

Full Title	Role of performance metrics in breast screening imaging - where are we and where should we be?
Article type	Review Article
Corresponding Author	Sue L Cohen MB BS, FFPH Public Health England, London, England
Corresponding Author's Institution	Public Health England Wellington House, 133-155 Waterloo Road London SE1 8UG
First Author	Cohen S
Order of Authors	Cohen S. Blanks RG, Kearins O, Jenkins J.
Abstract	<p>The NHS breast screening programme (NHSBSP) was started in 1988 and is a large organised cancer screening programme. It is delivered by 80 screening services across England and screens over 2 million women each year. As a screening programme, it must balance the detection of cancers against possible harm to women who do not have cancer. The NHSBSP was therefore designed with detailed information gathering and performance metrics right from the start. In this review paper, we examine how performance metrics in screening mammography have improved the national screening programme and the further developments and challenges that are expected in the years to come.</p>

Title Page

Role of performance metrics in breast screening imaging - where are we and where should we be?

Dr Sue L Cohen MBBS BSc FFPH

Consultant in Public Health Medicine, National Lead Screening QA Service,
Public Health England
Wellington House, 133-155 Waterloo Road
London
SE1 8UG
UK

Sue.cohen@phe.gov.uk

Tel: 07768646164

Roger Blanks

Olive Kearins

National Portfolio Lead, Breast Screening Programme & Head of QA Midlands and East, Screening QA Service
Public Health England
Wellington House, 133-155 Waterloo Road
London
SE1 8UG
UK

Jacquie Jenkins

National Programme Manager, Breast Screening Programme
Public Health England
Wellington House, 133-155 Waterloo Road
London
SE1 8UG
UK

Acknowledgements: We would like to thank Dr Matthew Wallis for his helpful comments on an early draft of this manuscript

We declare we have no conflicts of interest

Authors Contributions

	Sue Cohen	Roger Blanks	Olive Kearins	Jacque Jenkins
Guarantor of integrity of the entire study	yes	No	No	No
Study concepts and design	yes	yes	yes	yes
Literature research	yes	yes	yes	No
Clinical studies	n/a	n/a	n/a	n/a
Experimental studies/data analysis	n/a	n/a	n/a	n/a
Manuscript preparation	Yes	Yes	Yes	Yes
Manuscript editing	Yes	Yes	No	No

Abstract

The NHS breast screening programme (NHSBSP) was started in 1988 and is a large organised cancer screening programme. It is delivered by 80 screening services across England and screens over 2 million women each year. As a screening programme, it must balance the detection of cancers against possible harm to women who do not have cancer. The NHSBSP was therefore designed with detailed information gathering and performance metrics right from the start. In this review paper, we examine how performance metrics in screening mammography have improved the national screening programme and the further developments and challenges that are expected in the years to come.

Introduction

The NHS Breast Screening Programme (NHSBSP) was established in 1988 in response to the recommendations of a working group on breast cancer screening chaired by Professor Sir Patrick Forrest¹. The model proposed for England by the Forrest report was based on the results of a number of studies most notably the Swedish-Two County Study randomised controlled trial (STC RCT)² and was designed to deliver a similar reduction in mortality.

In England women aged 50 to 70 years are screened using mammography every 36 months. Women who are found to be screen positive are followed through assessment to the point of diagnosis and monitored into treatment. European countries operate a number of different models, screening women every two years and often not following up women beyond the screening result³.

Since the inception of the programme, there have been a number of important developments in the NHSBSP to improve the quality and effectiveness of the programme. Over subsequent years the programme introduced; increasing the age range from 50-64 to 50 -70 years, double reading, two views at the first then subsequent screens and digital mammography. These developments have been informed by high quality data collected along the pathway that can monitor the programme performance at national, service and more recently individual clinician/radiographer level.

Organisation of Breast Screening in England

The NHS Cancer Screening Programmes was responsible for developing the breast screening programme until this responsibility transferred to Public Health England (PHE) in 2013. Breast screening is delivered predominantly by hospital trusts, usually referring into

symptomatic services within the same hospital. There are currently 80 screening services in England.

The programme has been shaped through a process of clinical engagement with radiologists, surgeons, pathologists and other professional groups. Under the auspices of NHSBSP, these groups produced guidance which incorporated quality standards that described the parameters within which the screening services should operate.

From the outset, as recommended by the Forrest Report¹, monitoring quality and outcomes was integrated into the design of the programme.

This was achieved primarily through a network of regional Quality Assurance Reference Centres (QARCs) which were established to monitor the world's first national breast screening programme. QARCs operated autonomously, collecting data in their region to measure performance against the standards, carrying out three yearly peer review visits, assessing compliance with national guidance and driving continuous improvement in the quality of services.

