



REVIEW ARTICLE – BREAST ONCOLOGY

Diagnostic Accuracy of Nipple Aspirate Fluid Cytology in Asymptomatic Patients: A Meta-analysis and Systematic Review of the Literature

Natasha Jiwa, BSc, MRCS¹, Rishikesh Gandhewar, BSc¹, Hemali Chauhan, BSc, MRCS¹, Hutan Ashrafiyan, PhD, MRCS, MBA¹, Swathica Kumar², Corrina Wright, MB BCh, BAO³, Zoltan Takats, PhD¹, and Daniel Richard Leff, MS (Hons), PhD, FRCS¹

¹Department of Surgery and Cancer, Imperial College London, London, UK; ²Royal College of Surgeons Ireland, Dublin, Ireland; ³Imperial College Healthcare Trust, London, UK

ABSTRACT

Purpose. To calculate the diagnostic accuracy of nipple aspirate fluid (NAF) cytology.

Background. Evaluation of NAF cytology in asymptomatic patients conceptually offers a non-invasive method for either screening for breast cancer or else predicting or stratifying future cancer risk.

Methods. Studies were identified by performing electronic searches up to August 2019. A meta-analysis was conducted to attain an overall pooled sensitivity and specificity of NAF for breast cancer detection.

Results. A search through 938 studies yielded a total of 19 studies. Overall, 9308 patients were examined, with cytology results from 10,147 breasts [age (years), mean \pm SD = 49.73 \pm 4.09 years]. Diagnostic accuracy meta-analysis of NAF revealed a pooled specificity of 0.97 (95% CI 0.97–0.98), and sensitivity of 0.64 (95% CI 0.62–0.66).

Conclusions. The diagnostic accuracy of nipple smear cytology is limited by poor sensitivity. If nipple fluid assessment is to be used for diagnosis, then emerging

technologies for fluid biomarker analysis must supersede the current diagnostic accuracy of NAF cytology.

Methods for early breast cancer detection enable lesions to be treated at the earliest possible time point, increasing survival and improving patient outcomes (e.g. lumpectomy versus mastectomy). Early identification impacts surgical treatment plans. This is exemplified by breast cancer clinical outcome measures data, which indicate that a greater proportion of screening cancers undergo breast conserving surgery when compared with patients who present symptomatically.¹

The breast screening programme, however, has certain limitations. Screening is conducted once every 3 years in the UK and only detects two-thirds of all breast cancers that arise during the 3-year interval.² Therefore, one-third of patients will present with ‘interval breast cancers’ between two screening mammograms and half of these occur in the 3rd year after screening.² One in 2500 screened women will have a missed cancer.³ The reasons for missed cancers include breast density, absence of the radiographic classification of cancer or subtle radiographic signs (often seen in invasive lobular carcinomas), poor technique and misinterpretation.^{4,5} Digital mammography has a sensitivity of 89% but a specificity of only 72%.⁶ High-risk women undergo more frequent mammograms, which carries radiation risk, or regular MRIs, which have a number of contra-indications as well as a cost implication.⁷ In addition, data on screening uptake indicates that uptake to the screening programme has been decreasing over the last 10 years, approaching the acceptable 70% mark.⁸

Electronic supplementary material The online version of this article (<https://doi.org/10.1245/s10434-020-09313-9>) contains supplementary material, which is available to authorized users.

© The Author(s) 2020

First Received: 13 August 2020

Accepted: 13 October 2020;

Published Online: 9 November 2020

N. Jiwa, BSc, MRCS

e-mail: natasha.jiwa04@imperial.ac.uk

As we move into an era of personalised screening, we need techniques that better estimate an individual's breast cancer risk. Whilst models for estimating breast cancer risk do exist, such as the Gail,⁹ Claus¹⁰ and Tyrer–Cuzick models,¹¹ they are not without limitations.^{12–14} This is primarily due to limited discrimination accuracy.¹⁵ The Gail model, for example, tends to underestimate risk as it only takes into account first-degree relatives and does not consider age of onset of cancer.^{11,16} The Claus model, although accounting for many of the genetic variations of risk, does not consider any non-hereditary risk factors and reflects women in the USA in the 1980s, mirroring a population with an outdated incidence of breast cancer¹⁶ in both North America and Europe.^{17,18} Although the Tyrer–Cuzick model adjusts for many of these genetic factors, it focuses predominantly on penetrance of BRCA1/2 mutations rather than the plethora of non-BRCA mutations^{14,19} and tends to over-estimate risk in women with atypical hyperplasia.¹³ As has been suggested with breast density,¹⁵ this begs the question as to whether adding further adjuncts such as nipple fluid cytological findings, if predictive findings are significant, could strengthen current risk assessment models and offer a personalised approach to risk screening in the future.

Cytopathology was invented and utilised in the 1920s by George Papanicolaou for early diagnosis of reproductive tract cancer, where he was able to readily distinguish between normal and malignant cells in the cervix.²⁰ Cytological evaluation of breast tissue has been ongoing since the 1950s,²¹ illustrating its cellular composition using a five-tiered classification.²² Its potential as a screening tool for nipple fluid and has been evaluated by several groups.^{23–28} However, the evidence base for the relative risk of breast cancer following abnormal cytology comes largely from breast tissue fine needle aspirate cytology or breast biopsies.^{29–31} Whether the two can be compared is controversial. Atypia in nipple fluid is different to ADH in tissue biopsies, for example. Atypia is more suggestive and can be due to degenerative cellular changes, whereas ADH in tissue biopsies is based on established links to cancer from long-term follow-up studies.³²

With the growth of metabolomics over the past decade, the study of nipple aspirate fluid has expanded to investigate the presence and quantities of a variety of substances, including proteins, lipids, DNA and the microbiome.^{33–36} This has allowed detailed evaluation of nipple fluid biomarkers, from which personalised risk screening is possible. However, nipple fluid cytology is the starting point for all future tests. As the current gold standard clinical test, information about both its diagnostic capabilities and predictive validity are vital. To the authors' knowledge, a meta-analysis of studies containing diagnostic data on the accuracy of nipple aspirate fluid obtained

by various methods has not been conducted to date. With the explosion of tools available for examining nipple fluid, and the relatively small number of studies that report the predictive capability of abnormal cells in nipple aspirate fluid, these data are more important than ever in the assessment of nipple fluid's diagnostic and screening potential.

