NABCOP Annual Report Methodology 2019

Data receipt and processing

What data do the NABCOP receive for analysis?

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 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).

For patients in England, the NCRAS provided 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 include:

- National cancer registrations, which include information directly from hospital pathology systems.
- Cancer Outcomes and Services Dataset (COSD) data items.
- Systemic Anti-cancer therapy (SACT) data.
- Radiotherapy dataset (RTDS).
- Hospital Episode Statistics (HES) data.
- Date of death from the Office of national statistics (ONS).

Additionally the NABCOP received English Cancer Patient Experience Survey (CPES) data, completed by patients diagnosed in England in 2015.

For patients in Wales, the WCN has provided national cancer registrations data using the Cancer Network Information System Cymru (Canisc) electronic patient record system. Radiotherapy information was also available from Canisc for the 2019 Annual Report. The cancer record for each patient is then linked to the following data:

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

Appendix 1, Table A1_1 provides more detail on the data sources listed above and the information they contain. Additionally, a data specification document is published online at **www.NABCOP.org.uk**; this 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).

Which patients are data provided for?

The NCRAS and WCN extracted 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 codes of C50 (invasive breast cancer) or D05 (non-invasive breast cancer)
- With a diagnosis date between 01/01/2014 to 31/12/2017

Exclude:

- Women whose cancer was only reported on their death certificate

The NABCOP team then applies the following exclusion criteria 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. Diagnosed with more than one breast tumour.
4. Diagnosed and treated outside of an NHS organisation in England or Wales.
5. Unclear place of diagnosis.
6. 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.
7. ICD-10 code not recorded as C50 or D05.1 (ductal carcinoma in-situ).

How are data prepared for analysis?

The NABCOP project team, based at the Clinical Effectiveness Unit (CEU)[1] receives the national data from the NCRAS and WCN between October and December in the year prior to publication of the annual report. They then perform a series of steps to prepare the complex and large datasets for analysis.

Specifically, using specialised statistical software[2], 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 2 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).

Measures of fitness

We are interested in the fitness of a patient at the point of diagnosis, and when treatment decisions are being made. The reason for this is that the NABCOP aims to understand what patient and tumour factors influence the choice of treatment. These factors can then be taken into account when the audit produces information on NHS organisations so that their statistics can be compared.

This section provides information on measures of fitness the NABCOP receives or derives for the purpose of analysis.

Performance Status

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].[3]

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 (see **Appendix 1, Table A1_2** for the definition of each WHO PS value).

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:

- Value recorded is valid (i.e. 0–5).
- Provided with an associated date that is prior to the date of treatment starting[4] 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.

[1] 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.

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

[3] 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

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

Comorbidity

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][5] to describe these.

The CCI is a commonly used scoring system for medical comorbidities. It consists of a grouped score that is calculated based on the absence (0) and presence (≥1) of 14 pre-specified medical conditions. The CCI was calculated using information on secondary diagnoses (ICD-10 codes) in the hospital admission data (HES/PEDW) recorded within the 24-month period prior to a patient's diagnosis of breast cancer. The conditions covered are detailed in **Appendix 1, Table A1_3**.

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.

Frailty

Among older patients, frailty plays an important role in the selection of breast cancer treatments. This is because, in frail women, surgery may pose a significant risk, and the ability to tolerate chemotherapy and radiotherapy may be reduced.

NHS organisations are recommended to assess patients for frailty using a formal assessment tool, although services are limited by the lack of an agreed instrument and the potential inaccuracies of simple tools.

A common way of measuring frailty is to use a scale based on the "cumulative deficit model". This defines frailty in relation to a range of variables that include symptoms, signs, diseases, disabilities and abnormal laboratory values. These are collectively referred to as deficits [Mitnitski 2001].[6] Recently, Clegg et al [2016][7] proposed a method of deriving a frailty index from primary care electronic health records using Read codes to capture 36 individual variables that are biologically plausibly associated with frailty, which was named the electronic Frailty Index (eFI; see **Appendix 1, Table A1_4** for the list of deficits).

We used this approach, translating 35 of the 36 deficits into ICD-10 codes that could be identified within the diagnosis fields in hospital admissions data (note that it was not possible to find a translation for the deficit of poly-pharmacy). This methodology has been internally validated, and it produces the type of pattern that would be expected from a measure of frailty.

