NABCOP 2021 Annual Report Methodology

Contents

Version dated 05/08/2021

Data receipt and processing

Routine data collection

Patient-level data on many aspects of breast cancer care are routinely collected in hospitals and mandatorily submitted to national organisations. These existing electronic data flows are used by the National Audit of Breast Cancer in Older Patients (NABCOP) in order to reduce the burden of data collection on staff and patients. The NABCOP uses this patient data, collected by the National Cancer Registration and Analysis Service (NCRAS) in England and the Wales Cancer Network (WCN), to report on breast cancer care for older women.

For patients in England, the NCRAS provides data from its Cancer Analysis System (CAS), which collates patient data from a range of national data feeds across all NHS acute hospitals. These data feeds included:

- National cancer registrations, which include information directly from hospital pathology systems.
- A screening flag from the NHS Breast Screening Programme (NHSBSP) and Association of Breast Surgery (ABS) screening audit (previously provided by the Screening Histories Information Manager (SHIM) system).
- Cancer Outcomes and Services Dataset (COSD) data items.
- Systemic Anti-cancer therapy (SACT) data.
- Radiotherapy dataset (RTDS).
- Hospital Episode Statistics (HES) data, including Admitted Patient Care (APC); Outpatients (OP); Accident & Emergency (A&E).
- Date and cause of death from the Office for National Statistics (ONS).

Data from the above sources were provided for the cohort of women diagnosed from 01-Jan-2014 to 31-Dec-2018, for the NABCOP 2020 Annual Report; these data were used within the prescriptions and outcomes chapters of the NABCOP 2021 Annual Report, with a refresh of SACT, RTDS, HES and ONS data for the cohort.

In addition, a snapshot of endocrine therapy prescriptions dispensed in 2018 was provided from the Primary Care Prescription Database (PCPD), linked to the NABCOP 2014-18 cohort via the pseudonymised patient identifier.

For patients in Wales, the WCN provides national cancer registrations data using the Cancer Network Information System Cymru (Canisc) electronic patient record system. The cancer record for each patient is linked to the following data:

- Patient Episode Database for Wales (PEDW).
- Date and cause of death from the Office for National Statistics (ONS).

Data from the above sources were provided for the cohort of women diagnosed from 01-Jan-2014 to 31-Dec-2018, for the NABCOP 2020 Annual Report; these data were used within the outcomes chapters of the NABCOP 2021 Annual Report

A data specification document is published online at **www.nabcop.org.uk**; which provides a comprehensive list of those data items the NABCOP receives from the NCRAS and WCN, along with their data source (e.g. COSD, HES etc).

Data for women diagnosed in 2019/20

For the NABCOP 2021 Annual Report an update of the usual Cancer Registration data was not available for England; this would have included women diagnosed up to 31-Dec-2019. The NCRAS therefore provided data from the Rapid Cancer Registration Dataset (RCRD) for women diagnosed from 01-Jan-2018 to 31-Jul-2020, linked to RTDS, SACT, HES, and ONS data.

For patients diagnosed and treated in Wales the WCN provided the usual updated data, on women aged 50+ years diagnosed with breast cancer in Wales from 01-Jan-2019 to 31-Jul-2020, from CANISC. In order to provide data for the same time frame as that covered by the RCRD for England, data were released prior to being fully validated. Linked data from Cancer Standards and the Radiotherapy Data were also provided.

NABCOP 2020 Organisational Audit (OA)

The OA was designed to be delivered as an online survey questionnaire. The survey questions were developed by the NABCOP Project Team, and piloted among members of the NABCOP Clinical Steering Group. Questions were refined based on feedback from the pilot. The final survey consisted of 26

questions across five topics. Survey questions, and responses from each participating NHS organisation, can be viewed within the NABCOP 2021 Annual Report NHS Organisation Data Viewer[1].

Using the NABCOP contact list, individuals in breast cancer teams within NHS trusts in England and local health boards in Wales were sent the survey via email on 08-Oct-2020. NHS organisations who did not respond were contacted with a follow-up email or telephone reminder. Information about the survey was also included in the NABCOP newsletters, sent to all stakeholders in October 2020 and January 2021. The survey was open for 16 weeks, and closed to responses on 29-Jan-2021.

