NAoPri
National Audit of Primary Breast Cancer

NATCAN
National Cancer Audit Collaborating Centre

National Audit of Primary Breast Cancer State of the Nation Report:

Methodology Supplement

An audit of care received by people diagnosed with primary breast cancer between 1 January 2020 and 31 December 2022 in England and Wales.

Published September 2025

Citation for this document:
National Audit of Primary Breast Cancer (NAoPri) State of the Nation Report 2025. London:
National Cancer Audit Collaborating Centre, Royal College of Surgeons of England, 2025.

This document was prepared by members of the NAoPri Project Team:
Kieran Horgan, Surgical Lead
David Dodwell, Clinical Oncology Lead
Mark Verrill, Medical Oncology Lead
Sarah Blacker, Clinical Research Fellow
Jemma Boyle, Senior Clinical Research Fellow
Liyang Wang, Clinical Research Fellow
Christine Delon, Data Scientist
Diana Withrow, Methodologist
David Cromwell, Senior Methodologist

With review and input from:
NAoPri Audit Advisory Committee (Clinical Reference Group)
NATCAN Executive Team

Acknowledgements: Special thanks to Jibby Medina for her many contributions as Programme Manager for the NAoPri, the National Audit of Metastatic Breast Cancer (NAoMe) and the National Audit of Breast Cancer in Older People (NABCOP) from April 2016 until May 2025.

The Royal College of Surgeons of England is an independent professional body committed to enabling surgeons to achieve and maintain the highest standards of surgical practice and patient care. As part of this it supports audit and the evaluation of clinical effectiveness for surgery. Registered Charity no: 212808.

The National Cancer Audit Collaborating Centre (NATCAN) is commissioned by the Healthcare Quality Improvement Partnership (HQIP) and funded by NHS England and Welsh Government as part of the National Clinical Audit and Patient Outcomes Programme (NCAPOP). NATCAN delivers national audits in, breast (primary and metastatic), kidney, lung, non-Hodgkin lymphoma, oesophago-gastric, ovarian, pancreatic and prostate cancers.

The Association of Breast Surgery is a registered charity dedicated to advancing the practice of breast surgery and the management of breast conditions for the benefit of the public. It is a multi-professional membership association, which promotes training, education, clinical trials and guideline composition and adoption. For further information, please refer to the website www.associationofbreastsurgery.org.uk. Registered charity no: 1135699

The UK Breast Cancer Group (UKBCG) is a forum for Clinical and Medical Oncologists. The UKBCG acts as a stakeholder to NICE, NHS England and other organisations; and undertakes key pieces of work, at times in collaboration with other bodies, with the overriding endpoint of improving patient care.
The Group's objectives include advancing the education of clinical and medical oncologists in the subject of breast cancer, concerning its identification, diagnosis and treatment; promoting research for the public benefit in all aspects of breast cancer and publishing the results; and assisting in the treatment and care of persons suffering from breast cancer, or in need of rehabilitation, by the provision of education for healthcare professionals.
Further information on the work of the UKBCG is communicated via this website on a regular basis https://ukbcg.org/. Registered charity no: 1177296

This work uses data that have been provided by patients and collected by the NHS as part of their care and support. For patients diagnosed in England, the data are collated, maintained and quality assured by the National Disease Registration Service (NDRS), which is part of NHS England. Access to the data was facilitated by the NHS England Data Access Request Service.

NHS Wales is implementing a new cancer informatics system. As a result, the quality and completeness of data from Wales is likely to have been impacted due to implementation of this new system across multiple NHS organisations (Health Boards), which has resulted in data being supplied by both old and new systems. Additionally, and reflecting the uncertainty of data quality, the data submitted to the audit may not have undergone routine clinical validation prior to submission to the Wales Cancer Network (WCN), Public Health Wales.

Contents

1. Introduction

This document provides supporting material to the 2025 State of the Nation (SotN) Report for the National Audit of Primary Breast Cancer (NAoPri) and its data tables and data viewer. The document describes the data used in the report with details on sources of data, criteria for inclusion and how data completeness, patient characteristics and performance indicators are derived and reported.

2. Sources of Data

The audit uses information from routine national healthcare datasets in England and Wales. These datasets capture details on the diagnosis, management, treatment and outcome of every patient newly diagnosed with cancer in the NHS in England and Wales.

For England, the audit received information from the National Disease Registration Service (NDRS) at a tumour level for this State of the Nation report. The information held in the National Cancer Registration Data (NCRD) is compiled from a variety of sources including the Cancer Outcomes and Services Dataset (COSD), Hospital Episode Statistics admitted patient care (HES APC) records, the Systemic Anti-Cancer Therapy dataset (SACT), the Radiotherapy dataset (RTDS) and data submitted by pathology laboratories. The audit also received linked information from COSD (linked at tumour level), HES APC, HES Outpatients data (HES OP), SACT,RTDS and the Primary Care Prescription Database (PCPD) (all linked at patient level). **Appendix 1** provides more detail on the data sources listed below and the information they contain.