In 2013 the QARCs moved to PHE and became part of the national Screening Quality Assurance Service (SQAS) thus enabling for the first time to have a single approach to measuring quality across England's breast screening services and national rather than regional benchmarking for performance metrics.

Using standards and performance metrics to drive up quality

Donabedian describes the importance of carefully constructing standards that can accurately translate concepts and attributes of quality into actual measurements.⁴ He defined standards as having a criterion of either structure, process or outcome and the *“standard to mean a specified quantitative measure of magnitude or frequency that specifies what is good or less so”*.⁴ This approach to standard setting, which includes metrics that can be accurately and

consistently collected from all services has been critical in driving up quality in the English breast screening programme.

Process standards that monitor the quality of the radiological component of the screening programme have been in place from its establishment. Originally published as part of '*Quality assurance guidelines for breast cancer screening radiology*',⁵ the radiological standards have recently been incorporated into a single publication of consolidated standards that cover the screening pathway. There are now 17 core standards from identification of the cohort to the point of diagnosis⁶

Each standard and associated metric has a performance threshold of acceptable and achievable. The standards' thresholds are periodically reviewed and can be revised upwards based on the evidence review and analysis of normative standards. This drives up quality over time.

Standards will only be effective in driving up quality if the data is accurate, reliable and can be consistently collected from all services. In England, this has been achieved by use of a single system to run the breast screening service. The National Breast Screening System (NBSS) is used in England, Wales and Northern Ireland and collects information on the woman's screening pathway. This single system with nationally defined data fields and reporting tools has enabled the breast screening programme to produce high quality data that is comparable across services.

NBSS is a rich source of data that is used to generate national reports which describe performance against standards. It can also be interrogated locally. National data are published annually in the Breast Screening Programme statistical bulletin often referred to as the Korner KC62 returns⁷. The national bulletin provides service level data aggregating the outcomes of all women from the offer of screening, through the screening pathway from initial screen to treatment. This allows monitoring of the standards at a service level described in the national guidance and service specification.

Quality assurance staff can run further reports from NBSS on performance at both service and individual level. For example, Film Reading Quality Assurance (FRQA) provides individual data on mammogram reader performance. This includes positive predictive value of referral (PPV), cancer detection rates, recall rates & discrepant cancers (missed by one of two readers).

Since 1996, the NHSBSP has worked with the Association of Breast Surgeons to conduct annual audits of screen detected breast cancers. These have proved pivotal in driving up standards of performance in both radiology and surgery. Successive audits have demonstrated improvements in non-operative diagnosis, axillary lymph node management and surgical techniques appropriate to screening^{8, 9}.

Ideally, the performance of the breast screening programme and each service in it, would be measured by their achievement in reducing breast cancer mortality. In practice it is not

possible to measure this and our best proxy measure for a reduction in breast cancer

mortality is interval cancer rates. Interval cancers are those invasive cancers presenting between screens from negative screening tests and within a 3-year screening interval. They are a slow, but reliable indicator of screening performance nationally, regionally and can be informative for some larger screening services.

Detailed information about both interval and screen-detected cancers are collected from services by staff working in the Screening QA service. The screening history of women that develop breast cancer is linked to the cancer registry data held by PHE. Recently, linking these two datasets has been automated and collected on the Screening History Information Management System (SHIM).

Current practice – The National Picture

Routine monitoring

Quality assurance staff supported by clinical advisors closely monitor service metrics, share findings at regular meetings with services and their commissioners, provide networking activities where services can see how their performance compares to other services and national comparators. Peer review visits are carried out approximately every three years where the service's quality can be closely scrutinised and recommendations made to promote quality improvement in the service. These visits are led by the quality assurance staff supported by clinical advisors. Quality assurance staff will intervene where routine monitoring generates concerns about the quality of the service.

Epidemiological QA

As well as routine monitoring of performance metrics, less frequent and more definitive epidemiological quality assurance is also carried out, which allows both for confounding factors such as age and also statistical stability. In general, standards are about ensuring individual service performance reaches an appropriate level, whereas epidemiological quality assurance is focussed on the performance of the whole programme. Epidemiological quality assurance can at times be applied at a service level to look for outliers. Figs 1 & 2 show examples of epidemiological QA graphics where age has been adjusted for and a reasonable level of statistical stability has been ensured by measures not based on numbers that are too small. Note that in fig 1 the recall rate is highly stable, being based on very large numbers, but the invasive cancer less so and would still have to be considered further. Fig 2 shows an example of a service with a high false positive biopsy rate that became an outlier in the screening year 2011/12 before reducing these rates to become more in line with the national average.