The aim of this study was to systematically review the published literature to evaluate the diagnostic accuracy of nipple aspirate fluid cytology, against which the performance of new and emerging diagnostic and screening technologies may be compared.

METHODS

Literature Search

The literature review was conducted as per guidelines for the 'Preferred Reporting Items for Systematic Reviews and Meta-analyses' (PRISMA). The literature search was conducted using MEDLINE, EMBASE and SCOPUS databases. Relevant studies were identified using MeSH terms of key phrases from the research question (up to date as of October 2019). They were specific to 'nipple aspirate fluid' in various formats AND 'cytology' in its various forms. The following Medical Subject Headings (MeSH) and key words were used in combination with AND/OR operators: 'nipple aspirate fluid' OR ('breast' OR 'nipple' adjacent to 'secretion*' OR 'aspirat*') with ('cytodiagnosis' OR 'cytolog*' OR 'papanicolaou' OR 'cytodiagnos*').

Inclusion and Exclusion Criteria

Title and abstract review was conducted according to the pre-defined inclusion and exclusion criteria as detailed below, for each part of the review.

Inclusion Criteria

Studies reporting data on the diagnostic potential of the cytology of nipple fluid were included in the review. Both radiological and tissue histopathological confirmation of diagnosis were accepted as diagnostic modalities for validation of breast cancer. All methods of retrieving nipple fluid, including ductoscopy and ductal lavage were included, as well as those of various study designs, i.e. asymptomatic/high-risk patients and those with a known diagnosis of cancer. Only studies with full text available in the English language on human subjects, until August 2019 were included.

Exclusion Criteria

Articles were excluded by title and abstract screening if they were review articles or carried no original/primary data, had an irrelevant research question or involved processing fine needle aspirate cytology (FNAC) of breast tissue rather than the nipple aspirate fluid itself. Papers that were conducted outside a clinical environment were excluded, as well as studies without histopathological or follow-up radiological data to confirm cytological diagnoses. Diagnostic accuracy data such as true positive, true negative, etc. were not compulsory and were calculated from the raw data provided where possible.

Study Quality

Study quality was evaluated using the 'Quality Assessment of Diagnostic Accuracy Studies 2' (QUADAS-2) scoring system checklist,³⁷ and was conducted by two independent investigators (NJ and RG). All the QUADAS-2 questions were included in the scoring, providing a maximum score 14.³⁷ Each question was given a score of 0, 1 or 2 depending on whether the question was not answered, unclearly answered or fully answered. To consider the study accurately conducted and analysed, the studies had to report the type of patient included in the study (symptomatic, asymptomatic, high-risk or post-operative). The cytopathologist interpreting the results had to be defined (i.e. 1 or 2 pathologists; independently reporting) and it had to be stated whether they were blinded to the clinical results.

Data Collection

Independent assessment by two investigators (NJ and RG) was conducted using Covidence (Veritas Health Innovation, Melbourne, Australia) systematic review software.³⁸ Any conflicts were discussed and resolved with explanations of 'yes', 'no' or 'uncertain'. All 'uncertain' cases underwent full-text screening: justification for inclusion or exclusion was documented within the system (see supplementary table) and was discussed with senior authors (HA and DRL). Demographic and accuracy data from the included studies was recorded using a pre-defined Excel spreadsheet. In particular, data was collected on: (1) first author and year of publication; (2) number of patients; (3) true positives; (4) true negatives; (5) false positives; (6) false negatives; (7) average age of patient; (8) total QUADAS-2 scoring; (9) method of collection; (10) sensitivity; (11) specificity and (12) positive predictive values. Following data extraction, studies were subdivided by their method of collection, e.g. ductal lavage, manual compression etc. for subsequent sub-analysis of sensitivity and

specificity by method. Sensitivity and specificity data were calculated to 2 decimal points. Median \pm IQR age (years) was recorded where available, otherwise mean \pm SD age was used.

Meta-analysis

A bivariate and hierarchical model was used to calculate the overall diagnostic accuracy of the studies included in the meta-analysis. This allowed identification of any statistical differences between the models and provided an internal cross-reference for results produced. The sensitivity and specificity of the results was assessed using a hierarchical summary receiver-operating characteristic (HSROC) model. Pooled diagnostic sensitivity and specificity was calculated using 11 and 19 of the studies alike (all studies with sensitivities of 0 were excluded). A diagnostic HSROC curve was produced for 6 studies from within the 19 included and this allowed demonstration of the diagnostic performance of cytology of nipple aspirate fluid. For utilisation of this model, studies containing the values '0' and '1' were excluded. The trapezoid rule was utilised to calculate the pooled area under the curve. In the model, a value 1.0 indicates a perfect test with 100% accuracy and 0.5 indicates that the test is equally likely as it is unlikely to be true.

RESULTS

Malignant Diagnostic Cytology

A total of 19 studies^{24,39–56} were included in the diagnostic arm of the systematic review. These all contained clinical data on the diagnoses acquired from NAF cytology, which were correlated with either imaging or histology following a biopsy. From this, sensitivity and specificity data were either extracted or calculated. Results for 19 studies that met the criteria are included in Table 1. The publication dates included in these studies ranged from 1958 to 2009, statistics summarised in Table 2. Mean or median age was available for 16 of the 19 studies with overall ages ranging from 20 to 87. The mean or median ages ranged from 40.3 to 57.0. The calculated true positives, true negatives, false positives and false negatives, as well as the sensitivity, specificity and positive predictive value for each study are also included in Table 1. The overall sensitivity of cytology was calculated to be 0.64 (95% CI 0.62–0.66) (Fig. 1) and the specificity was 0.97 (95% CI 0.97–0.98) (Fig. 2). The diagnostic accuracy curve for these studies is illustrated in Fig. 3a, b.