The English data received by the National Cancer Audit Collaborating Centre (NATCAN) included data on patients registered with cancer from 01 January 2015 up to 31 December 2022.

As with cancer registries in other countries, cancer registrations in England can take up to 5 years after the end of a given calendar year to approach 100% completeness and stability. NDRS uses an active system of gathering information on cancer diagnoses from multiple sources across the patient pathway. Completeness varies by tumour type because different patient pathways provide different opportunities for data flows into NDRS. The 'Gold standard' cancer registration dataset that is used in cancer statistics bulletins and available for analysis outside of NDRS contains over 98% of all the people that will eventually be found by the registration process, and the completeness for a calendar year of data increases over time. More information about the cancer registration process can be found here.

For Wales, the audit was provided with a registration dataset at patient level for patients diagnosed with cancer from 01 January 2015 up to 31 December 2022. Welsh cancer registration data is captured through a national system, Cancer Information System for Wales (CaNISC) and the new Welsh Clinical Portal. The audit also received linked datasets of records from the Patient Episode Database for Wales (PEDW) containing information on inpatient and day case activity, and mortality data from the Office for National Statistics (ONS). Data from England and Wales were harmonised and analysed together.

3. Inclusion and Exclusion Criteria

The data submitted by NDRS and WCN is checked and filtered for eligible participants, tables 3.1 and 3.2 explain the process in defining the final cohort to be used in the audit.

People were included for analysis within the SotN Report based on the following inclusion and exclusion criteria:

Table 3.1: Audit Inclusion Criteria

Inclusion Criteria	**Details**
Type of cancer	ICD-10 diagnosis codes: C50 (invasive breast cancer); D05 (in-situ breast cancer).
Adults	Age >=18
Valid Diagnosis Date	01 January 2020 to 31 December 2022
Cancer Stage	Stages 0 to 3C breast cancer or with an "unknown" staging.

Table 3.2: Audit Exclusion Criteria	
Exclusion Criteria	**Details**
Metastatic disease recorded	ICD-10 diagnosis recording secondary cancer within hospital admissions data within 6 months of date of diagnosis.
Histology	Disease not invasive or DCIS (e.g., lobular carcinoma in situ)
Reported by death certificate only or Date of diagnosis corresponds to date of death	**For English data:** Using NCRD: basisofdiagnosis = 0 (Death certificate) and/or dco = Y (tumour registered from a death certificate only) and/or date_diagnosis=ONS_date_death **For Welsh data:** Date_diagnosis = ONS_date_death and/or basisofdiagnosis=0
Diagnosed outside of an eligible NHS organisation in England or Wales.	Diagnosed outside of an NHS organisation in England or Wales OR Place of diagnosis not provided OR Diagnosed at an NHS organisation with no active breast unit OR Diagnosed within an NHS organisation with less than 30 allocated registrations of breast cancer per year OR Diagnosed at a tertiary centre
Multiple cancer registrations during the audit period	>1 tumour ID in the cancer registration dataset (e.g., bilateral cancer)

4. Key Data Items

Details of the variables and datasets used to compile the data completeness information are shown below in Table 4.1.

Table 4.1: Data Completeness Variables				
Data Item	**Source**			
	England		**Wales**	
	Data field	**Dataset(s)**	**Data field**	**Dataset**
Performance status[A]	CR0510	COSD	PerformanceStatus	CaNISC
Clinical Nurse Specialist	CR2050	COSD	HasSeenClinicalNurseSpecialist	CaNISC
Triple Diagnostic Assessment	BR4400 (not provided)	COSD	BreastTripleDiagnosticAssessment	CaNISC
ER status[B]	ER_STATUS; ER_SCORE / pBR4220	NCRD /COSD	ErStatus	CaNISC
HER2 status	HER2_STATUS; pBR4280	NCRD/COSD	Her2Status; Her2FishStatus	CaNISC
PR status	PR_STATUS; PR_SCORE / pBR4290	NCRD/COSD	PrStatus	CaNISC
Stage (overall)	STAGE_BEST, T-stage, N-stage, M-stage	NCRD	StageGroupIntegrated	CaNISC
T-stage	T_BEST	NCRD	FirstTStagePathological	CaNISC
N-stage	N_BEST	NCRD	FirstNumberOfNodesExamined FirstNumberOfNodesPositive	CaNISC
Grade	GRADE	NCRD	FirstGradeOfDifferentiationPathology	CaNISC

[A] See Appendix 2. [B] The percentage reported for data completeness reflects the data quality as received by the NAoPri, without augmentation with data for endocrine therapy prescription, to highlight the need for improved data quality. When oestrogen receptor status was used for risk adjustment or for subgroup analyses, this was augmented with data from Primary Care Prescription Database, where persons receiving endocrine therapy were assumed to be ER-positive.

