14. Borrelli C, Cohen S, Duncan A, Given-Wilson R, Jenkins J, Kearins O, Pinder S, Sharma N, Sibbering M, Steel J, Turnbull A, Wallis M. NHS Breast Screening Programme Clinical guidance for breast cancer screening assessment (Publication No 49), November 2016. Public Health England: London, PHE publications gateway number: 2016426
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/567600/Clinical_guidance_for_breast_cancer_screening_assessment_Nov_2016.pdf
15. Stroup D.F., Berlin J.A., Morton S.C., Olkin I., Williamson G.D., Rennie D., Moher D., Becker B.J., Sipe T.A., Thacker S.B. 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 283(15):2008-2012.
16. Burbank, F. 1997. Stereotactic breast biopsy of atypical ductal hyperplasia and ductal carcinoma in situ lesions: improved accuracy with directional, vacuum-assisted biopsy. *Radiology*. 202(3):843–847.
17. Maxwell, A.J., Mataka, G. and Pearson, J.M. 2013. Benign papilloma diagnosed on image-guided 14 G core biopsy of the breast: Effect of lesion type on likelihood of malignancy at excision. *Clinical Radiology*. 68(4):383–387.
18. Tennant SL, Evans A, Hamilton LJ, James J, Lee AH, Hodi Z, Ellis IO, Rakha EA, Wilson AR.

2008. Vacuum-assisted excision of breast lesions of uncertain malignant potential (B3) - an alternative to surgery in selected cases. *Breast*. 17(6):546-549.
19. Youk JH, Kim MJ, Son EJ, Kwak JY, Kim EK. 2012. US-guided vacuum-assisted percutaneous excision for management of benign papilloma without atypia diagnosed at US-guided 14-gauge core needle biopsy. *Ann Surg Oncol*. 19(3):922-928.
20. Wells, G.A., Shea, B., O'connell, D., Peterson, J.E.A., Welch, V., Losos, M. and Tugwell, P. 2000. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available online:
http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
21. Higgins, J.P.T. and Thompson, S.G. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine*. 21(11):1539–1558.
22. Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. 2003. Measuring inconsistency in meta-analyses. *BMJ (Clinical Research ed.)*. 327(7414):557–560.
23. Nyaga, V.N., Arbyn, M. and Aerts, M. 2014. Metaprop: a Stata command to perform meta-analysis of binomial data. *Archives of Public Health*. 72(1):39.
24. de Beça, F.F., Rasteiro, C., Correia, A., Costa, S. and Amendoeira, I. 2013. Improved malignancy prediction by B3 breast lesions subclassification. *Annals of Diagnostic Pathology*. 17(5):434–436.
25. Yu, Y.-H., Liang, C. and Yuan, X.-Z. 2010. Diagnostic value of vacuum-assisted breast biopsy for breast carcinoma: a meta-analysis and systematic review. *Breast Cancer Research and Treatment*. 120(2):469–479.

26. Rageth CJ, O'Flynn EA, Comstock C, Kurtz C, Kubik R, Madjar H, Lepori D, Kampmann G, Mundinger A, Baege A, Decker T, Hosch S, Tausch C, Delaloye JF, Morris E, Varga Z. 2016. First International Consensus Conference on lesions of uncertain malignant potential in the breast (B3 lesions). *Breast Cancer Res Treat.* 159(2):203-213.
27. Lourenco AP, Khalil H, Sanford M, Donegan L. 2014. High-risk lesions at MRI-guided breast biopsy: frequency and rate of underestimation. *Am J Roentgenol* 203: 682-686.
28. Oliveira, T.M.G., Elias, J., Melo, A.F., Teixeira, S.R., Filho, S.C., Gonçalves, L.M., Faria, F.M., Tiezzi, D.G., Andrade, J.M. and Muglia, V. 2014. Evolving concepts in breast lobular neoplasia and invasive lobular carcinoma, and their impact on imaging methods. *Insights into Imaging.* 5(2):183–194.
29. Hwang, H., Barke, L.D., Mendelson, E.B. and Susnik, B. 2008. Atypical lobular hyperplasia and classic lobular carcinoma in situ in core biopsy specimens: routine excision is not necessary. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc.* 21(10):1208–1216.
30. Cangiarella, J., Guth, A., Axelrod, D., Darvishian, F., Singh, B., Sinsir, A., Roses, D. and Mercado, C. 2008. Is surgical excision necessary for the management of atypical lobular hyperplasia and lobular carcinoma in situ diagnosed on core needle biopsy?: a report of 38 cases and review of the literature. *Archives of Pathology & Laboratory Medicine.* 132(6):979–983.
31. Houssami N, Abraham LA, Onega T, Collins LC, Sprague BL, Hill DA, Miglioretti DL. 2014. Accuracy of screening mammography in women with a history of lobular carcinoma in situ

or atypical hyperplasia of the breast. *Breast Cancer Res Treat.* 145(3):765-773.

32. Rajan, S., Shaaban, A.M., Dall, B.J.G. and Sharma, N. 2012. New patient pathway using vacuum-assisted biopsy reduces diagnostic surgery for B3 lesions. *Clinical Radiology.* 67(3):244–249.
33. Rakha, E.A., Lee, A.H.S., Jenkins, J.A., Murphy, A.E., Hamilton, L.J. and Ellis, I.O. 2011. Characterization and outcome of breast needle core biopsy diagnoses of lesions of uncertain malignant potential (B3) in abnormalities detected by mammographic screening. *International journal of cancer. Journal international du cancer.* 129(6):1417–1424.
34. MacLean, G.M., Courtney, S.P., Umeh, H., Sanjeev, S., McCormick, C. and Smith, B.M. 2013. Is Mode of Presentation of B3 Breast Core Biopsies (Screen-Detected or Symptomatic) a Distinguishing Factor in the Final Histopathologic Result or Risk of Diagnosis of Malignancy? *World Journal of Surgery.* 37(11):2607–2612.
35. Jackman, R.J., Francis A Marzoni, J. and Rosenberg, J. 2012. False-Negative Diagnoses at Stereotactic Vacuum-Assisted Needle Breast Biopsy: Long-Term Follow-Up of 1,280 Lesions and Review of the Literature. *American Journal of Roentgenology.* 192(2):341–351.
36. Wen, X. and Cheng, W. 2013. Nonmalignant breast papillary lesions at core-needle biopsy: a meta-analysis of underestimation and influencing factors. *Annals of Surgical Oncology.* 20(1):94–101.
37. Elmore, J.G., Nelson, H.D., Pepe, M.S., Longton, G.M., Tosteson, A.N.A., Geller, B., Onega, T., Carney, P.A., Jackson, S.L., Allison, K.H. and Weaver, D.L. 2016. Variability in Pathologists' Interpretations of Individual Breast Biopsy Slides: A Population Perspective Variability in

Interpretations of Individual Breast Biopsy Slides. *Annals of Internal Medicine*. 164(10):649–655.