TABLE 1 Demographics and outcome data—studies containing diagnostic data for meta-analysis

Author, year	n	tp	tn	fp	fn	age	qtot	meth	sens	spec	Ppv
Wood et al. 2009	23	5	18	0	13	51.0	10	1	0.3	1.0	1.0
West et al. 2006	22	0	8	4	0	51.5	12	1	0.0	0.7	0.0
Khan et al. 2005	20	0	7	0	5	54.2	12	1	0.0	1.0	0.0
Krishnamurthy et al. 2003	91	5	8	0	31	54.0	10	2	0.1	1.0	1.0
Dooley et al. 2001	507	0	370	40	0	51.9	8	5	0.0	0.9	0.0
Sauter et al. 1999	95	14	32	1	25	52.0	10	3	0.4	1.0	0.9
Zimmerman et al. 1977	4685	8	190	30	3		8	4	0.7	0.9	0.2
King et al. 1983	796	20	16	3	20		6	3	0.5	0.8	0.9
Sauter et al. 1997	152	4	20	2	19	51.7	14	3	0.2	0.9	0.7
Sauter et al. 2007	177	28	54	6	72		8	3	0.3	0.9	0.8
Konstandiadou et al. 2012	80	0	78	2	0	45.8	10	1	0.0	1.0	0.0
Loud et al. 2009	171	0	92	0	92	40.3	12	1	0.0	1.0	0.0
Visvanathan et al. 2007	69	0	25	3	0	46.6	10	1	0.0	0.9	0.0
Bushnaq et al. 2007	150	10	264	17	0	48.0	8	1	1.0	0.9	0.4
Danforth et al. 2006	25	0	16	5	0	57.0	6	1	0.0	0.8	0.0
Mitchell et al. 2005	52	0	17	4	4	43.0	12	5	0.0	0.8	0.0
Redlich et al. 2004	37	1	19	5	0	51.7	10	1	1.0	0.8	0.2
Papanicolaou et al. 1958	412	1	607	1	0	47.3	8	2	1.0	1.0	0.5
Buerhring et al. 2006	744	1	529	7	8		12	5	0.1	1.0	0.1

N, number of patients; tp, true positives; tn, true negatives; fp, false positives; fn, false negatives; qtot, QUADAS total score; meth, method of collection of nipple fluid (1: ductal lavage, 2: manual compression, 3: manual pump, 4: not specified, 5: two methods combined); sens, sensitivity; spec, specificity; ppv, positive predictive value

TABLE 2 Summary statistics of the 19 studies included and overall diagnostic accuracy data of nipple aspirate fluid cytology

Summary statistics

Total number of breasts: 10,147	Methods of collection: 11 DL; 6MP; 4MC; 3DT; 1NS
Total number of patients: 9308	Non-yielder: $30.5\% \pm 26.4$ (mean)
Mean age: 49.73 ± 4.09 years	Insufficient sample: 38.87% (mean)

Parameter	Estimate	95% CI
Sensitivity ^a	0.64	0.621–0.659
Specificity ^b	0.971	0.966–0.977
Positive likelihood ratio	4.70	2.80–7.80
Negative likelihood ratio	0.74	0.63–0.87
Diagnostic odds ratio	6.00	4–11

Overall summary statistics from the studies included in the meta-analysis

DL, ductal lavage; MP, manual pump; MC, manual compression; DT, dual technique; NS, not specified

^a11 of the 19 studies were utilised to calculate the overall sensitivity (studies that reported a sensitivity of 0 were excluded)

^bAll 19 studies were included to calculate the overall specificity

DISCUSSION

This meta-analysis assimilates the diagnostic accuracy of nipple aspirate fluid cytology across published clinical studies as well as the future risk of developing breast cancer depending on the type of previous nipple aspirate

fluid cytology finding. The results reveal an overall low sensitivity of 0.64 (95% CI 0.62–0.66) and high specificity of 0.97 (95% CI 0.97–0.98), when NAF cytology is used as a diagnostic tool.

FIG. 1 ES, effect size; 95% CI, 95% confidence interval; % weight, percentage weight carried by the study calculated from the random effects analysis. Overall sensitivity from 11/19 studies 0.64 [0.62–0.64]