Sensitivity and specificity of the test and assessment

Screening is a two-stage process; firstly, the screening test (mammography) decides if women are at low risk (return to routine recall) or at high risk (refer for assessment) of breast cancer. The screening test therefore will have both false positive and false negative results.

In contrast, once a woman is referred to assessment the proportion of invasive cancers is somewhere between; 1 in 10 to 1 in 6, therefore a false negative at assessment will be of greater concern.

In determining overall performance, we are interested in both sensitivity and specificity, but most importantly the trade-off between sensitivity and specificity. In theory, we could estimate invasive cancer sensitivity, or total cancer sensitivity. For invasive cancer sensitivity, we are interested in estimating the total proportion of invasive cancers present at the screening test that were correctly referred to assessment.

For invasive cancer specificity, we are interested in the total proportion of women without invasive cancer who were correctly returned to routine recall. However, as we are monitoring performance in real time, we can only give an estimate of both sensitivity & specificity, as the status of woman as true negative or positive will only be confirmed some years later when women with interval cancers, or next round screen-detected cancers have been reviewed and can be taken account of in calculations.

Therefore, in practice, as both sensitivity and specificity are difficult to estimate we rely on cancer detection rate and recall rate.

Recall rates are an important measure of screening programme performance and are associated closely with specificity. Recall rate standards have an achievable threshold of less than 7% and an acceptable level of less than 10% at prevalent (first) screens and less than 5% and 7% respectively at incident screens. To maximise benefit and minimise harm it is important that the cancer detection rates are achieved at acceptable recall rates. Services with high recall rates are identified by routine quality assurance monitoring and efforts made to reduce the rate. This may not be easy as high recall rates can be driven by an anxiety within a service of missing cancers, or as a response to a low cancer detection rate. Quality assurance staff will need to understand the reason for the high rate and may need to provide appropriate support over a considerable time to bring about a reduction.

Invasive cancer detection rates

We can look at a screening service's performance at both an absolute and relative level. The absolute and relative levels are well illustrated using the standardised detection ratio (SDR) measure for invasive cancer detection¹⁰. This is an age standardised measure based on achieving an equivalent invasive cancer detection rate to the Swedish-Two County Study randomised controlled trial STC RCT² and therefore potentially an equivalent mortality reduction. We make the assumption that as the STC RCT achieved a 31% mortality reduction with a 90% uptake that the national programme will achieve about a 25% mortality reduction with a 70% uptake. This is an absolute measure of screening performance, which was arguably most **valid** in the earlier years of the screening programme prior to two-view mammography and digital technology. When first developed the SDR measure showed that the invasive cancer detection rate from most services was too low in the early years of screening. This was thought to be due to the greater experience of the Swedish film readers in the trial compared to the readers in the newly developing national programme which detected a higher proportion of cancers as DCIS.

As a result of the introduction of two view mammography, double reading and intervention by quality assurance staff to services with low detection rate, the invasive cancer detection rate increased by about 30% by 2003 from the level being achieved in 1994¹¹. This illustrates the importance of high quality performance measurement being available right from the start of the implementation phase (roll-out) of a new screening programme when major issues are most likely to occur.

Whilst invasive cancer detection rates have been well in excess of the STC RCT derived target for many years, with most services from 2003 onwards performing well above the achievable standards, now concern is more about assessment performance and screening specificity with failure of some services to achieve acceptable levels of recall.

Monitoring Interval cancers

Performance monitoring using interval cancer statistics is always retrospective and takes some time for complete ascertainment. This limits its use to pick up issues in real time, but it is useful in monitoring trends in national programme performance. In fact, it was the relatively high interval cancer rates at regional level in the early 1990s that alerted the national programme to the relatively poor detection of invasive cancers ^{12, 13}. From around 2003 with the introduction of two view mammography, the screening sensitivity in the national screening programme has improved considerably and is now considered as being very good. The effect of the increased cancer detection rate on reducing interval cancers has been demonstrated in recent years ¹⁴.

Interval cancers continue to be monitored at both national and service level and on occasions have flagged up possible quality issues in assessment practice in a service. However, small numbers and difficulty in achieving consistency across services in how they review interval cancers, means that interpreting interval cancer statistics at a service level is fraught with difficulty. If a service does appear to be an outlier, much more detailed investigations are needed to determine whether or not outlier status represents true quality issues.

Impact of clinical practice on interpretation of metrics

Given, that services and individuals are being asked to achieve improvements in both sensitivity and specificity, but that there is a recognised trade-off between these two measures, care is needed giving recommendations to services and individuals based on these metrics. Encouraging a service to improve its CDR can result in unacceptably high recall and biopsy rate, conversely, asking a service to reduce its recall rate may lead to a drop in CDR.