ASA score

The American Society of Anaesthesiologists (ASA) classification is a scoring system based on perioperative health and comorbidities of a surgical patient. It is used to assess the physical status of patients before surgery, and patients are given a score ranging from 1 to 5. A higher ASA score denotes a higher risk of perioperative complications in the short and long term.

As the score is predominantly assigned to patients having surgery, it is of limited use to the NABCOP towards understanding treatment decisions as those patients not receiving surgery will not be assessed by an anaesthetist and assigned an ASA score.

[5] 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.
[6] Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. The Scientific World. 2001; 1.
[7] Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E, et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age and Ageing. 2016;45(3):353-60.

Statistical methods

Statistical analysis

For the majority of the indicators presented within the NABCOP Annual Reports, the results are reported as percentages (%). 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.

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). 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.

All analyses were undertaken by the NABCOP project team at the Clinical Effectiveness Unit, Royal College of Surgeons of England.

Assigning NHS organisation of diagnosis

Details of the organisation at which each patient was diagnosed are provided in both the England and Wales datasets. For England, the NHS trust of diagnosis is assigned by the NCRAS.

There are some patients, however, where either no organisation is assigned or the organisation assigned does not fulfil the criteria for including the patient in the NABCOP (e.g. private hospital code; small trust numbers; tertiary centre; hospital in a different country to the data provider). For such cases, 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 inclusion criteria; use the provider code associated with earliest record of primary surgery (breast conserving surgery the provider or mastectomy).
2. Use the MDT provider code for English patients, which fulfils the NABCOP inclusion criteria; use the provider associated with 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.

Coding of key patient characteristics

The NABCOP uses data on patient characteristics provided from several data sources. Broadly, information on patient characteristics are captured within the cancer registry datasets, typically being measured around the time of diagnosis. The NABCOP focuses on measures of fitness, as well as method of presentation. This latter characteristic is grouped as screen-detected status “Yes”/”No”; specifically a woman is classed as having screen detected cancer where the data item screen detected is reported as “Yes” or where the referral route reported is screening. Conversely information on screen detected status influences the coding of route to diagnosis (section; which is considered to be screening where screen detected status is reported as “Yes”.

Coding of key tumour characteristics

The NABCOP uses data on tumour characteristics provided from several data sources. **Appendix 2, Table A2_1** defines the key tumour characteristics in terms of the data source and what analyses they are used in. More specifically, where a woman’s breast cancer stage is not reported in the primary data sources, this is calculated from available information their T,N,M stage, using the UICC TNM classification system (**Appendix 1, Table A1_5**).

In the NABCOP 2019 Annual Report, women are reported as an “unknown” overall stage, if there is lack of full information on all three components; or if the stage recorded in the datasets contradicts the ICD-10 diagnosis (e.g. stage 0 recorded for ICD-10 code of C50, invasive cancer).

Coding receipt of treatment

Primary surgery

Information on patterns of surgery was derived using the data in the routine hospital datasets (HES for English patients; PEDW for Welsh patients) as well as the cancer registration treatment datasets. We identified when a patient underwent different types of surgery by searching for admissions in which the relevant OPCS procedure codes appear (such as for mastectomy). A list of procedure codes is provided in **Appendix 1, Table A1_6**.

For the NABCOP analyses considering use of surgery, those women for whom there was no breast surgical information reported in HES/PEDW are described here 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.

Adjuvant therapy

Use of radiotherapy was determined from the RTDS, for those women diagnosed and treated in England, and from the Welsh radiotherapy dataset for those women diagnosed and treated in Wales.

Use of chemotherapy was determined from the SACT data, for those women diagnosed and treated in England, and from the Canisc data for those women diagnosed and treated in Wales.

For the NABCOP analyses, receipt of adjuvant treatment within a pre-specified time frame following surgery (see **Appendix 2, Table A2_1**) was used to code receipt of adjuvant treatment as "Yes". Where the adjuvant treatment was either not reported, or it was reported as starting outside of the pre-specified time frame, the patient was considered not to have received the relevant treatment. The time frames were used in order to more accurately conclude that the treatments were given for the index breast cancer episode and not for recurrence.

Dates

Diagnosis date

The date of diagnosis[8] used to define the audit group, and subsequently used within relevant analyses, was provided within the NCRAS Registry dataset for English patients and within the Canisc dataset for Welsh patients. This is calculated using a methodology in accordance with the European Network of Cancer Registries.