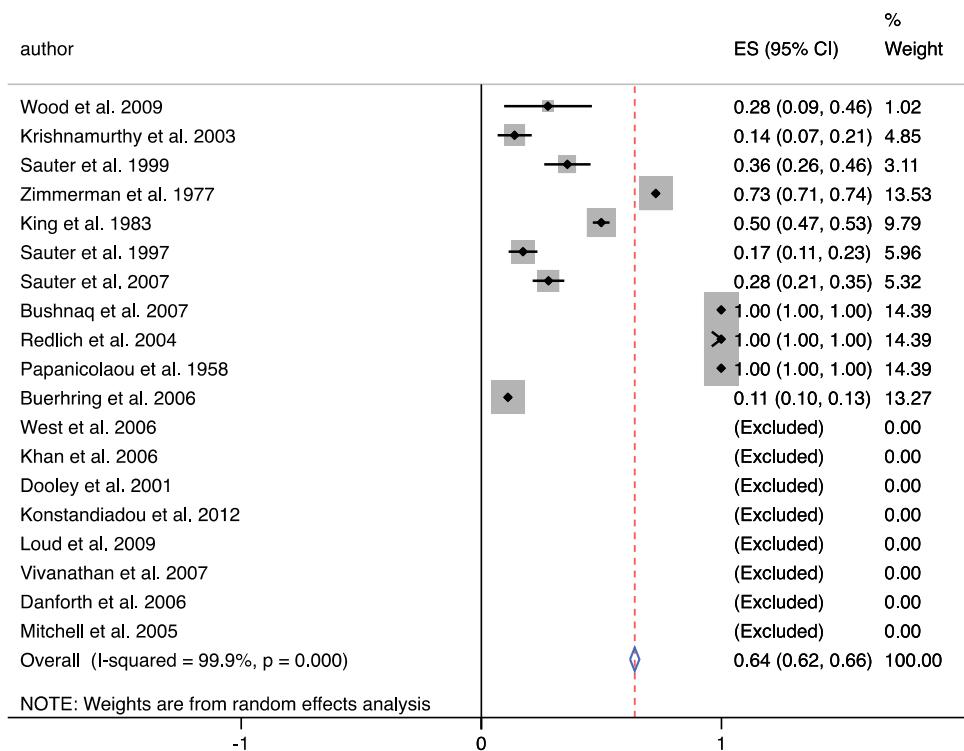
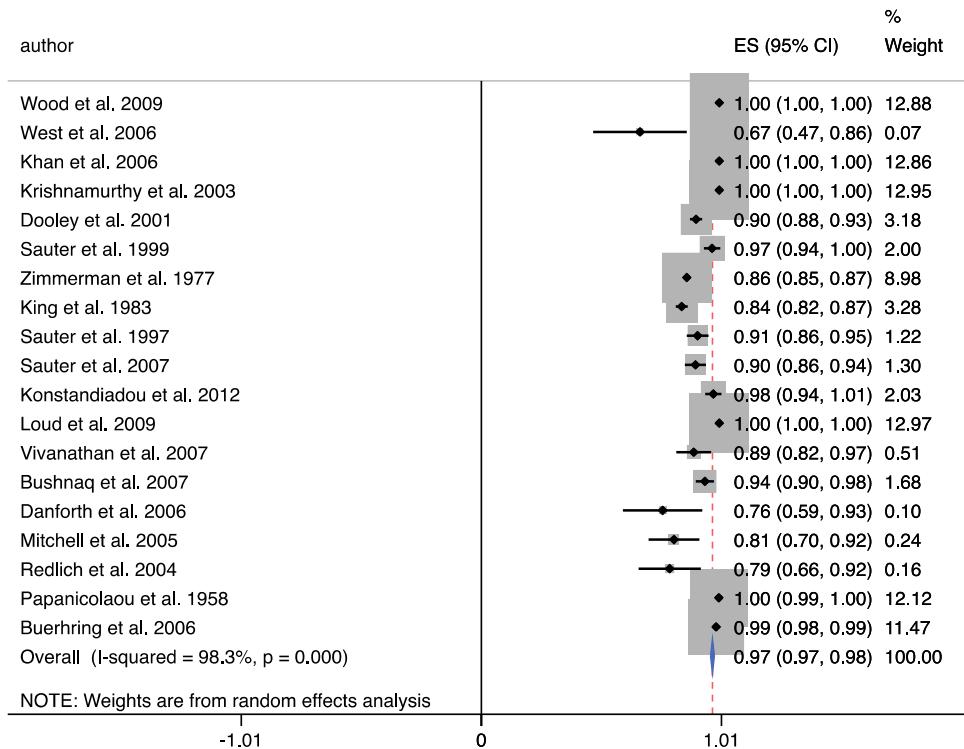


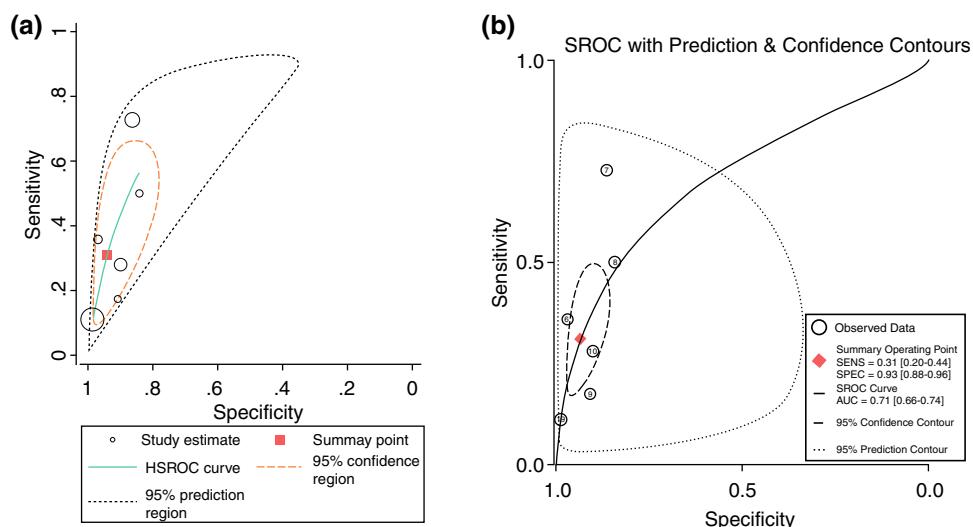
FIG. 2 ES, effect size; 95% CI, 95% confidence interval; % weight, percentage weight carried by the study calculated from the random effects analysis. Overall specificity from 19/19 studies 0.97 [0.97–0.98]



Despite the continued use of cytology for evaluating symptomatic nipple discharge fluid, NAF cytology has not successfully penetrated clinical practice as a method of assessment for asymptomatic women at high risk of breast cancer, or indeed as an adjunct to screening. The reasons

for this may include both the limited diagnostic accuracy in this setting, as highlighted by our findings, as well as the challenges involved in collection of fluid, in particular the potential for non-producers, the fact that not all terminal duct-lobular units drain to the nipple,⁵⁷ the proportion of

FIG. 3 **a** Meta-analysis HSROC curve for cytology and **b** prediction and confidence contours for the six studies included in the meta-analysis. The x-axis demonstrates 1-specificity (the true negatives) and the y-axis demonstrates the sensitivity (the true positives). The curve delineates the true positive rate of nipple cytology at each true negative value. A perfect test is one in which the results are closest to 1, i.e. 100% accuracy



smears deemed inadequate,⁵⁸ and whether methods for NAF harvest are deemed acceptable by patients and the public.⁵⁹

The primary finding of this meta-analysis is that the sensitivity of NAF cytology for cancer detection in asymptomatic women is poor [0.64 (95% CI 0.62–0.66)] and yet the specificity is extremely high [0.97 (95% CI 0.97–0.98)]. Overall, high specificity and low sensitivity data is echoed across individual studies and also in those presenting with symptomatic nipple discharge,^{60,61} with a diagnostic accuracy for malignancy of only 31%.⁶⁰ The reason for such low sensitivity of nipple aspirate fluid is that discharge is often “physiological” in nature, has a generally high acellular or paucicellular composition,^{58,62,63} and therefore is thought to consist almost exclusively of background debris and/or proteinaceous material.^{58,64}

