Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.

MELISSA RUTH GANNON

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

OCTOBER 2023

Department of Health Services Research and Policy
Faculty of Public Health and Policy
London School of Hygiene and Tropical Medicine

No funding was received

Declaration

I Melissa Ruth Gannon, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Name in Full: Melissa Ruth Gannon

Student ID Number: P1805830

Date: 20/10/2023

Acknowledgements

I would firstly like to thank my supervisors, Professor David Cromwell and Dr Min Hae Park, for their guidance, advice and support throughout this time of development for me undertaking and completing the PhD.

The work presented in this thesis was embedded within the National Audit of Breast Cancer in Older Patients (NABCOP) and would not have been possible without the support of all members of the Project Team over the years (Professor Kieran Horgan, Dr David Dodwell, Ms Yasmin Jauhari, Ms Katie Miller, Ms Jibby Medina and Ms Karen Clements). Thank you each for the part you played in my journey into the world of breast cancer and routine data. I am incredibly grateful for all the reviews of draft manuscripts, particularly over this past year, and for sticking with me as I aimed high with the journal submissions. Additionally, I want to acknowledge Dr Ajay Aggarwal who inspired and contributed to work looking at the use of NICE-approved drugs in practice.

I am particularly grateful to Dr David Dodwell for his valued wisdom and oncological experience, along with his support and encouragement over the years. His recognition of my learnings and experience with the routine data so central to this thesis has provided the much-needed uplifting during those more difficult times.

Finally, I am incredibly thankful to my close colleagues, Jibby Medina and Katie Miller, who provided me with the daily support needed to undertake the endeavours of a PhD alongside the ever increasing and variable pressures of conducting a national audit. You have been with me through the highs and the lows, moves to home-working, and all that COVID and these “post-pandemic” years have thrown at me. You have both most definitely kept me sane and for this I will be forever thankful.

Abstract

Background

Evidence on oncological treatment efficacy from randomised controlled trials (RCTs) forms the backbone to national guidance on clinical practice but their findings may have limited generalisability. This is a particular concern when applying guidance about treatments to older patient populations who can be poorly represented within RCTs.

This thesis examined the value of routinely-collected data sources in evaluating oncological treatments for patients with invasive breast cancer (IBC).

Methods

Cancer Registration records for women aged 50+ years newly-diagnosed with IBC in England from 2014-2019 were used, linked at patient/tumour-level to routine national datasets providing information on patient and tumour characteristics, treatment and survival outcomes.

Initial work examined methodological challenges in understanding oncological treatments using routine data sources. Work then investigated clinical aspects of oncological treatments received in practice, focusing on their uptake, and the safety and benefit of trastuzumab-based treatment for HER2-positive early invasive breast cancer (EIBC) among older, less fit patients.

Results

There were systematic differences in oncological treatment recording within available national data sources, with records less complete for older patients. For endocrine therapy, completeness was excellent in primary care data but linkage to secondary care data identified initial hospital-based prescriptions, providing more comprehensive information about treatment timings.

Use of oncological treatments was consistently lowest among older women, independent of other relevant factors. Trastuzumab-based treatment for HER2-positive EIBC had comparable overall safety among both younger and older women, although increasing age was associated with increased odds of cardiovascular problems. There was no evidence of differences in the effect of trastuzumab on survival by age.

Conclusions

This research has demonstrated the value of routine national data in understanding treatment use and associated outcomes among patients with IBC treated in routine care. Similar studies may address evidence gaps in other treatment areas where patient representation in RCTs is an issue.

List of Abbreviations and Acronyms

ABS	Association of Breast Surgery
APC	Admitted Patient Care
BC	Breast cancer
BCSS	Breast cancer-specific survival
CAS	Cancer Analysis System
CCI	Charlson Comorbidity Index
COSD	Cancer Outcomes and Services Dataset
HER2	Human epidermal growth factor receptor 2
HES	Hospital Episode Statistics
ICD-10	International Classification of Diseases, 10th Revision
IMD	Index of Multiple Deprivation
ITT	Intention to treat
MDT	Multidisciplinary team
NABCOP	National Audit of Breast Cancer in Older Patients
NCRAS	National Cancer Registration and Analysis Service
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OS	Overall Survival
RCS	Royal College of Surgeons of England
RCT	Randomised controlled trial
RTDS	National radiotherapy dataset
SACT	Systemic anti-cancer therapy
SCARF Index	Secondary Care Administrative Records Frailty Index
UK	United Kingdom