As film readers work in teams for double reading, it is difficult to determine exactly the skill of an individual reader whose data only reflects one part of the process. However, particularly poor performance (low cancer detection rate and high recall rate) or good performance (high cancer detection rate with low recall rate) from the first (unbiased) read is highly suggestive of film reading skill.

Services deploy many different approaches to double reading and arbitration. Some services will double read or arbitrate blind, others will see what the first reader has recorded, other services will consensus read all images recalled to assessment.

Services will also employ different strategies at times to ameliorate high or low recall rates. This may involve pairing individuals with high recall rates and very high sensitivity with individuals with very high specificity and lower recall rates but possibly with lower screening sensitivity.

All these factors can impact on both service and individual performance metrics, understanding the setting and protocols operating within a service are important in correctly interpreting and advising on radiological performance metrics

Where are we going? Performance metrics for an individual practitioner

After years of monitoring performance at a service level where individual performance was masked, the ability to analyse individual practitioner level data can provide a greater insight into practice. Individual statistics now give the potential for even greater improvements in performance at the local and national levels. Comparative statistics should be a powerful steer to drive service improvement. Detailed analysis of the individual components of screening will provide a greater understanding of what makes a high-quality service

excellent. This learning can be disseminated and should result in the optimisation of national programme performance.

In order to understand performance at an individual level, data relating to clinical practice needs to be attributed to the correct clinician. In assessment practice in the past, this has proved difficult, as often it was unclear from NBSS entries which clinician was taking responsibility for the woman's care. As a result, new guidance has been developed that requires services to identify a responsible assessor on NBSS for each woman assessed¹⁵.

In the future, this will provide more information on performance by linking radiological findings to and may enable analysis of outcomes by an individual clinician, if challenges from for example statistical stability from smaller numbers and case mix can be overcome.

Breast Screening Information System – an opportunity to compare individual and national performance

Professionals working within the programme should have easy access to reliable data and intelligence on their performance. To address this Public Health England has developed a new data sharing platform BSIS (Breast Screening Information System). BSIS is available to all breast screening services in England. Users will be able to access data on all aspects of the breast screening pathway to facilitate a comprehensive review of their programme. The system will facilitate the standardisation of data analysis and statistical methods employed within the programme. The data provided will primarily be displayed in the form of funnel plots to ease interpretation. Users will be able to export their data to support local use such as inclusion in their appraisal record. Dedicated resources are being produced in partnership with the professional bodies to support users in interpreting their data and providing suggestions on how to act on their results.

Assessment data – new analysis of Wide Bore Needle rates

Serious incidents occurring within the screening programme are carefully managed and a critical part of the process is to learn from such events. Some of the most significant incidents to affect the breast screening programme in the past have been in the area of assessment. This is where women have been recalled following a double read of their screening mammogram to a dedicated clinic where they may undergo further imaging, ultrasound, clinical examination and/or biopsy.

Performance in assessment is notoriously difficult to determine. Interval cancer reviews or rates are poor indicators. A new measure is being explored to provide some insight into practice that could act as a flag for further investigation. Data from the **annual statistical bulletin** enables analysis of the rate of biopsy based on either the number of women screened or assessed. PPV of referral to assessment and conversion rates of biopsy to cancer detection can also be produced for comparison.

Further analysis may be possible using data on the localisation method employed for each biopsy. Additionally, data on the number of visits to assessment and the positive predictive value of ultrasound and needle biopsy of the axilla can be extracted from the annual audit of screen-detected cancers undertaken by the programme in partnership with the ABS.

Opportunities to improve cancer detection rate

The detection of cancer is in many ways the conclusion of a pathway for screen positive women. A service's failure to achieve a high detection rate is the summation of a series of processes carried out by a number of practitioners. Improving rates requires an ability to optimise the quality of each step. Going forwards, data on each aspect of the pathway will be reviewed. To measure radiological performance individual data on recall rates, PPV of recall, discrepant cancers and individual cancer detection will be produced. Users will be encouraged to initially look at the performance of their service against the national mean and then look at their own personal performance and practice.