Details of the variables and datasets used to compile the patient and tumour characteristics are shown below in Table 4.2.

Table 4.2: Patient and Tumour Characteristics Variables

Data Item	Source			
	England		Wales	
	Data field	Dataset	Data field	Dataset
Year of Diagnosis	DIAGNOSISDATEBEST / CR2030	NCRD/COSD	DateOfPrimaryDiagnosis	CaNISC
Age at Diagnosis	Birthmonth; birthyear / CR0100	NCRD/COSD	AgeAtDiagnosis	CaNISC
Sex	SEX	NCRD	Gender	CaNISC
Screened	SCREEN_DETECTED / CR1600	NCRD/COSD	CancerReferralSource	CaNISC
Ethnic Group	ETHNICITY / CR0150	NCRD/COSD	N/A	CaNISC
Index of Multiple Deprivation Quintile	QUINTILE_2019 / CR0080	NCRD/COSD	LSOA_ Deprivation Quintile Group	CaNISC
SCARF index[A]	DIAG_nn	HES APC	Diagnosis_n	PEDW
Charlson co-morbidity score	DIAG_nn	HES APC	Diagnosis_n	PEDW
Performance status[B]	CR0510	COSD	PerformanceStatus	CaNISC
Stage (overall)	STAGE_BEST, M_Best	NCRD	StageGroupIntegrated	CaNISC
T-stage	T_BEST	NCRD	FirstTStagePathological	CaNISC
N-stage	N_BEST	NCRD	FirstNumberOfNodesExamined, FirstNumberOfNodesPositive	CaNISC
Grade	GRADE	NCRD	FirstGradeOfDifferentiationPathology	CaNISC
Histology	HISTOLOGY_CODED	NCRD	DiagnosisICD10	CaNISC
Disease Group	Derived from stage, grade and histology above	NCRD	Derived from stage, grade and histology above	CaNISC
Date of Death	Vitalstatus, vitalstatusdate, deathdatebest	NCRS/ONS	DeathDate	CaNISC/ONS

[A] See Appendix 3. [B] See Appendix 2.

5. Indicator Definitions

The audit uses key indicators to monitor progress against its healthcare improvement goals. These indicators align with national guidelines and standards. Definitions of how the indicators included in the SotN report were derived from data for England and Wales are described below.

Some indicators are focused on subgroups of patients as defined by route to diagnosis, stage, histopathology, or treatment, as these factors are important determinants of whether particular treatments or processes are suitable for patients.

Performance Indicator 1: Triple Diagnostic Assessment (TDA) in a single hospital visit

Percentage of patients who underwent Triple Diagnostic Assessment (TDA) (clinical examination, imaging and biopsy) within a single clinic visit.

For England, there is a bespoke TDA data item (BR4400) which was introduced in NDRS data in 2022. However, as this report covers patients diagnosed between 2020 and 2022, information for this data item was not provided by NDRS. For England, TDA was therefore estimated from the date of first contact (from the Cancer Waiting Times dataset) and the date of the biopsy sample (from COSD). Because most biopsies are performed under image guidance, we assumed that if the date of first contact matches the biopsy date, the patient had undergone TDA. While reliable and timely imaging data would have further increased the validity of this indicator, such data were not available at the time. In future, we hope to use the bespoke TDA data item for England. In Wales, the bespoke data item for TDA was used.

Table 5.1: Percentage of patients who underwent Triple Diagnostic Assessment (TDA) in a single hospital visit		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of people who undergo Triple Diagnostic Assessment (TDA) in a single hospital visit.	Number of patients who have a biopsy date (COSD_date_sampletaken1, COSD_date_samplereceipt or COSD_date_samplereceipt-1) in COSD data which matches the date first seen (date_first_seen) in Cancer Waiting Times data.	Number of patients who have TDA item completed as "yes".
Denominator: Number of people diagnosed with primary breast cancer outside of screening.	Final cohort as described in patient inclusion / exclusion, further excluding those patients detected via screening.	Final cohort as described in patient inclusion / exclusion but excluding those patients detected via screening.
Construction notes:	Whether a patient originated from screening is indicated from two sources. We use the routes to diagnosis data item which is derived by NDRS using multiple sources. The screening status is also determined by NDRS using data supplied to it by the NHS Breast Screening Programme. This data item is seen as the most reliable source of information on whether a tumour was detected through routine screening	-
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Gender, age-group and diagnosis year.	Gender, age-group and diagnosis year.
Risk adjusted:	No – not required as should be near 100% regardless of case-mix.	
Outlier reporting:	No	

Performance Indicator 2: Contact with a Clinical Nurse Specialist (CNS)

Percentage of patients who had contact with a Clinical Nurse Specialist (CNS) recorded after diagnosis.