In the analysis, a single response is included for each participating NHS organisation. In cases where more than one response was received from the same NHS organisation, but there were differences in the levels of completeness, the more complete response was taken forwards. Where multiple complete responses were received from the same NHS organisation, respondents were contacted and provided with the following options, to agree one complete submission:

- **Option 1:** Responses to multiple-choice questions, from multiple individual submissions, could be combined to capture all options selected by the team. Where discrepancies arose between responses to single-choice questions, teams were instructed to agree a single submission.
- **Option 2:** Survey respondents could select the results from one survey as their submission.
- **Option 3:** A subsequent survey could be submitted, as a co-ordinated response.

The OA also elicited free-text responses to open questions (such as those given to 'other, please specify'). These responses underwent thematic analysis to identify key themes; which were tabulated and their frequency presented or described within the OA chapter.

The NABCOP cohort – patient inclusion

The NCRAS and WCN extract all the data, described in the previous section for patients fulfilling the following criteria:

Include:

- Women
- Aged 50 years or over at the point of diagnosis (no upper age limit)
- Registered diagnostic ICD-10 code of C50 (invasive breast cancer) or D05 (non-invasive breast cancer)
- With a valid diagnosis date (typically from 01/01/2014 to the most recently available date)

Exclude:

- Women whose cancer was only reported on their death certificate

Note: Inclusion of data on men are not considered for the NABCOP at this time, primarily due to the low incidence meaning analyses considering variation by age and across NHS organisations would be infeasible. Other sources provide information on annual incidence of male breast cancer.

The NABCOP team then applies the following exclusion criteria[2] to define the cohort for analysis:

1. Date of diagnosis is the same as ONS date of death.
2. There is a previous diagnosis of breast cancer before 01/01/2014.
3. The registration is for bilateral breast cancer.
4. The patient has multiple cancer registrations during the audit period.
5. Diagnosed and treated outside of an NHS organisation in England or Wales.
6. Place of diagnosis is not provided or the patient is assigned to an NHS organisation with no active breast unit.
7. Diagnosed and treated within an NHS organisation with less than 30 allocated registrations of breast cancer, in women aged 50 years and over, per year.
8. ICD-10 code not recorded as C50 or D05.1 (ductal carcinoma in situ).

[1] https://www.nabcop.org.uk/resources/nabcop-2021-annual-report-supplementary-materials/

[2] For analysis using the RCRD for England it was not possible to apply exclusions 2, 3 or 4; additionally it was not possible to distinguish D05.1 tumours from D05 tumours

Preparation for analysis

The NABCOP project team, based at the Clinical Effectiveness Unit (CEU)[3] received the national data from the NCRAS and WCN between October and December in the year prior to publication of the annual report. A series of steps are performed to prepare the complex and large datasets for analysis.

Specifically, using specialised statistical software[4], the project team:

Clean the datasets received.

- Checking the datasets for discrepancies
- Identifying and removing duplicate records
- Data augmentation (combining multiple sources of information).

↓

Merge the relevant datasets.

This involves restructuring the English and Welsh datasets so that they have the same format and can be analysed simultaneously.

↓

Where necessary, derive new information (data items) by combining different data items.

For example, the Charlson comorbidity index is calculated using patient diagnosis information in HES and PEDW in the two years prior to the cancer diagnosis.

↓

Conduct analyses and present audit results.

In aggregated tables and graphs for annual reports and other outputs (such as peer reviewed articles and papers).

[3] The CEU is an academic collaboration between The Royal College of Surgeons of England and the London School of Hygiene and Tropical Medicine, and undertakes national clinical audits and research. Since its inception in 1998, the CEU has become a national centre of expertise in methods, organisation, and logistics of large-scale studies of the quality of surgical care.

[4] Stata® is a statistical package for data analysis, data management, and graphics (https://www.stata.com/)

Definition of variables

Patient fitness

We are interested in the fitness of a patient at the point of diagnosis, and when treatment decisions are made. This is because the NABCOP aims to understand what patient and tumour factors influence the choice of treatment(s) offered to a patient. These factors are taken into account when the audit produces information by individual NHS organisation so their statistics can be compared.