Methods for NAF evaluation, whether they be cytological assessment or otherwise, suffer as a result of the relatively high proportion of “non-yielders”: women in whom techniques to obtain NAF fail. Statistics from all the studies included in this review demonstrate a non-yielding percentage of $30.5\% \pm 26.4\%$ across the various methods of fluid collection of nipple aspirate fluid. The lowest non-yielders were those that underwent ductal lavage studies.^{40,65,66} However, ductal lavage is subject to selection bias, as those in whom cannulation for ductoscopy was not possible are unlikely to have been selected for ductal lavage. Similarly, a low non-yielder rate was observed in those that underwent expression using a manual pump.^{46,66–68} In these studies, non-yielders ranged from 0 to 5.9%, and three out of four of the studies utilised a modified breast pump.^{67–69} However, these samples were acquired from mastectomy specimens, immediately following its excision from the chest wall, rather than from the awake patient. This allowed for a reasonable amount of negative pressure to be applied to the nipple-areolar

complex without concern of causing pain or discomfort. These figures, whilst likely to be an underestimation of non-yielding status using the same method in a patient who is conscious, provide reassuring evidence that physiological nipple aspirate fluid is present and easily accessible in patients under general anaesthetic.

One of the greatest limitations of nipple fluid cytology according to the results from the studies included in the current analysis, is the percentage of samples produced that are ‘inadequate’. Here we demonstrate a mean of 38.9% of analysed samples that were deemed inadequate. This can often be the case with ductal lavage samples, in which fluid content is diluted by saline. This has a major cost implication. If almost 40% of all samples processed are inadequate, and the cost of processing and reporting each slide is £75, then this represents substantial waste and ultimately reduces the value of the test. Repeating NAF assessments is also problematic. For example, ductal lavage may be painful⁷⁰ or difficult to tolerate under topical/local anaesthetic, and manual pumps and/or automated pumps may cause discomfort or skin surface irritation.²⁶ In addition, the process of collecting and reporting smears is confounded by inappropriately prepared slides in the absence of a cytopathologist in the outpatient department, which can lead to sub-optimal slides due to air drying artefacts, contamination or inadequate fluid distribution. Reporting may be subject to inter-reporter variability or relative inexperience, as well as the presence of atypical cellular changes unrelated to a malignancy, leading to either a higher degree of false positives or false negative findings.

Next is the challenge of proportion of ducts whose biocomposition can accurately be evaluated within NAF. Most breast cancers arise from the epithelial lining of the terminal ducts—invading ductal carcinomas. NAF therefore provides a mirror of what is occurring in the tumour

microenvironment or, in high-risk individuals, in the lead-up to cancer.⁷¹ However, it has also been shown that not all ducts drain to the nipple surface,⁵⁷ suggesting that even if NAF cytological analysis has superior diagnostic accuracy to that demonstrated here it could still miss a proportion of breast cancers.

As an important adjunct, the predictive validity of nipple aspirate fluid has been investigated by Wrensch and her group^{72–76} at various time points between 1992 and 2010, using several patient groups recruited at different time points. Results have yielded an overall threefold increase in the risk of breast cancer with a NAF finding of atypia. Although findings are limited by a potential overlap in patient cohorts within the studies and therefore a duplication of datasets, a threefold increase in risk of developing breast cancer should not be overlooked. Long-term follow-up studies, such as that by Page et al.,³² highlight that detection of ALH/ADH has an associated 4–5 times risk of breast cancer. However, there is a relative scarcity of similar evidence with nipple smear cytology and even ductal lavage.⁷⁷ Evidence tends to be almost completely extrapolated from FNAC studies with tissue^{29–32} and, in fact, it is this body of evidence which has supported commercial products such as HALO (NeoMatrix, Irvine, CA), the nipple aspiration device. In nipple fluid, atypia is more suggestive than diagnostic and can even be due to degenerative cellular changes. Its significance is therefore inconclusive and would have to be carefully considered prior to incorporation into existing and established models of risk prediction, such as the Gail model or Tyrer–Cuzick. Integrating a clinical assessment tool into an existing model could both strengthen and personalise risk in asymptomatic, high-risk women, as illustrated by Vilmun et al.¹⁵ when reviewing the impact of adding breast density to breast cancer risk models.⁷⁸ In order to offer a personalised approach to risk screening, implementation of a stratification protocol offering varying screening regimes according to their elicited risk, following the interrogation of nipple smear cytology, would be required to be undertaken in various patient cohorts. This would be undertaken with a view to a reduction in the diagnosis of interval cancers, an increase in the diagnosis of early breast cancer and a reduction in mortality. For example, the “WISDOM trial”⁷⁹ was a multi-centre randomised controlled trial that allowed for both risk-based and observational screening. It takes into account personal and genetic risk, including mutations such as BRCA1/2, ATM, CDH1, CHEK2, PALB2, PTEN, STK11 and TP52, as well as a polygenic risk score from 96 lower-risk common genetic variants (SNPs) with known association with breast cancer, and an updated polygenic risk model, including ethnicity- and race-specific SNPs that are shown to confer risk, to calculate a personalised risk score.

Moreover, with promising early data emerging from the interrogation of nipple fluid using innovative technologies such as mass spectrometry, as well as the known limitations of cytology, it is reasonable to suggest that various metabolites found within nipple fluid have great diagnostic potential. With the growth of metabolomics over the past decade, the study of nipple aspirate fluid has expanded to investigate the presence and quantities of a variety of substances, including proteins, lipids, DNA and the microbiome. Proteomic analysis of nipple aspirate fluid works on the principle that it contains a concentrated source of proteins from cancerous ducts, which may identify tumour-specific protein patterns.⁸⁰ In 2004, Alexander et al.⁸¹ identified candidate markers using matrix-assisted laser desorption ionization time-of-flight (ESI Q-TOF) proteomic analysis and validated the markers identified using quantitative, high-throughput ELISA analysis. Among their subjects, GCDFP-15 levels were lower and AAG levels correlated with presence and stage of breast cancer disease. Similarly, in 2007, He et al.⁸² identified a set of 8 protein markers which collectively gave a 63% sensitivity, 89% specificity and 76% accuracy for distinguishing between cancer and normal pathologies.⁸² Further work on the proteome was conducted by Pavlou et al.³⁶ in 2010. The authors utilised LCMS-MS to generate an extensive nipple aspirate fluid proteome identifying over 800 unique proteins, more than 50% of which were extracellular plasma membrane proteins. In 2017 Shaheed et al.⁸³ used manual expressing techniques to acquire NAF samples for proteomic analysis with 2D LC-MS separation. The results demonstrated that in the majority of individuals with bilateral samples, paired samples illustrated protein profiles that were similar. They reported an average of 1374 proteins per sample with significant progress, identifying 1374 new proteins from those previously seen by Pavlou et al.³⁶ in NAF. In terms of lipidomics, a study by Matos Do Canto et al.³³ in 2016 identified up to 83 ions with a fold change — metabolites included endogenous metabolites such as amino acid derivatives, products of lipid metabolism, glycerophospholipids and phosphatidylserine. It is the first known study that demonstrates the feasibility of conducting a comprehensive metabolomic profiling of breast tumours using ductal lavage. Due to the contemporaneous nature of these studies, the process of biomarker development is still in its early stages and is therefore a promising field of study.