For England and Wales, contact with a CNS is captured in bespoke data items from COSD and CaNISC, respectively.

Table 5.2: Percentage of patients who had contact with a Clinical Nurse Specialist (CNS) recorded as 'yes' within 30 days of the date of diagnosis.

	England	Wales
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Percentage of patients who had contact with a Clinical Nurse Specialist (CNS) recorded as 'yes' within 30 days of the date of diagnosis	Number of patients who have COSD data item CR2050 (*Clinical Nurse Specialist Indication Code)* recorded as yes.	Number of patients who have CaNISC data item HasSeenClinicalNurseSpecialist recorded as yes.
Denominator: Number of people diagnosed with primary breast cancer.	Final cohort as described in patient inclusion / exclusion.	Final cohort as described in patient inclusion / exclusion.
Construction notes:		
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Gender, age-group and diagnosis year.	Gender, age-group and diagnosis year.
Risk adjusted:	No – not required as we expect this to be yes, regardless of case-mix.	
Outlier reporting:	No	

Performance Indicator 3: Breast-conserving surgery or mastectomy within 12 months of diagnosis.

Percentage of patients who have surgery (breast-conserving or mastectomy) within 12 months of diagnosis.

Receipt of surgery is derived from the Cancer Registry Treatment dataset and HES data for England and CaNISC and PEDW data for Wales. The type of surgery is established using OPCS codes (See **Appendix 4**).

Table 5.3: Percentage of patients who had i) breast-conserving surgery or ii) mastectomy within 12 months of diagnosis.		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of people who had i) breast-conserving surgery or ii) mastectomy within 12 months of diagnosis.	Number of patients who have surgery (in HES or Cancer Registry Treatment Dataset) recorded up to 90 days before date of diagnosis and up to 12 months following the date of diagnosis.	Number of patients who have surgery recorded (in CaNISC) up to 90 days before date of diagnosis and up to 12 months following the date of diagnosis.
Denominator: Number of people diagnosed with early breast cancer (stage 0 to stage 3A or "unknown" staging).	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer.
Construction notes:	The first recorded breast surgery is used for this indicator.	The first recorded breast surgery is used for this indicator.
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Gender, age-group, diagnosis year, stage and hormone receptor status.	Gender, age-group, diagnosis year, stage and hormone receptor status.
Risk adjusted:	Yes - age (spline), tumour grade, Charlson co-morbidity score[A], SCARF index[B], invasive cancers only: ER status T-stage, N-stage, and diagnosis year.	
Outlier reporting:	No	

[A] See Appendix 5 [B] See Appendix 3

Performance Indicator 4: Neo-adjuvant chemotherapy.

Percentage of patients who receive neo-adjuvant chemotherapy for triple negative or HER2-positive stage 2 to 3A breast cancer.

Information regarding the use of neo-adjuvant chemotherapy is derived from Systemic Anti-Cancer Therapy and HES data for England and CaNISC data for Wales.

Table 5.4: Percentage of patients who received neo-adjuvant chemotherapy.		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of people who receive neo-adjuvant chemotherapy.	Number of people who receive neo-adjuvant chemotherapy (chemotherapy before the date of surgery) using the SACT dataset.	Number of people who receive neo-adjuvant chemotherapy (chemotherapy before the date of surgery) using the CaNISC dataset.
Denominator: Number of people with early invasive breast cancer (stage 2 to stage 3A), who are triple negative or HER2 positive who have had surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with stage 2 to 3A, with triple negative or HER2-positive disease and patients who have had surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with stage 2 to 3A, with triple negative or HER2-positive disease and patients who have had surgery within 12 months of diagnosis.
Construction notes:	Date of chemotherapy is constructed from date of administration in SACT where chemotherapy is one of: cabazitaxel, capecitabine, carboplatin, cisplatin, cyclophosphamide, docetaxel, doxorubicin, epirubicin, eribulin, etoposide, fluorouracil, methotrexate, mitomycin, mitoxantrone, paclitaxel, vindesine, vinorelbine	Use of data item: DateOfChemotherapy
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Gender, age-group, diagnosis year, and hormone receptor status.	Gender, age-group, diagnosis year, and hormone receptor status.
Risk adjusted:	Yes - age (spline), T-stage, N-stage, tumour grade, Charlson co-morbidity score[A], SCARF index[B], ER status, HER2 status, adjuvant radiotherapy and diagnosis year.	
Outlier reporting:	No	

[A] See Appendix 5 [B] See Appendix 3

Performance Indicator 5: Adjuvant radiotherapy following breast-conserving surgery or mastectomy.

Percentage of patients who received adjuvant radiotherapy following breast-conserving surgery or mastectomy.

Information for this indicator is derived from RTDS and HES data for England and CaNISC and radiotherapy data for Wales.