Examples of individual reader data expressed as funnel plots are shown in figs 3 & 4. In figure 3 we can see that the service is performing slightly better than the nation with control limits above that achieved by the country as a whole. However there are two readers which demonstrate a lower PPV when their performance is compared nationally. Reader J is a low outlier for PPV for recall to assessment. In this case, the recall rate and discrepant cancer rate should also be considered. A low PPV may be because they are sending more cases than their colleagues to assessment but detecting the same number of cancers or that they are not recalling sufficient number of cases. This performance could be because they are newly qualified or inexperienced. However, we can note that this reader is reading a relatively large number of first read images (~7,000) so there may be issues with regards excessive workload. In other cases a low or average PPV may be associated with a high discrepant cancer rate as illustrated in figure 4 with reader E. Action in this scenario may include the Director of Breast Screening and the individual reader reviewing the images that were not identified as cancer, identifying common themes such as missing irregular or spiculate masses and offering targeted training and support. It may be necessary to pool several years to get sufficient cases to inform these discussions. Reader I in this graph is achieving a high PPV and low discordant rate. Opportunities to learn from this reader within the service should be maximised with possibilities such as them acting as a mentor explored.

Having this granularity of data will provide teams with the information to optimise image reading practice. For example, by avoiding double reading by pairs where both individuals have high recall rates. Pairing these individuals with others who have high levels of screening sensitivity and specificity can improve overall service performance and individuals can learn from detailed internal audit where this has occurred.

Opportunities and challenges

Cancer detection rate and non-invasive disease

The challenge of any cancer screening programme is that cancer exists in the population as a continuum of disease and that any screening programme only wants to detect and remove a lesion that could become life-threatening for that individual person and leave all other lesions alone. This is the ultimate goal of any screening programme, such that benefit is fully maximised and harms fully minimised.

An unexpected difficulty at the outset of the national screening programme was the high amount of non-invasive cancers (mostly ductal carcinoma in-situ (DCIS)) detected. The STC RCT detected about 8% DCIS, but the national programme in England detects over 20% DCIS¹¹. **Of what** Given there is no general agreement on whether the programme should only detect high grade, high and intermediate grade only or all grades of DCIS, this creates a particular difficulty for the programme defining good performance in terms of non-invasive cancer detection. is

To further develop performance measures, we may need to look more closely at grade (aggressiveness) of cancer, rather than just whether a cancer is invasive or non-invasive. We could, for example, redefine detection rate in terms of a high sensitivity for invasive cancers and high-risk DCIS.

Further work would be needed to produce a similar stable measure to SDR which is based on invasive cancers only. As there are fewer non-invasive cancers (mostly DCIS) detection rates would be prone to substantial statistical instability

However, until we have further evidence from studies such as the Sloane project ¹⁶ that can shed light on the natural history of this condition, the programme will not be able to develop any standards or metrics to shape this aspect of programme performance.

Responding to new data sets with individual level data

The transition from service level data to individual clinician statistics with national comparative data requires careful thought and implementation. Historically the programme has run schemes that used test data to assess individual competence such as PERFORMs¹⁷ in radiology and EQA in pathology. The analysis of real-life data provides a new perspective, offering the individual insight into their own practice. As systems such as BSIS roll out, practitioners will need support to interpret the findings and access advice. Quality assurance staff and clinical advisors need to be on hand for directors of breast screening and individuals interpreting the first data sets as they become available.

National comparative performance in breast screening radiology is a new area and while for some these comparative statistics may endorse their opinion of their own standard of image reading. For others, comparison may prove less palatable and will be a surprise finding.

The challenge will be how to mentor and support individuals whose performance may be towards the bottom quintile or who appear to be outliers, and make sure they receive adequate support and re-training if needed.

Transparency and Duty of candour

There has been a concerted effort within the NHS to increase levels of transparency. This takes a number of forms such as; sharing information about performance, providing users with information about the benefits and harms of a procedure and being open and honest when things go wrong. The latter is a requirement that is captured in the duty of candour legislation¹⁸.

Specialities as diverse as bariatric surgery and interventional cardiology now publish relevant data at an individual consultant level. In breast screening, consultant outcome data

measured as the outcome for an individual woman does not provide a complete picture. If there is a desire to move to consultant outcomes in breast screening radiology, measures will need to reflect important outcomes for the population screened and the ability of the radiologist to be sensitive *and* specific in the detection of breast cancer.

Being open and transparent should be no different in the breast screening programmes to other NHS care. However, inherent in screening is the recognition that false negatives and false positives will always occur. This feature of screening is often not appreciated by those working outside of screening or the public. Breast screening clinicians may become increasingly reluctant to engage in collecting more information that focusses on individual practice if this could expose them to repeated requests, albeit unwarranted, to subject all cases of interval cancer to duty of candour regulations.

The quality of the NHSBSP has improved over time because of the commitment of the clinicians working in the programme to support data collection, to look critically at their own performance and learn and change their practice. It is imperative that a drive for transparency and duty of candour does not inhibit our ability to engage clinicians in providing individual level performance data which can be such a powerful tool in improving quality.