Triple diagnostic assessment

In order to determine whether, for those patients not presenting through routine screening, triple diagnostic assessment was received in a single visit, the following conditions have to be met:

- Patient has a reported date of biopsy or cytology.
- Patient has a matching date of mammogram.
- OR patient has no mammogram date but has a matching date first seen (English patients only; reported within COSD).

Censor date for patients alive at the end of the Audit period

For those patients with no ONS date of death, a "date last known alive" or censoring date is calculated for use in survival analyses.

- For English patients provided by the NCRAS, this is taken to be the vital status date provided; where this date is missing, the day after the last reported date of death is used.
- For Welsh patients, the day after the last reported date of death is used.

Risk adjustments

For indicators evaluating receipt of treatment, such as having surgery, statistical models were fitted to calculate a "risk adjusted" percentage to account for differences in case-mix, allowing comparison across organisations. To account for any clustering within NHS organisation multi-level, mixed-effects logistic regression models were fitted to the data. 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 experiencing the outcome (e.g., receiving treatment); these individual probabilities were summed to calculate an expected number of outcomes. This was combined with the observed

[8] Based on the data available this was the date of biopsy for most cases.

outcomes to produce the risk-adjusted indicator value for each NHS organisation (a method known as indirect standardisation). Details of the patient and tumour characteristics adjusted for are provided within **Appendix 2, Table A2_2**. 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.

Patient experience

We analysed data captured by the following seven questions collected by the CPES 2015 questionnaire: 12, 16, 17, 18, 45, 48 and 59. These are listed in full, along with the possible responses, in **Appendix 3**. Results are reported as percentages (%),typically provided as an overall figure, but further broken down by age at diagnosis where a difference was observed.

Appendix 1

Table A1_1: Overview of the data sources and content provided for the NABCOP 2019 Annual Report.

Country	Data source	Content
England	Cancer registry	Data on all aspects of the cancer registration including information from hospital pathology systems.
England	COSD	Cancer Outcomes and Servives 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, regime(s) and dose.
England	RTDS	Radiotherapy dataset (RTDS) contains information on radiotherapy treatment including dates, prescripion 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	CPES	Cancer Patient Experience Survey (CPES), completed by patients diagnosed in England in 2015
Wales	Canisc	Cancer Network Information System Cymru (Canisc) contains data on all aspects of the cancer registration including investigations
Wales	PEDW	Patient Episode Database for Wales (PEDW) is the administrative database of all NHS hospital admissions in Wales.
England & Wales	ONS	Office of National Statistics (ONS) death data including date of death and cause of death.

Table A1_2: WHO Performance Status values and corresponding definition.

WHO PS	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

Table A1_3: Charlson Comorbidity Index – pre-specified conditions.

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	

Table A1_4: electronic Frailty Index – pre-specified deficits.

Deficit			
Activity limitation	Falls	Ischaemic heart disease	Respiratory disease
Anaemia & haematinic deficiency	Foot problems	Memory & cognitive problems	Skin ulcer
Arthritis	Fragility fracture	Mobility & transfer problems	Sleep disturbance
Atrial fibrillation	Hearing impairment	Osteoporosis	Social vulnerability
Cerebrovascular disease	Heart failure	Parkinsonism & tremor	Thyroid disease
Chronic kidney disease	Heart valve disease	Peptic ulcer	Urinary incontinence
Diabetes	Housebound	Peripheral vascular disease	Urinary system disease
Dizziness	Hypertension	*Polypharmacy*	Visual impairment
Dyspnoea	Hyotension/syncope	Requirement for care	Weight loss & anorexia

Table A1_5: TNM stage groupings.

Stage grouping	T stage	N stage	M stage
DCIS / Stage 0	Tis	N0	M0
Early breast cancer			
IA	T1	N0	M0
IB	T0 / T1	N1(mi)	M0
IIA	T0 / T1 T2	N1 N0	M0
IIB	T2 T3	N1 N0	M0
IIIA	T0, T1, T2 T3	N2 N1, N2	M0
Locally advanced disease			
IIIB	T4	N0, N1, N2	M0
IIIC	Any T	N3	M0
Metastatic disease			
IV	Any T	Any N	M1

Key:

Tumour size –
T1 = 1-20mm;
T2 = 21-50mm;
T3 = 51+mm;
T4 = tumour spread to skin or chest wall.