Table 5.5: Percentage of patients who received adjuvant radiotherapy following i) breast-conserving surgery and ii) mastectomy.		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator:	The number of patients who are recorded as receiving any adjuvant	The number of patients who are recorded as receiving any adjuvant

Number of persons who received adjuvant radiotherapy following breast-conserving surgery or mastectomy	radiotherapy regimen within the 6-month period after surgery OR with no time limit after surgery if adjuvant chemotherapy was given.	radiotherapy regimen within the 6-month period after surgery OR with no time limit after surgery if adjuvant chemotherapy was given.
Denominator: Number of persons with early breast cancer (stage 0 to stage 3A or "unknown" staging) who have had surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing surgery within 12 months of diagnosis.
Construction notes:		
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Gender, age-group, diagnosis year, stage and hormone receptor status.	Gender, age-group, diagnosis year, stage and hormone receptor status.
Risk adjusted:	Yes - age (spline), tumour grade, Charlson co-morbidity score[A], SCARF index[B] and diagnosis year; invasive only: ER status, T-stage, N-stage.	
Outlier reporting:	No	

[A] See Appendix 5 [B] See Appendix 3

Performance Indicator 6: Chemotherapy.

Percentage of patients who received any chemotherapy (neo-adjuvant or adjuvant).

Information regarding the use of neo-adjuvant or adjuvant chemotherapy is derived from Systemic Anti-Cancer Therapy data for England and CaNISC data for Wales.

Table 5.6: Percentage of patients who received any chemotherapy (neo-adjuvant or adjuvant).		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of people who have chemotherapy.	Number of people recorded as having received any chemotherapy using the SACT dataset.	Number of people recorded as having received any chemotherapy using the CaNISC dataset.
Denominator: Number of people with early breast cancer (stage 1 to stage 3A or "unknown" staging) who have had surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early invasive breast cancer (stage 1-stage 3A including stage unknown) and undergoing surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early invasive breast cancer (stage 1-stage 3A including stage unknown) and undergoing surgery within 12 months of diagnosis.
Construction notes:	Date of chemotherapy is constructed from date of administration in SACT where chemotherapy is one of: cabazitaxel, capecitabine, carboplatin, cisplatin, cyclophosphamide, docetaxel, doxorubicin, epirubicin, eribulin, etoposide, fluorouracil, methotrexate, mitomycin, mitoxantrone, paclitaxel, vindesine, vinorelbine	Use of data item: DateOfChemotherapy
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards

Subgroup Reporting:	Gender, age-group, diagnosis year, stage and hormone receptor status.	Gender, age-group, diagnosis year, stage and hormone receptor status.
Risk adjusted:	Yes - age (spline), T-stage, N-stage, tumour grade, Charlson co-morbidity score[A], SCARF index[B], ER status HER2 status and diagnosis year.	
Outlier reporting:	No	

[A] See Appendix 5 [B] See Appendix 3

Performance Indicator 7: Immediate reconstruction following a mastectomy.

Percentage of patients who have an immediate reconstruction at the same time as their mastectomy.

Receipt of immediate reconstruction is derived from the Cancer Registry Treatment dataset and HES data for England, and CaNISC and PEDW data for Wales. Immediate reconstruction is identified using relevant OPCS codes (see Appendix 4).

Table 5.7: Percentage of patients who had an immediate reconstruction following a mastectomy.		
	England	**Wales**
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of women who have mastectomy with immediate breast reconstruction.	The number of patients who are recorded as having an immediate breast reconstruction – defined as reconstruction happening on the same date and laterality (where available) as the mastectomy.	The number of patients who are recorded as having an immediate breast reconstruction – defined as reconstruction happening on the same date and laterality (where available) as the mastectomy.
Denominator: Number of women with early breast cancer (stage 0 to stage 3A or "unknown" staging) who have had a mastectomy within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing mastectomy within 12 months of diagnosis. Women only.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing mastectomy within 12 months of diagnosis. Women only.
Construction notes:	Identify occurrence of relevant OPCS codes[A] in HES episodes, with date of episode.	Identify occurrence of relevant OPCS codes[A] in HES episodes, with date of episode.
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Age-group, diagnosis year and stage.	Age-group, diagnosis year and stage.
Risk adjusted:	Yes - age (spline), T-stage, N-stage, tumour grade, Charlson co-morbidity score[B], SCARF index[C], ER status, diagnosis year and adjuvant radiotherapy use.	
Outlier reporting:	No	

[A] See Appendix 4 [B] See Appendix 5 [C] See Appendix 3

Performance Indicator 8: Re-operation within 12 months of their initial breast-conserving surgery.

Percentage of patients who have a re-operation within 12 months of their initial breast-conserving surgery.

Re-operations are derived from the Cancer Registry and Treatment Dataset and HES data for England, and CaNISC and PEDW data for Wales. Re-operation is identified using relevant OPCS codes and laterality (where available).