Conclusion

Performance metrics have been pivotal in making the NHSBSP a world class breast screening programme. Embracing new metrics at an individual level will ensure that the programme continues to improve and provide a high-quality service for women in England.

References

- 1 Breast cancer screening: report to the health ministers of England Wales, Scotland and Northern Ireland. Sir Patrick Forrest: London HMSO 1986.
- 2 Tabar L, Fagerberg G, Duffy S, Day N, Gad A, Grontoft O. Update of the Swedish-Two County programme of mammographic screening for breast cancer. *Radiol Clin North Am.* 1992;30:187-210.
- 3 IARC Handbook of Cancer Prevention: Breast Cancer Screening. 2002 IARC Press
- 4 Donabedian A. An Introduction to Quality Assurance in Health Care, USA: Oxford University Press, 2003: 60 - 64
- 5 Public Health England. NHSBSP Publication 59: Quality assurance guidelines for breast cancer screening radiology 2011: <https://www.gov.uk/government/publications/breast-screening-quality-assurance-standards-in-radiology>
- 6 Public Health England. Breast screening: consolidated programme standards. PHE publications gateway number: 2016720, 15 August 2017: <https://www.gov.uk/government/publications/breast-screening-consolidated-programme-standards>
- 7 Information Standards Board for Health and Social Care, Breast Screening Programme Dataset (KC63 and KC62) Standard Specification Document ID ISB 1597 29/08/2013
- 8 Sauven, P., Bishop, H., Patnick, J., Walton, J., Wheeler, E. and Lawrence, G. (2003), The National Health Service Breast Screening Programme and British Association of Surgical Oncology audit of quality assurance in breast screening 1996–2001. *Br J Surg*, 90: 82–87. doi:10.1002/bjs.4013)
- 9 Public Health England: NHS breast screening programme and Association of Breast Surgery; An audit of screen detected breast cancers for the year April 2015 to March

2016. PHE publications gateway number: 2017043, June 2017:

<http://associationofbreastsurgery.org.uk/audit/nhs-breast-screening-programme-audit/>

10 Blanks RG, Day NE, Moss SM. Monitoring the performance of breast screening programmes: use of indirect standardisation in evaluating the invasive cancer detection rate. J Med Screening 1996; 3:79-81

11 Gray JAM, Patnick J, Blanks RG. Maximising benefit and minimising harm of screening BMJ 2008;Mar1 336(7642) 480-83.

12 Woodman CBJ, Threlfall A, Boggis CRM, Prior P. Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's north western region. BMJ 1995, 310-224-6,

13 Day N, McCann J, Camilleri-Ferrante C, Britton P, Hurst G, Cush S et al. Monitoring interval cancers in breast screening programmes: the East Anglian experience. J Med Screening 1995; 2:180-5

14 Dibden A, Offman J, Parmer D et al. Reduction in interval cancer rates following the introduction of two-view mammography in the UK breast screening programme. Br J Cancer 2014 Feb 4 110(3) 56—4.

15 Public Health England: NHSBSP Publication No 49; Breast screening: clinical guidelines for screening assessment: Fourth Edition. PHE publications gateway number: 2016426, November 2016: <https://www.gov.uk/government/publications/breast-screening-clinical-guidelines-for-screening-management>

16 Thomas J, Evans A, Macartney J, Pinder S E, Hanby A, Ellis I, Kearins O, Roberts T, Clements K, Lawrence G, Bishop H on behalf of the Sloane Project Steering Group. Radiological and pathological size estimations of pure ductal carcinoma in situ of the breast, specimen handling and the influence on the success of breast conservation surgery:

a review of 2564 cases from the Sloane Project. [British Journal of Cancer](#) 2010; 102: 285-293

17 [Scott HJ](#), [Gale AG](#). Breast screening: PERFORMS identifies key mammographic training needs. [Br J Radiol](#). 2006 Dec;79 Spec No 2:S127-33.

18 Care Quality Commission; Regulation 20: 8 June 2017; <http://www.cqc.org.uk/guidance-providers/regulations-enforcement/regulation-20-duty-candour>

Fig 1 PPV-referral diagram by services for all years from 2009/10 to 2015/16 (80 services with 7 years data i.e. 560 data points) for invasive cancers with target recall of 5% shown on diagram. All services for all years are within the 7% minimum standard. Note that some services with higher recall rates can be shown to be technically outliers even though they are within the minimum standard.