Table A1_6: OPCS codes used to define surgical procedures for analysis.

Surgical procedure	OPCS code(s)
BCS	B28.1, B28.2, B28.3, B28.5, B28.7, B28.8, B28.9
Mastectomy	B27
Reconstruction	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
SLNB	T86.2, T87.3, T91.1
AND	T85.2

Appendix 2

Table A2_1: Details of data items used within the NABCOP 2019 Annual Report; data source and where they are used.

Item	Where data comes from		Indicator
	England	Wales	
Non-invasive grade	COSD BR4160	Canisc	Data completeness
Invasive grade	COSD BR4170	Canisc	Data completeness; risk-adjustment
ER status	COSD BR4220 COSD BR4230 (ER Score)	Canisc	Recorded molecular marker status; risk-adjustment
HER2 status	COSD BR4280 COSD BR4310 (HER2 ISH)	Canisc	Recorded molecular marker status; risk-adjustment
PR status	COSD BR4290 COSD BR4300 (PR Score)	Canisc	Data completeness
Whole tumour size	COSD BR4190	Canisc	Data completeness
DCIS size	COSD BR4180	Canisc	Data completeness
Tumour stage	COSD CR0520 COSD CR0620 COSD CR0910	Canisc	Data completeness; risk-adjustment
Nodal stage	COSD CR0540 COSD CR0630 COSD CR0920	Canisc	Data completeness; risk-adjustment
Metastases stage	COSD CR0560 COSD CR0640 COSD CR0930	Canisc	Data completeness
Overall stage	COSD CR0580 COSD CR0610 COSD CR0940	*Not available*	Data completeness; risk-adjustment
WHO performance status	COSD CR0510 SACT	Canisc	Data completeness; Receipt of surgery by age and patient fitness
Nodes excised	COSD CR0890	Canisc	Data completeness
Nodes positive	COSD CR0900	Canisc	Data completeness
Source of referral	COSD CR1600	Canisc	Route to diagnosis; risk-adjustment
Screen-detected status	SHIM COSD CR1600 = screening	Canisc	Route to diagnosis; risk-adjustment
Date of biopsy	COSD CR1010 COSD CR0780	Canisc	Triple assessment in a single visit
Date of mammogram	COSD CR0320 COSD BR4030	Canisc	Triple assessment in a single visit
Date first seen	COSD CR0230	*Not available*	Triple assessment in a single visit
Clinical Nurse Specialist indication code	COSD CR2050	Canisc	Seen by a breast CNS
Receipt of surgery	OPCS codes in HES	OPCS codes in PEDW	Treatment
Receipt of chemotherapy	SACT – Start date within nine months of primary surgical procedure (BCS or mastectomy).	Canisc – Start date within nine months of primary surgical procedure (BCS or mastectomy)	Treatment
Receipt of trastuzumab		*Not available*	Treatment
Receipt of radiotherapy	RTDS – Start date within six months of primary surgical procedure (BCS or mastectomy) OR start date over six months after the primary surgical procedure, BUT only if chemotherapy is given in the interim.	Radiotherapy dataset – Start date within six months of primary surgical procedure (BCS or mastectomy) OR start date over six months after the primary surgical procedure, BUT only if chemotherapy is given in the interim.	Treatment

Item	Where data comes from		Indicator
	England	Wales	
Charlson comorbidity index (CCI)	ICD-10 codes in HES	ICD-10 codes in PEDW	Receipt of surgery by age and patient fitness; risk-adjustment
electronic Frailty Index (eFI)	ICD-10 codes in HES	ICD-10 codes in PEDW	Receipt of surgery by age and patient fitness; risk-adjustment
Deprivation	IMD quintiles from LSOA where not reported as IMD	WIMD quintiles calculated from LSOA in Canisc	Risk adjustment

Table A2_2: Details of the content of figures within the NABCOP 2019 Annual Report