Table 5.8: Percentage of patients who had re-operation within 12 months of their initial breast-conserving surgery.

	England	Wales
Dates of diagnosis:	1/1/2020 to 31/12/2022	1/1/2020 to 31/12/2022
Numerator: Number of persons who have had a re-operation within 12 months of their initial breast-conserving surgery.	The number of persons who have had an ipsilateral re-operation (where laterality available) defined as re-excision, repeat breast-conserving surgery or completion mastectomy within 12 months of their initial breast-conserving surgery. Uses HES and the Cancer Registry Treatment Dataset.	The number of persons who have had an ipsilateral re-operation (where laterality available) defined as re-excision, repeat breast-conserving surgery or completion mastectomy within 12 months of their initial breast-conserving surgery.
Denominator: Number of persons with early breast cancer (stage 0 to stage 3A or "unknown" staging) who have had breast conserving surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing breast-conserving surgery as their first breast surgery within 12 months of diagnosis.	Final cohort as described in patient inclusion / exclusion. Restricted to those with early breast cancer (stage 0-stage 3A including stage unknown) and undergoing breast-conserving surgery as their first breast surgery within 12 months of diagnosis.
Construction notes:	Re-operation[A] is defined as the presence of OPCS codes indicating either of the following: 1. Re-excision (OPCS code: B284) 2. Repeated breast-conserving surgery on the same side where laterality information is available. 3. Completion mastectomy on the same side where laterality information is available. Re-operation must be after the initial breast-conserving surgery but excluding the first 7 days to avoid capturing post-operative complications. Laterality of re-operation is checked against the laterality of the initial breast-conserving surgery where available.	
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	Age-group and diagnosis year.	Age-group and diagnosis year.
Risk adjusted:	Age, grade, Charlson co-morbidity score[B], SCARF index[C], diagnosis year; invasive only: ER status, HER2 status, T-stage, N-stage	
Outlier reporting:	No	

[A] See Appendix 4 [B] See Appendix 5 [C] See Appendix 3

Performance Indicator 9: Overnight hospital admission for treatment-related toxicity within 30 days of a systemic anti-cancer therapy (SACT) cycle.

This performance indicator remains under development and is not yet reported on.

Performance Indicator 10: 3 year survival

3-year breast cancer-specific survival for patients with primary breast cancer.

Mortality data for England and Wales is compiled by the Office for National Statistics.

Table 5.10: Percentage of patients who survived their breast cancer for at least 3 years from their initial breast cancer diagnosis		
	England	Wales
Dates of diagnosis:	1/1/2021 to 31/12/2022	1/1/2021 to 31/12/2022
Numerator:	Number of people who are alive and free from breast cancer death within	Number of people who are alive and free from breast cancer death within

	the specified time period. Uses date of diagnosis from the Cancer Registry and ONS mortality data.	the specified time period. Uses date of diagnosis from CaNISC and ONS mortality data.
Denominator: Number of people diagnosed with early invasive primary breast cancer (stage 1 to stage 3A).	Final cohort as described in patient inclusion / exclusion, restricted to early invasive disease (stage 1-stage 3A)	Final cohort as described in patient inclusion / exclusion, restricted to early invasive disease (stage 1-stage 3A)
Construction notes:	Flexible parametric survival models and competing risk analysis were used to estimate risk-adjusted 3-year cause-specific survival.	
Country reporting:	England & Wales – combined and separate.	
Organisational Reporting level:	English NHS Trusts	Welsh Health Boards
Subgroup Reporting:	None	None
Risk adjusted:	Age, grade, Charlson co-morbidity score[A], SCARF index[B], ER status, HER2 status, T-stage, N-stage	
Outlier reporting:	Yes	

[A] See Appendix 5 [B] See Appendix 3

6. NHS organisations

The audit presents organisation-level findings by the NHS organisation of diagnosis. This is because this is the organisation where diagnosis and multidisciplinary team decisions are likely to be made.

There are some tertiary centres that mainly provide oncological treatment for people with breast cancer. This includes the Christie NHS Foundation Trust, Clatterbridge Cancer Centre NHS Foundation Trust, and Velindre NHS Trust. These tertiary centres are not included directly within audit outputs where findings are reported by the diagnosing NHS organisation. This is because patients are not diagnosed at these centres.

7. Statistical Analysis

All statistical analyses were conducted using *Stata version 17.*

Most results in the SotN Report are descriptive. The results of categorical data items are reported as percentages (%). Results are typically provided as an overall figure and broken down by NHS organisation of diagnosis (see NHS organisations section). Note that within tables in the SotN Report, the total percentage may not equal 100%, due to rounding.