CONCLUSION

The current systematic review and meta-analysis provides new diagnostic accuracy data for nipple aspirate fluid cytology, including pooled data overall, whilst taking into

account the collection method. The results demonstrate that the diagnostic accuracy of nipple fluid cytology is limited due to poor sensitivity secondary to a paucicellular material. Emerging techniques for surveillance and screening of patients who carry a risk of breast cancer will need to have a personalised approach and surpass the present diagnostic accuracy of cytology, whilst taking into account cost effectiveness, reproducibility of results, user dependency and turn-around time in the laboratory. The sensitivity and specificity should be high enough to warrant further assessment in the form of imaging or a confirmatory biopsy (histopathology).

ACKNOWLEDGEMENT The research was funded by the NIHR Imperial Biomedical Research Centre (BRC). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

AUTHOR CONTRIBUTIONS Natasha Jiwa and Daniel Leff conceived the ideas and framework of the paper. Natasha Jiwa, Rishikesh Gandhewar and Swathica Kumar were involved in the study design and systematic searches. Hutan Ashrafiyan performed the statistical analyses tests for the meta-analysis. Natasha Jiwa and Hemali Chauhan prepared the manuscript. Corina Wright, Zoltan Takats and Daniel Leff prepared and reviewed the manuscript as well as interpreted the dataset.

FUNDING The research was funded by the NIHR Imperial Biomedical Research Centre (BRC). No commercial interest, no external financial or material support.

DISCLOSURES N. Jiwa is a Ph.D. candidate investigating nipple biofluid in the early detection of breast cancer, under the supervision of D. Leff and Z. Takats.

OPEN ACCESS This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- West Midlands Cancer Intelligence Unit. Breast Cancer Clinical Outcome Measures (BCCOM) Project. 2007. https://associationofbreastsurgery.org.uk/media/63921/final-bccom_ar07_prf5.pdf.
- Dixon JM, Barber M. *Breast Surgery: A Companion to Specialist Surgical Practice, Sixth Edition*. Elsevier, Amsterdam, 2018.
- Cancer Research UK. *Breast Cancer Statistics*. 2015.
- Martin JE, Moskowitz M, Milbrath JR. Breast cancer missed by mammography. *Am J Roentgenol*. 1979;132(5):737–9.
- Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. *Radiographics*. 2003;23(4):881–95.
- Lei J, Yang P, Zhang L, Wang Y, Yang K. Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis. *Eur Radiol*. 2014;24(3):595–602.
- Saslow D, Boetes C, Burke W et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin*. 2007;57(2):75–89.
- Public Health England Screening. Breast screening performance reports show we are detecting cancers early but uptake is falling.
- Tice JA, Miike R, Adduci K et al. Nipple aspirate fluid cytology and the Gail model for breast cancer risk assessment in a screening population. *Cancer Epidemiol Biomark Prev*. 2005;14(2):324–8.
- Jacobi CE, de Bock GH, Siegerink B, van Asperen CJ. Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose? *Breast Cancer Res Treatment*. 2009;115(2):381–90.
- Amir E, Evans D, Shenton A et al. Evaluation of breast cancer risk assessment packages in the family history evaluation and screening programme. *J Med Genet*. 2003;40(11):807–14.
- Euhus DM, Leitch AM, Huth JF, Peters GN. Limitations of the Gail model in the specialized breast cancer risk assessment clinic. *Breast J*. 2002;8(1):23–7.
- Boughey JC, Hartmann LC, Anderson SS et al. Evaluation of the Tyrer–Cuzick (International Breast Cancer Intervention Study) model for breast cancer risk prediction in women with atypical hyperplasia. *J Clin Oncol*. 2010;28(22):3591.
- Himes DO, Root AE, Gammon A, Luthy KE. Breast cancer risk assessment: calculating lifetime risk using the Tyrer–Cuzick model. *J Nurse Pract*. 2016;12(9):581–92.
- Vilmun BM, Vejborg I, Lyng E et al. Impact of adding breast density to breast cancer risk models: a systematic review. *Eur J Radiol*. 2020;109019.
- Evans DGR, Howell A. Breast cancer risk-assessment models. *Breast Cancer Res*. 2007;9(5):213.
- Breastcancer.org. U.S Breast Cancer Statistics: 2020. Available from: https://www.breastcancer.org/symptoms/understand_bc/statistics.
- Office for National Statistics. Cancer registration statistics, England: 2017. 2019;1–16. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017>.
- Fasching PA, Bani MR, Nestle-Krämling C et al. Evaluation of mathematical models for breast cancer risk assessment in routine clinical use. *Eur J Cancer Prev*. 2007;16(3):216–24.
- Tan SY, Tatsumura YJS. George Papanicolaou (1883–1962): discoverer of the Pap Smear. *Singapore Med J*. 2015;56(10):586.
- Papanicolaou GN, Bader GM, Holmquist DG, Falk EA. Cytologic evaluation of breast secretions. *Ann NY Acad Sci*. 1956;63(6):1409–21.
- Cibas ES, Ducatman BS. *Cytology: diagnostic principles and clinical correlates*. Elsevier, Amsterdam, 2013.
- Fabian CJ, Kimler BF. Breast cancer risk prediction: should nipple aspiration fluid cytology be incorporated into clinical practice? *J Natl Cancer Inst*. 2001;93(23):1762–63.
- Dooley WC, Ljung B-M, Veronesi U, et al. Ductal lavage for detection of cellular atypia in women at high risk for breast cancer. *J Natl Cancer Inst*. 2001;93(21):1624–32.
- Filassi JR, Zonta MA, Trinconi A, et al. Can breast nipple fluid collected with automated aspiration and preserved in based-liquid solution improve the cytological samples? *Acta Cytol*. 2013;57(3):276–80.