The **World Health Organization (WHO) performance status (PS)** classification is a measure of how disease(s) impact(s) a patient's ability to manage on a daily basis, [Oken *et al* 1982].[5] The NABCOP uses all available data on WHO PS to understand treatment decisions for a patient; the table below highlights where the value is recorded in the data the NABCOP receives.

WHO Performance Status sources		
Country	**Source**	**Associated date**
England	COSD	MDT discussion date
England	SACT	Regimen/cycle start date
Wales	Canisc	Investigation date

WHO PS at diagnosis is then calculated for a patient based on the following criteria, that:

- the value recorded is valid (i.e. 0–4).
- the value provided has an associated date that is prior to the date of treatment starting[6] and within two months of diagnosis.

Where there are multiple records of a patient's WHO PS that fulfil the above criteria the value closest to diagnosis is taken. Where multiple values have the same date the highest (i.e. worst) is taken. Historically this information is poorly recorded within routine data.

The presence of comorbidities is not captured within a single data item by the national registration services. The NABCOP team therefore uses the Royal College of Surgeons of England (RCS) modified **Charlson Comorbidity Index (CCI)** [Armitage *et al* 2010][7] to describe these. The CCI 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 1). The CCI was calculated using information on secondary diagnoses (ICD-10 codes) recorded in HES APC/PEDW within the 24-month period prior to a patient's diagnosis. For the purpose of analysis, the CCI is grouped into three categories:

- (0) none of the 14 pre-specified comorbidities;
- (1) only 1 of the 14 pre-specified comorbidities;
- (2+) 2 or more of the 14 pre-specified comorbidities.

Among older patients, frailty plays an important role in what breast cancer treatments are offered to patients. This is because in frail women, the ability to tolerate stressors such as surgery, radiotherapy or chemotherapy may be significantly reduced, leading to morbidity and mortality. NHS organisations are recommended to screen for frailty using a formal assessment tool, although assessment is limited by the lack of an agreed instrument and the potential inaccuracies of simple tools. The **Secondary Care Administrative Records Frailty (SCARF) Index**[8] is based on the 'cumulative deficit' model [Clegg et al 2016], and describes frailty in relation to 32 different symptoms, signs, diseases and disabilities (referred to as deficits; Appendix 2). The index translates the 32 deficits into ICD-10 codes and counts the number of deficits in HES APC/PEDW records within the 24-month period prior to a patient's diagnosis. This methodology has been internally validated, and it produces the type of pattern that would be expected from a measure of frailty.

[5] 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

[6] Based on dates for surgery or anti-cancer treatments.

[7] Armitage JN, van der Meulen JH, Royal College of Surgeons Co-morbidity Consensus G. Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score. Br J Surg. 2010;97(5):772-81.

[8] 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

Socioeconomic status

The Index of Multiple Deprivation (IMD) is a relative measure of deprivation which ranks every small area in England (Lower-layer Super Output Areas (LSOAs), containing ~1,500 residents/650 households) from 1 (most deprived area) to 28,844 (least deprived)[9]. The five categories used within the NABCOP analyses are calculated by dividing the ranks into five equal groups and correspond to the most deprived 20 percent down to the least deprived 20 percent.

NHS organisation of diagnosis

The NABCOP presents organisation-level findings by the NHS organisation of diagnosis. Where this information is not provided for a patient, or where the organisation assigned does not fulfil the pre-specified inclusion criteria[10] for including the patient in the NABCOP, the following steps are followed in order to assign a diagnosing NHS organisation:

1. Use the surgery provider code (as provided within HES/PEDW) which fulfils the NABCOP pre-specified inclusion criteria; use the provider code associated with the earliest record of primary surgery (breast conserving surgery or mastectomy).
2. Use the MDT provider code for English patients, which fulfils the NABCOP pre-specified inclusion criteria; use the provider associated with the earliest MDT discussion date.

Patients provided by the NCRAS can have a Welsh local health board code assigned, with no further record of treatment within an English NHS trust. These patients cannot be included in the NABCOP analysis. This is due to the uncertainty around whether the full care pathway for such a patient is captured within the data provided. The same is true for patients provided in the WCN data with an English trust code assigned as the place of diagnosis and no record of further treatment within a Welsh local health board.