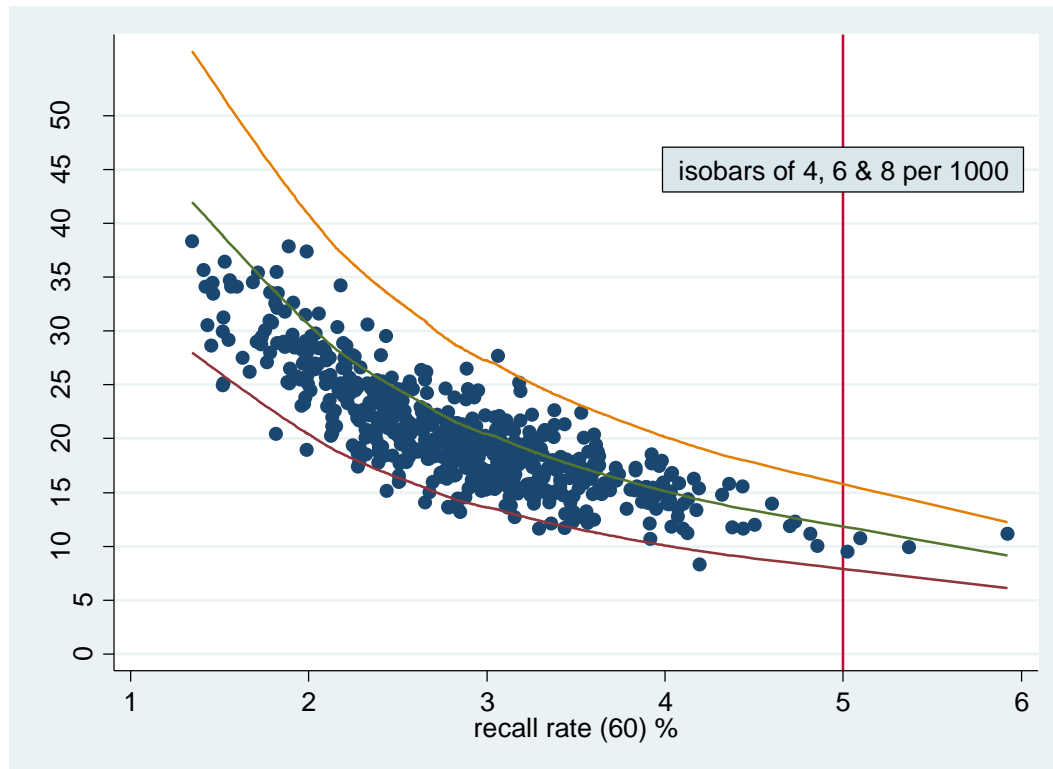


Fig 2 Graph of standardised false positive biopsy ratio by year labelled with screening service “BY” which was an outlier in the screening year 2011 (1st April 2011 to 31st March 2012) before reducing its false negative biopsy rates closer to the national average figure