26. Proctor KA, Rowe LR, Bentz JS. Cytologic features of nipple aspirate fluid using an automated non-invasive collection device: a prospective observational study. *BMC Women's Health.* 2005;5(1):10.
27. Sartorius OW, Smith HS, Morris P, Benedict D, Friesen L. Cytologic evaluation of breast fluid in the detection of breast disease. *J Natl Cancer Inst.* 1977;59(4):1073–80.
28. Twelves D, Nerurkar A, Osin P, Ward A, Isacke CM, Gui GPH. The feasibility of nipple aspiration and duct lavage to evaluate the breast duct epithelium of women with increased breast cancer risk. *Eur J Cancer.* 2013;49(1):65–71.
29. Hartmann LC, Sellers TA, Frost MH et al. Benign breast disease and the risk of breast cancer. *N Engl J Med.* 2005;353(3):229–37.
30. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med.* 1985;312(3):146–51.
31. Fabian CJ, Kimler BF, Zalles CM, et al. Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model. *J Natl Cancer Inst.* 2000;92(15):1217–27.
32. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer.* 1985;55(11):2698–708.
33. Matos Do Canto L, Marian C, Varghese RS et al. Metabolomic profiling of breast tumors using ductal fluid. *Int J Oncol.* 2016;49(6):2245–54.
34. Shaheed SU, Tait C, Kyriacou K, Linforth R, Salhab M, Sutton C. Evaluation of nipple aspirate fluid as a diagnostic tool for early detection of breast cancer. *Clin Proteomics.* 2018;15:3.
35. Chan AA, Bashir M, Rivas MN et al. Characterization of the microbiome of nipple aspirate fluid of breast cancer survivors. *Nature.* 2016;6:28061.
36. Pavlou MP, Kulasingam V, Sauter ER, Kliethermes B, Diamandis EP. Nipple aspirate fluid proteome of healthy females and patients with breast cancer. *Clin Chem.* 2010;56(5):848–55.
37. Whiting PF, Rutjes AW, Westwood ME et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Int Med.* 2011;155(8):529–36.
38. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia.
39. Wood ME, Stanley MA, Crocker AM, Kingsley FS, Leiman G. Ductal lavage of cancerous and unaffected breasts: procedure success rate and cancer detection. *Acta Cytolog.* 2009;53(4):410.
40. West KE, Wojcik EM, Dougherty TA, Siziopikou KP, Albain KS, Gabram SGJTAjtos. Correlation of nipple aspiration and ductal lavage cytology with histopathologic findings for patients before scheduled breast biopsy examination. *Am J Surg.* 2006;191(1):57–60.
41. Khan SA, Wolfman JA, Segal L, Benjamin S, Nayar R, Wiley EL, et al. Ductal lavage findings in women with mammographic microcalcifications undergoing biopsy. *Ann Surg Oncol.* 2005;12(9):689–96.
42. Krishnamurthy S, Sneige N, Thompson PA, et al. Nipple aspirate fluid cytology in breast carcinoma. *Cancer.* 2003;99(2):97–104.
43. Sauter E, Ehya H, Babb J, et al. Biologic markers of risk in nipple aspirate fluid are associated with residual cancer and tumour size. *Br J Cancer.* 1999;81(7):1222.
44. Zimmerman AL, King EB, Barrett DL, Petrakis NL. The incidence and significance of intracytoplasmic calcifications in nipple aspirate specimens. *Acta Cytolog.* 1977;21(5):685–92.
45. King EB, Chew KL, Petrakis NL, Emster VL. Nipple aspirate cytology for the study of breast cancer precursors. *J Natl Cancer I.* 1983;71(6):1115–21.
46. Sauter ER, Ross E, Daly M, et al. Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk. *Br J Cancer.* 1997;76(4):494–501.
47. Sauter ER, Wagner-Mann C, Ehya H, Klein-Szanto A. Biologic markers of breast cancer in nipple aspirate fluid and nipple discharge are associated with clinical findings. *Cancer Detect Prev.* 2007;31(1):50–8.
48. Konstandiadou I, Mastoraki A, Kotsilianou O et al. Does ductal lavage assert its role as a noninvasive diagnostic modality to identify women at low risk of breast cancer development? *Gynecol Oncol.* 2012;123(2):110–4.
49. Loud JT, Thiebaut AC, Abati AD et al. Ductal lavage in women from BRCA1/2 families: is there a future for ductal lavage in women at increased genetic risk of breast cancer? *Cancer Epidemiol Biomark.* 2009;18(4):1243–51.
50. Visvanathan K, Santor D, Ali S et al. The reliability of nipple aspirate and ductal lavage in women at increased risk for breast cancer—a potential tool for breast cancer risk assessment and biomarker evaluation. *Cancer Epidemiol Biomark.* 2007;16(5):950–5.
51. Bushnaq ZI, Ashfaq R, Leitch AM, Euhus D. Patient variables that predict atypical cytology by nipple duct lavage. *Cancer.* 2007;109(7):1247–54.
52. Danforth DN, Abati A, Filie A et al. Combined breast ductal lavage and ductal endoscopy for the evaluation of the high-risk breast: a feasibility study. *J Surg Oncol.* 2006;94(7):555–64.
53. Mitchell G, Antill YC, Murray W et al. Nipple aspiration and ductal lavage in women with a germline BRCA1 or BRCA2 mutation. *Breast Cancer Res.* 2005;7(6):1122–31.
54. Redlich PN, Purdy AC, Shidham VB et al. Feasibility of breast duct lavage performed by a physician extender. *J Surg.* 2004;136(5):1077–80.
55. Papanicolaou GN, Holmquist DG, Bader GM, Falk EA. Exfoliative cytology of the human mammary gland and its value in the diagnosis of cancer and other diseases of the breast. *Cancer.* 1958;11(2):377–409.
56. Buehring GC, Letscher A, McGirr KM et al. Presence of epithelial cells in nipple aspirate fluid is associated with subsequent breast cancer: a 25-year prospective study. *Breast Cancer Res Treatment.* 2006;98(1):63–70.
57. Wang S-C, Gombos E, Birdwell RL, Berg WA. *Diagnostic Imaging: Breast.* Amirsys; 2006.
58. Bibbo M, Wilbur D. *Comprehensive Cytopathology.* E-Book: Elsevier Health Sciences; 2014.
59. Proctor KA, Rowe LR, Bentz JS. Cytologic features of nipple aspirate fluid using an automated non-invasive collection device: a prospective observational study. *BMC Women's Health.* 2005;5:10.
60. Dinkel H-P, Gassel AM, Müller T, Lourens S, Rominger M, Tschaumller AJO, et al. Galactography and exfoliative cytology in women with abnormal nipple discharge. *Obstet Gynecol.* 2001;97(4):625–9.
61. Kan W, Chen C, Kwong AJHKMJ. Implications of nipple discharge in Hong Kong Chinese women. *Hong Kong Med J.* 2018;24(1):18–24.
62. Yu SN, Li J, Wong SI, Tsang JYS, Ni YB, Chen J, et al. Atypical aspirates of the breast: a dilemma in current cytology practice. *J Clin Pathol.* 2017;70(12):1024–32.
63. Sakorafas G. Nipple discharge: current diagnostic and therapeutic approaches. *Cancer Treatment Rev.* 2001;27(5):275–82.
64. Das DK, Al-Ayadhy B, Ajrawi MTG, Shaheen AA, Sheikh ZA, Malik M, et al. Cytodiagnosis of nipple discharge: a study of 602 samples from 484 cases. *Diagn. Cytopatho.* 2001;25(1):25–37.
65. Bushnaq ZI, Ashfaq R, Leitch AM, Euhus DJCIIJotACS. Patient variables that predict atypical cytology by nipple duct lavage. *Cancer.* 2007;109(7):1247–54.
66. King EB, Barrett D, Petrakis NLJAJocp. Cellular composition of the nipple aspirate specimen of breast fluid: II. Abnormal findings. *Am J Clin Pathol.* 1975;64(6):739–48.