7.1 Cause-specific survival

3-year cause-specific survival was calculated from the date of breast cancer diagnosis using ONS mortality data. To ensure the data was timely and to decrease the impact of the COVID-19 pandemic, breast-cancer specific survival metrics were limited to individuals diagnosed in 2021 and 2022. For those patients with no ONS date of death, a “date last known alive” or censoring date was calculated for use in survival analyses. For English patients provided by the NCRAS, this was taken to be the vital status date provided; where this date was missing, the day after the last reported date of death was used. For Welsh patients, the day after the last reported date of death was used. Standardised breast cancer-specific cumulative incidence function (CIF) for the risk of breast cancer death in the presence of the competing risk of death from other causes was estimated.

7.2 Suppression

Data suppression is primarily used to protect patient privacy by preventing the disclosure of sensitive information when dealing with small sample sizes or specific combinations of data points. It involves withholding certain data elements to minimize the risk of identifying individuals. For data quality and completeness, disclosure risk is deemed negligible and results have not been suppressed. For performance indicators, organisations with indicator denominator values of less than 10 have suppressed. Because small numbers are instable and may lead to spurious conclusions if not provided within context and with appropriate uncertainties, risk-adjusted percentages for indicators with fewer than 10 in the denominator have also been suppressed.

To avoid residual disclosure which could occur by summing across subcategories, where only a single trust within a cancer alliance has been supressed, a second trust has had the number and unadjusted percentage supressed to prevent residual disclosure however the risk adjusted percentage is provided.

In the patient characteristics table, for categories with small numbers (1-4) the number has been supressed by converting it to 2 and editing the numbers of a larger group to maintain the overall total. Therefore, the percentage provided for <5 is similar to the real value and the other numbers in the indicator group may contain a small amount of 'jitter'. Where the category value is 'Unknown', 'Not in HESAPC', 'Not in PEDW' or 'Not Recorded', i.e. it represents missing data, we have not suppressed small numbers.

7.3 Risk-adjustment of indicators

Multivariable logistic regression was performed to 'risk adjust' relevant indicators. The regression model is used to estimate the probability of a patient having an outcome (e.g., a given treatment) given patient and tumour characteristics. When summed, these individual probabilities produce the expected number of events at an organisation. The adjusted indicator value for an organisation is then calculated as: the observed number of events at an institution divided by the expected number (based on the national model), multiplied by the overall national average. This is called indirect standardisation. The tables of performance indicators state whether risk adjustment has been performed. Table 7.1 provides details on the datasets and variables used for risk adjustment.

Table 7.1: Risk Adjustment Variables

Data Item	**Additional detail**	
	England	**Wales**
Age at diagnosis	Age, included as a spline (birthmonth; birthyear from NCRD)	Age, included as a spline (AgeAtDiagnosis from CaNISC)
Grade	GRADE from NCRD	FirstGradeOfDifferentiationPathology from CaNISC
Charlson co-morbidity score[A]	The Charlson co-morbidity score is a commonly used scoring system for medical comorbidities, consisting of a grouped score calculated based on the absence (0) and presence (≥1) of 14 pre-specified medical conditions (**Appendix 5**). The score was calculated using information on secondary diagnoses (ICD-10 codes) recorded in HES APC (England) / PEDW (Wales) within the 24-month period prior to a person's diagnosis. For analysis, the Charlson co-morbidity score is grouped into four categories: • **0:** none of the 14 pre-specified comorbidities. • **1:** 1 of the 14 pre-specified comorbidities. • **2:** 2 of the 14 pre-specified comorbidities. • **3+:** 3 or more of the 14 pre-specified comorbidities (some risk adjustment may use 2+ in the event that very few people have a score of 3+).	
SCARF Index[B]	Using DIAG_01-DIAG20 from HES APC	Using Diagnosis01-12 from PEDW

Diagnosis year	Year extracted from diagnosisdatebest (NCRD)	Year extracted from DIAGNOSIS_DATE (Cohort data)
ER Status (Invasive only)	Using ER_STATUS; ER_SCORE from NCRD See Footnote of Table 4.1	Using ERStatus from CaNISC
HER2 Status (Invasive only)	Using HER2_STATUS from NCRD	Using Her2Status and Her2FIshStatus from CaNISC.
T-Stage (Invasive only)	Using T_BEST from NCRD	FirstTStagePathological from CaNISC
N-Stage (Invasive only)	Using N_BEST from NCRD	Using FirstNumberOfNodesExamined and FirstNumberOfNodesPositive from CaNISC
Adjuvant radiotherapy	Using treatment information from RTDS	Using CaNISC data for radiotherapy

[A] See Appendix 5 [B] See Appendix 3

7.4 Handling of missing data

For the risk-adjustment, multiple imputation techniques are used to allow us to make fair comparisons between hospitals without excluding patients with incomplete records. Specifically, missing values within the risk adjustment model were imputed with multiple imputation using chained equations, creating ten datasets and pooling model estimates using Rubin's Rules. The imputation models included all the variables in the risk adjustment model.

8. Outlier Process

The outlier process can be found in the separate audit outlier policy.