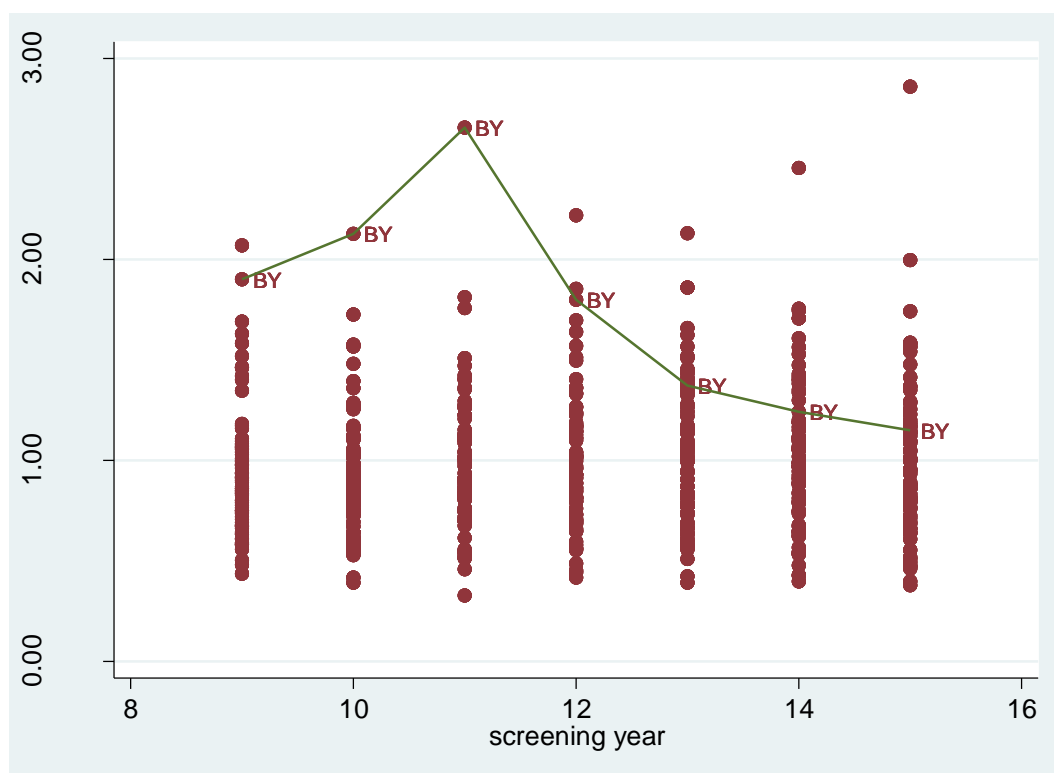


Fig 3 PPV of recall

Positive Predictive Value (PPV) is the percentage of cases the reader recommended be recalled to assessment which were diagnosed with cancer.

= (Number of women recommended be recalled to assessment and were diagnosed with cancer) x 100 / (Number of women recommended be recalled to assessment)

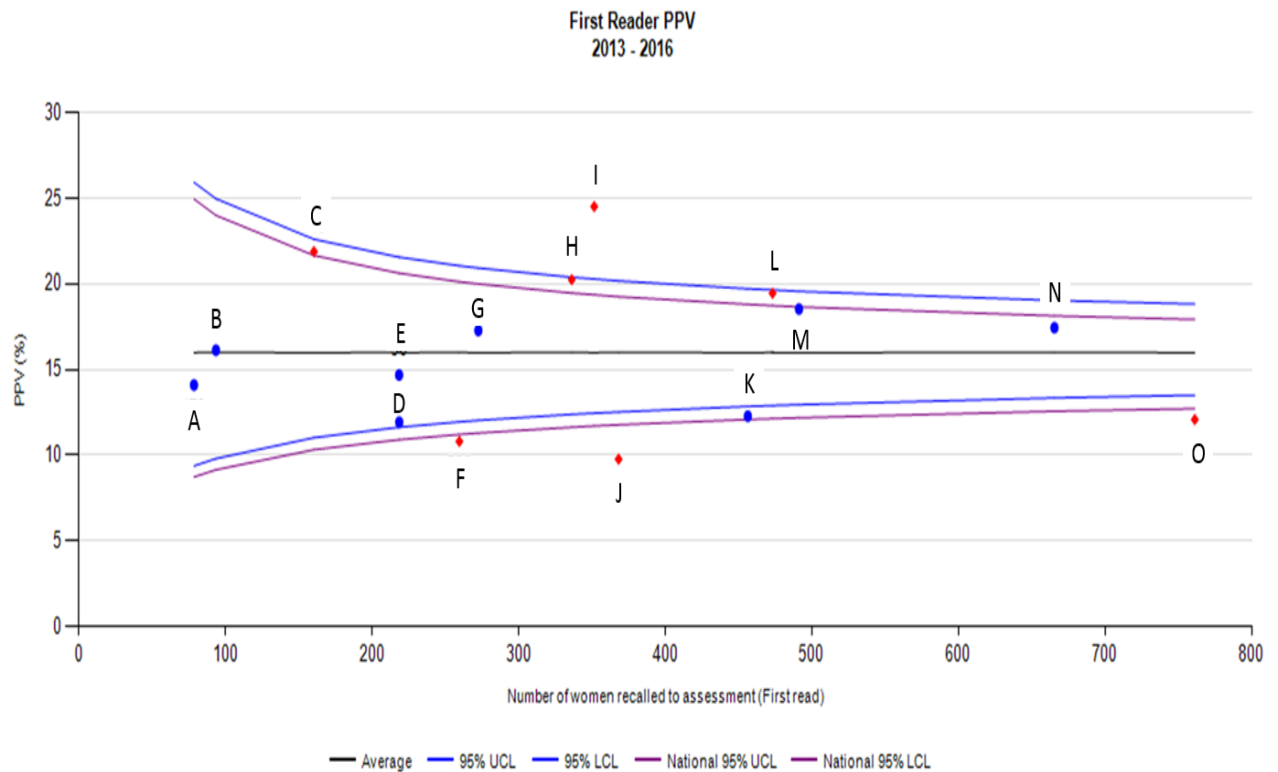


Fig 4 Discrepant rate

Discrepant Cancer Rate is the proportion of cases reported which the reader recommended be returned to routine recall and were recalled to assessment by another reader and diagnosed with cancer.

= (Number of women recommended be returned to routine recall and were diagnosed with cancer)
x 1,000 / (Number of cases reported)