67. Sauter ER, Wagner-Mann C, Ehya H, Klein-Szanto AJCd, prevention. Biologic markers of breast cancer in nipple aspirate fluid and nipple discharge are associated with clinical findings. *Cancer Detect Prev.* 2007;31(1):50–8.
68. Sauter E, Ehya H, Babb J, Diamandis E, Daly M, Klein-Szanto A, et al. Biologic markers of risk in nipple aspirate fluid are associated with residual cancer and tumour size. *Br J Cancer.* 1999;81(7):1222.
69. Sauter ER, Ross E, Daly M, Klein-Szanto A, Engstrom PF, Sorling A, et al. Nipple aspirate fluid: a promising non-invasive method to identify cellular markers of breast cancer risk. *Br J Cancer.* 1997;76(4):494–501.
70. Loud JT, Beckjord EB, Nichols K, Peters J, Giusti R, Greene MH. Tolerability of breast ductal lavage in women from families at high genetic risk of breast cancer. *BMC Women's Health.* 2009;9.
71. Djuric Z, Chen G, Ren J, Venkatramanamoothy R, Covington CY, Kucuk O, et al. Effects of high fruit–vegetable and/or low-fat intervention on breast nipple aspirate fluid micronutrient levels. *Cancer Epidemiol Biomark Prevent.* 2007;16(7):1393–9.
72. Wrensch MR, Petrakis NL, King EB, Miike R, et al. Breast cancer incidence in women with abnormal cytology in nipple aspirates of breast fluid. *Am J Epidemiol.* 1992;135(2):130–41.
73. Wrensch M, Petrakis NL, King EB, Lee MM, Miike RJAjoe. Breast cancer risk associated with abnormal cytology in nipple aspirates of breast fluid and prior history of breast. *Biopsy.* 1993;137(8):829–33.
74. Wrensch MR, Petrakis NL, Miike R, et al. Breast cancer risk in women with abnormal cytology in nipple aspirates of breast fluid. *J Natl Cancer Inst.* 2001;93(23):1791–8.
75. Stivers D, Tice J, Wrensch M, Miike R, Berry D. Quantifying breast cancer risk from atypia and hyperplasia via nipple aspiration: A synthesis of evidence. *Ann Surg Oncol.* 2010;17:S188.
76. Baltzell KA, Moghadassi M, Rice T, Sison JD, Wrensch M. Epithelial cells in nipple aspirate fluid and subsequent breast cancer risk: a historic prospective study. *BMC Cancer.* 2008;8:75.
77. Johnson-Maddux A, Ashfaq R, Cler L, et al. Reproducibility of cytologic atypia in repeat nipple duct lavage. *Cancer.* 2005;103(6):1129–36.
78. Tice JA, Cummings SR, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. *Breast Cancer Res Treatment.* 2005;94(2):115–22.
79. Esserman LJ. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer.* 2017;3(1):1–7.
80. Debal M, Wolfgarten M, Walgenbach-Brünagel G, Kuhn W, Braun MJEJ. Non-invasive proteomics—thinking about personalized breast cancer screening and treatment. *EPMA J.* 2010;1(3):413–20.
81. Alexander H, Stegner AL, Wagner-Mann C, Du Bois GC, Alexander S, Sauter ERJCCR. Proteomic analysis to identify breast cancer biomarkers in nipple aspirate fluid. *Clin Cancer Res.* 2004;10(22):7500–10.
82. He J, Gornbein J, Shen D, Lu M, Rovai LE, Shau H, et al. Detection of breast cancer biomarkers in nipple aspirate fluid by SELDI-TOF and their identification by combined liquid chromatography–tandem mass spectrometry. *Int J Oncol.* 2007;30(1):145–54.
83. Shaheed S, Tait C, Kyriacou K, Mullarkey J, Burrill W, Patterson LH, et al. Nipple aspirate fluid—a liquid biopsy for diagnosing breast health. *Proteomics Clin Appl.* 2017;11(9–10):1700015.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.