Section	Figure Number	Denominator	Observed or risk-adjusted
4.1	4.2	Women diagnosed in 2017.	Observed number of women.
4.1	4.3	All women.	Observed % of women.
5.1	5.1	All women.	Observed % of women.
5.2	5.2	Women. ICD-10 code C50. Method of presentation not screening.	Observed % of women.
	5.3		Strict = Observed % of women; Relaxed = Observed % of women.
5.3	5.4	All women with data on CNS contact reported.	Observed % of women.
	5.5		Observed % of women.
6.1	6.1	Women. ICD-10 code D05.	Risk-adjusted % of women. Logistic regression models adjusted for age, age^2, CCI, eFI, (W)IMD, method of presentation. Only units with >10 women diagnosed with DCIS breast cancer included.
	6.2	Women. ICD-10 code D05. No surgery.	Observed % of women. Kaplan-Meier curves censoring women alive.
7.1	7.1	Women. ICD-10 code C50. Stage 1-3A.	Observed % of women.
	7.2	Women. ICD-10 code C50. Stage 1-3A. ER status reported.	Risk-adjusted % of women. Random effects logistic regression models adjusted for age, age^2, overall stage, invasive grade, nodes positive, HER2 status, ER status, CCI, eFI, (W)IMD, method of presentation.
	7.3	Women. ICD-10 code C50. Stage 1-3A. ER status reported. No surgery.	Observed % of women. Kaplan-Meier curves censoring women alive.
7.2	7.4	Women. ICD-10 code C50. Stage 1-3A. Surgery within 6m of diagnosis.	Observed % of women.
	7.5		Observed % of women.
	7.6		Observed % of women.
7.3	7.7	Women. ICD-10 code C50. Stage 1-3A. Surgery within 6m of diagnosis. No prior chemo. ER status reported.	Observed % of women.
	7.8	Women. Diagnosed in England. ICD-10 code C50. Stage 1-3A. Surgery within 6m of diagnosis. No prior chemo. HER2-positive.	Observed % of women.
	7.9	Women. Diagnosed in England. ICD-10 code C50. Stage 1-3A. Surgery within 6m of diagnosis. No prior chemo. HER2-positive. CCI calculated.	Observed % of women.
	7.10	Women. Diagnosed in England. ICD-10 code C50. Stage 1-3A. Surgery within 6m of diagnosis. No prior chemo. HER2-positive.	Risk-adjusted % of women. Random effects logistic regression models adjusted for age, tumour stage, nodal stage, ER status, invasive grade, CCI, IMD.
8.1	8.1	Women, ICD-10 code C50, Stage 4.	Risk-adjusted % of women. Random effects logistic regression models adjusted for age, age^2, tumour stage, nodal stage, HER2 status, ER status, invasive grade, CCI, eFI, (W)IMD, method of presentation. Only units with >10 women diagnosed with DCIS breast cancer included.
	8.2	Women, ICD-10 code C50, Stage 4. ER status reported. CCI/eFI calculated.	Observed % of women.
	8.3	Women, ICD-10 code C50, Stage 4.	Observed % of women. Kaplan-Meier curves censoring women alive.

Appendix 3

Details of CPES questions used for analysis of patient experience.

12. Before your cancer treatment started, were your treatment options explained to you?

1. ☐ Yes, completely
2. ☐ Yes, to some extent
3. ☐ No
4. ☐ There was only one type of treatment that was suitable for me
5. ☐ Don't know / can't remember

16. Were you involved as much as you wanted to be in decisions about your care and treatment?

1. ☐ Yes, definitely
2. ☐ Yes, to some extent
3. ☐ No, but I would like to have been more involved
4. ☐ Don't know / can't remember

17. Were you given the name of a Clinical Nurse Specialist who would support you through your treatment?

1. ☐ Yes → **Go to Q18**
2. ☐ No → **Go to Q20**
3. ☐ Don't know / can't remember → **Go to Q20**

18. How easy or difficult has it been for you to contact your Clinical Nurse Specialist?

1. ☐ Very easy
2. ☐ Quite easy
3. ☐ Neither easy nor difficult
4. ☐ Quite difficult
5. ☐ Very difficult
6. ☐ I have not tried to contact my Clinical Nurse Specialist

45. Once you started your treatment, were you given enough information about whether your radiotherapy was working in a way you could understand?

1. ☐ Yes, completely
2. ☐ Yes, to some extent
3. ☐ No
4. ☐ It is too early to know if my radiotherapy is working
5. ☐ I did not need any information

48. Once you started your treatment, were you given enough information about whether your chemotherapy was working in a way you could understand?

1. ☐ Yes, completely
2. ☐ Yes, to some extent
3. ☐ No
4. ☐ It is too early to know if my chemotherapy is working
5. ☐ I did not need any information

59. Overall, how would you rate your care? (**Please circle a number**)