Appendix 1: Routine data sources

Overview of the data sources used for the SotN Report.

Country	Data source	Content
England	Cancer registry (NCRD and RCRD)	Data on all aspects of the cancer registration including information from hospital pathology systems.
England	COSD	Cancer Outcomes and Services dataset (COSD) items, are submitted routinely by service providers via multidisciplinary team (MDT) electronic data collection systems to the National Cancer Data Repository (NCDR) on a monthly basis.
England	SACT	Systemic Anti-Cancer Therapy (SACT) data contains information on chemotherapy dates, regimen(s) and dose(s).
England	RTDS	Radiotherapy dataset (RTDS) contains information on radiotherapy treatment including dates, prescription region and dose.
England	HES	Hospital Episode Statistics (HES) is the administrative database of all NHS hospital admissions in England; records were supplied by NHS Digital to NCRAS.
England	PCPD	Primary Care Prescription Database (PCPD) contains information on the use of endocrine therapy.
Wales	CaNISC	Cancer Network Information System Cymru (Canisc) contains data on all aspects of the cancer registration including investigations. (OLD SYSTEM)
Wales	CDF	Clinical Dataset Form (CDF) contains data on all aspects of the cancer registration including investigations (NEW SYSTEM)
Wales	PEDW	Patient Episode Database for Wales (PEDW) is the administrative database of all NHS hospital admissions in Wales.
Wales	RTH	Radiotherapy data (RTH) contains information on radiotherapy treatment.
England & Wales	ONS	Office for National Statistics (ONS) death data including date of death and cause of death.

Appendix 2: WHO Performance Status

Reference:

Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American Journal of Clinical Oncology. 1982;5(6):649-56

WHO Performance Status	**Definition**
0	Fully active, able to carry on all pre-disease performance without restriction.
1	Restricted in physically strenuous activity but ambulatory & able to carry out work of a light or sedentary nature.
2	Ambulatory and capable of all self-care but unable to carry out any work activities. Up & about more than 50% of waking hours.
3	Capable of only limited self-care, confined to bed or chair more than 50% of waking hours.
4	Completely disabled. Cannot carry on any self-care. Totally confined to bed or chair.
5	Dead.

Appendix 3: Secondary Care Administrative Records Frailty (SCARF) Index

Reference:

Jauhari Y, Gannon MR, Dodwell D, et al. Construction of the secondary care administrative records frailty (SCARF) index and validation on older women with operable invasive breast cancer in England and Wales: a cohort study. BMJ Open 2020;10:e035395. doi: 10.1136/bmjopen-2019-035395

Deficit			
Activity limitation	Diabetic complications	Hypotension	Requirement for care
Anaemia	Falls	Ischaemic heart disease	Respiratory disease
Arthritis	Foot problems	Incontinence	Skin ulcer
Cardiac arrhythmias	Fragility fracture	Neurodegenerative disorders	Sleep disturbance
Cerebrovascular disease	Hearing impairment	Nutritional Problems	Social vulnerability
Chronic kidney disease	Heart failure	Osteoporosis	Thyroid disease
Cognitive and mental health problems	Heart valve disease	Peptic ulcer	Urinary system disease
Diabetes	Hypertension	Peripheral vascular disease	Visual impairment

Appendix 4: OPCS Codes

Breast conserving surgery	Mastectomy	Reconstruction
B28.1, B28.2, B28.3, B28.5, B28.7, B28.8, B28.9, B41.1, B41.2, B41.8, and B41.9	B27	B29.1, B29.2, B29.3, B29.4, B29.8 B29.9, B30.1, B30.8, B30.9, B38.1, B38.2, B38.8, B38.9, B39.1, B39.2, B39.3, B39.4, B39.5, B39.8, B39.9, and S48.2.

Laterality of surgery was identified via Z94 codes. A reconstruction was defined as immediate if the OPCS-4 codes for mastectomy and reconstruction occurred on the same date of surgery and laterality matches where this is available.

Appendix 5: Charlson co-morbidity score

Reference:

Armitage JN, van der Meulen JH. Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score. Br J Surg 2010;97:772-81. doi https://doi.org/10.1002/bjs.6930

Pre-specified conditions included in the assignment of Charlson co-morbidity Score.

Charlson co-morbidity score conditions
Myocardial infarction
Dementia
Diabetes mellitus
Metastatic solid tumour
Congestive cardiac failure
Chronic pulmonary disease
Hemiplegia or paraplegia
AIDS/HIV infection
Peripheral vascular disease
Rheumatological disease
Renal disease
Cerebrovascular disease
Liver disease
Any malignancy

Note: AIDS/HIV diagnoses cannot be identified in HES APC data because of legal requirements for NHS trusts to remove patient identifiers from legally restricted records, including those containing diagnoses of HIV/AIDS. These diagnoses are also not found in linked PEDW data.