Treatment allocation

A patient was considered to have received surgery for breast cancer where they were identified as having received a mastectomy or breast conserving surgery within 12 months of their diagnosis date.

Those women for whom there was no breast surgical information reported in HES/PEDW, or for whom surgery was more than 12 months after diagnosis, are described as having 'no surgery'. In many cases, this will be because women had another course of treatment, such as primary endocrine therapy (PET). However, in some cases, it will be because the surgery was performed in independent healthcare providers in England and Wales. Independent hospitals do not generally contribute treatment information to the national cancer registration services datasets received by the NABCOP.

Breast conserving surgery

HES APC (England) and PEDW (Wales) records were used to identify patients who had breast conserving surgery (BCS) using the OPCS-4 procedure codes B28.1, B28.2, B28.3, B28.5, B28.7, B28.8, B28.9, B41.1, B41.2, B41.9. Where information was missing in HES/PEDW the Cancer Registration treatment records were used to identify receipt of BCS, with the same OPCS-4 codes used.

Mastectomy (with reconstruction)

HES APC (England) and PEDW (Wales) records were used to identify patients who had a mastectomy using the OPCS-4 procedure code B27 (for reconstruction the codes are B29.1-4, B29.8 B29.9, B30.1, B30.8, B30.9, B38.1, B38.2, B38.8, B38.9, B39.1-5, B39.8, B39.9, S48.2). Where information was missing in HES/PEDW the Cancer Registration treatment records were used to identify receipt of mastectomy (with reconstruction), with the same OPCS-4 codes used.

Adjuvant radiotherapy

For England, use of radiotherapy was determined from the RTDS. For Wales, Canisc was used to identify women receiving radiotherapy, along with the radiotherapy dataset provided for women diagnosed in 2019-2020.

[9] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/853811/IoD2019_FAQ_v4.pdf
[10] A private hospital code provided; the organisations diagnoses less than 30 patients aged 50+ years with breast cancer each year; the organisation is a tertiary centre; the hospital is in a different country to the data provider; the organisation has no active breast unit.

Chemotherapy

For England, the SACT data item “drug group” was used to identify those women who received chemotherapy. For Wales, Canisc data were used; within these data there was no information on drug used or cycle dates so analysis beyond a “Yes/No” receipt of chemotherapy was not possible.

Assigning outcomes

Reoperation

This outcome was derived from HES APC/PEDW for women diagnosed with ductal carcinoma in situ or early invasive breast cancer, in England and Wales, between 1 January 2014 and 31 December 2018, who had breast conserving surgery (BCS). To create a variable for those patients who had a reoperation within 3 months of initial BCS we identified those patients who had a first BCS within 12 months of diagnosis, calculated the difference in days between the first surgery date and any subsequent BCS or mastectomy date, and flagged those patients with a reoperation recorded within 3 months. Subsequent BCS or mastectomy procedures dated within seven days of the initial BCS were considered to most likely be due to a complication from the original surgery and so not counted as a reoperation.

Chemotherapy related overnight admission

This outcome was derived from HES APC/PEDW for women diagnosed with early invasive breast cancer in England, between 1 January 2014 and 31 December 2018, who had adjuvant chemotherapy and had at least one related overnight admission within 30 days of a cycle. Patients were flagged has having a chemotherapy related overnight admission where an overnight admission, recorded with a diagnostic (ICD-10) code (Appendix 3) indicating a chemotherapy related admission recorded, was within 30 days of a chemotherapy cycle.

Death

Record of death for an individual patient was coded where a date of death was provided within the ONS data.

Statistical analysis

All statistical analyses were conducted using Stata version 15.1.

Most results in the NABCOP 2021 Annual Report are descriptive. The results of categorical data items are reported as percentages (%). In descriptive analyses of continuous variables, the distribution of values is described using appropriate statistics (e.g. mean and standard deviation or median and interquartile range). Results are typically provided as an overall figure and broken down by age at diagnosis, (and by diagnosing NHS organisation in the online NHS tables). Note that within tables in the annual report, the total percentage may not equal 100%, owing to rounding errors.

We follow the Office for National Statistics (ONS) policy on the publication of small numbers to minimise the risk of patient identification from these aggregate results.

Adjusted outcomes

For analyses looking at reoperation rates following initial breast conserving surgery, statistical models were fitted to calculate a "risk adjusted" percentage to account for differences in case-mix, allowing comparison across organisations. Such models included clinically relevant patient and tumour factors likely to contribute to treatment decisions. The models were then used to estimate the probability of an individual having the outcome (at least one reoperation); these individual probabilities were summed to calculate an expected number of outcomes. This was combined with the observed outcomes to produce the risk-adjusted indicator value for each NHS organisation (a method known as indirect standardisation). Categories of "unknown" were created where data items had missing, unintelligible or conflicting information, in order to ensure all patients contributed to the statistical models.

Funnel plots

Funnel plots are used to make comparisons, and graphically display variation, between NHS organisations. The plots are generated by plotting the rate for each NHS organisation against the total number of patients used to estimate the rate. The 'All NABCOP NHS Orgs %' is the average rate across all NHS organisations.

The funnel plots include control limits defining differences corresponding to two standard deviations (inner limits) and three standard deviations (outer limits) from the overall average. These limits get wider where organisations have a lower volume of patients and narrower where there is higher volume, reflecting the increased variability in results when there are fewer patients per organisation.

Relative survival

Estimates of relative survival were obtained using stpm2 and its post estimation commands, within Stata, with population mortality data from ONS to provide the baseline survival.

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 last reported date of death was used.
- For Welsh patients, the last reported date of death was used.

Appendix 1: Charlson Comorbidity Index

Pre-specified conditions included in the assignment of Charlson Comorbidity Index.

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	

Appendix 2: Secondary Care Administrative Records Frailty Index

Pre-specified deficits included in the calculation of the Secondary Care Administrative Records Frailty Index.

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 3: Coding for chemotherapy-related admission

The table below provides details of the diagnostic codes used to identify chemotherapy-related acute care visits in administrative data among patients receiving chemotherapy for early-stage breast cancer. The codes were validated in work by Krzyzanowska et al (2018)[11] which looked at using administrative data to accurately identify treatment-related complications.

Toxicity	Description	ICD-10 code
Neutropenia	Agranulocytosis- Including drug induced	D70
Fever	Other Specified Fever (Chills with fever; Persistent fever; Fever with rigors)	R508
	Fever unspecified (Fever NOS; FUO; Hyperpyrexia NOS ; Pyrexia NOS ; Pyrexia UO)	R509
Infection	Infectious and parasitic diseases	A00-B99
	Line associated Infection	T82.7
	Bronchitis	J20-J22
	Pneumonia	J12-J18
	Kidney Infection	N10, N390
	Acute cystitis	N300
	Cellulitis	L00-L08
	Empyema	J86
	Abscess of lung/mediastium	J85
	Other septicaemia	A41
	Septicaemia unspecified	A419
	Septicaemia other	A418
GI Toxicity	Diarrhea	K52
	Functional diarrhea	K59.1
	Nausea/emesis	R11
	Heartburn	R12
	Constipation	K59.0
	Obstruction	K56
	Stomatitis	K12
	Cachexia	R64.0
	Anorexia	R63.0
Other Systemic Treatment Related	Hyponatremia	E87.1
	Hypokalemia	E87.6
	Electrolyte disorder Magnesium disorder	E87.0, 2, 3, 4, 5, 7, 8 E834
	Dehydration/hypovolemia	E86
	Malaise/Fatigue	R53
	Syncope	R55
	Dizziness	R42
	Hypotension	I959
	Fe deficiency anaemia	D50
	Other deficiency anaemia	D51-D53
	Aplastic anemia	D60, D61
	Other and unspecified anemia	D62-D64
	Thrombocytopenia	D69.5, D69.6
	Other venous embolism and thrombosis	I82
	Rash and non-specific skin eruptions	R21
	Hyperglycemia	R73
	Phlebitis	I808

Note: ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th revision; NOS= not otherwise specified; FUO= fever of unknown origin; UO= unknown origin

[11] Krzyzanowska MK, Enright K, Moineddin R, Yun L, Powis M et al. Can Chemotherapy-Related Acute Care Visits Be Accurately Identified in Administrative Data? *J Oncol Pract* 2018 Jan;14(1):e51-e58.