Contents

1. Introduction

1.1. Background

Oncological therapies are widely used for the treatment of invasive breast cancer. The majority of published evidence on the efficacy of these treatments is from randomised controlled trials (RCTs), which have applied strict inclusion or exclusion criteria, sometimes with an upper age limit. With the resulting trial populations differing from the general population of patients, RCT findings risk not being generalisable outside of a trial setting, to the "real-world" patient population.[1] Previous work in this area has demonstrated that older, less fit patients are the most likely to be poorly represented within RCTs, despite being the largest patient group diagnosed with breast cancer; common reasons for this include protocols excluding patients based on their pattern of comorbidity and physician attributes having an impact on recruitment and subsequent trial entry among older patient groups.[2-10]

The poor representation of older patients in RCTs means there is subsequently limited evidence on treatment effectiveness for this patient population.[11,12] Clinical guidelines in older, more frail patients (with breast cancer) as such are scarce, leading to a lack of consensus in what treatment pathway to follow.[13] Deciding on the appropriate treatment options for older patients often requires the treating clinician to extend RCT findings from a younger, fitter trial population, resulting in variation in the management of older patient groups.[14] Previous work has highlighted such variation in the treatment received by older patients with breast cancer in routine care.[15-23] A 2012 Department of Health report looking at the impact of age on decision-making for cancer treatment found that "the absence of clinical trial data on the efficacy of treatments in older patients" was an influential factor in the challenges highlighted in treating older patients with cancer.[24]

Along with a need for understanding the benefit and risk of treatments in patient groups that are underrepresented in RCTs, there is also a need for evidence on the uptake and "real-world" effectiveness of oncological treatments for breast cancer.[25,26] Routinely-collected data present a potentially valuable resource for studies to address these evidence gaps.[27]

The research presented within this thesis examines the value of routinely-collected data in the evaluation of systemic oncological treatments for patients newly-diagnosed with invasive breast cancer. The work approached this aim from two complementary perspectives. The first was to examine the recording of systemic oncological treatment information within routinely-collected national data sources, with a view to understanding the strengths and limitations of the datasets for evaluation and to ensure the information is accurate. The second was to evaluate the uptake of systemic oncological treatments, as recorded in the routine data, and look at associated clinical outcomes in terms of safety and effectiveness.

The work within the thesis includes two methodological studies and four clinical studies, each of which used routine national data collected by the National Health Service (NHS) as part of their process of care and support. One of the main sources of treatment data was the Systemic Anti-Cancer Therapy (SACT) dataset; this is a

national dataset of systemic oncological treatment, which is designed to have whole population coverage in England and contains data items capturing drug name and administration date.[28] This was the primary source of data on systemic oncological treatments used within this thesis and it provides comprehensive and detailed information on a national scale.

The next sections within this Introduction provide an overview of invasive breast cancer in women and its treatment, an overview of the national routine data collection, including the data sources used within the work, and an overview of the evidence generation process and key considerations for addressing this with an observational design using routine data.

1.2. Overview of breast cancer in women

1.2.1. Epidemiology

Breast cancer occurs where abnormal cells in the breast grow and divide in an uncontrolled manner, eventually forming a growth, known as a tumour. Breast cancer is typically grouped into two main types: non-invasive or invasive. Invasive breast cancer is where the cancer cells have spread beyond the lining of the ducts into the surrounding breast tissue. Based on the size of the primary tumour (tumour stage), the extent to which the cancer has spread to lymph nodes outside the breast (nodal stage) and if it has spread to other parts of the body (metastatic stage), invasive breast cancer is given a stage from 1 to 4 (Table 1).

Table 1: TNM stage groupings

Stage grouping	Overall stage	Tumour (T) stage	Nodal (N) stage	Metastasis (M) stage
Ductal carcinoma in situ	Stage 0	Tis	N0	M0
Early invasive breast cancer	Stage 1A	T1	N0	M0
	Stage 1B	T0 / T1	N1(mi)	M0
	Stage 2A	T0 / T1 T2	N1 N0	M0
	Stage 2B	T2 T3	N1 N0	M0
	Stage 3A	T0,T1,T2 T3	N2 N1,N2	M0
Locally advanced breast cancer	Stage 3B	T4	N0, N1, N2	M0
	Stage 3C	Any T	N3	M0
Metastatic (advanced) breast cancer	Stage 4	Any T	Any N	M1

KEY:
Tumour stage (based on tumour size): Tis = in situ (cancer confined within the breast tissue ducts); T0 = no evidence of cancer in the breast; T1 = 1-20mm; T2 = 21-50mm; T3 = 51+ mm; T4 = tumour spread to skin or chest wall.
Nodal stage: N0 = No cancer cells in lymph nodes; N1, N2, N3 increasing spread of cancer within the lymphatic system.
mi = micrometastases

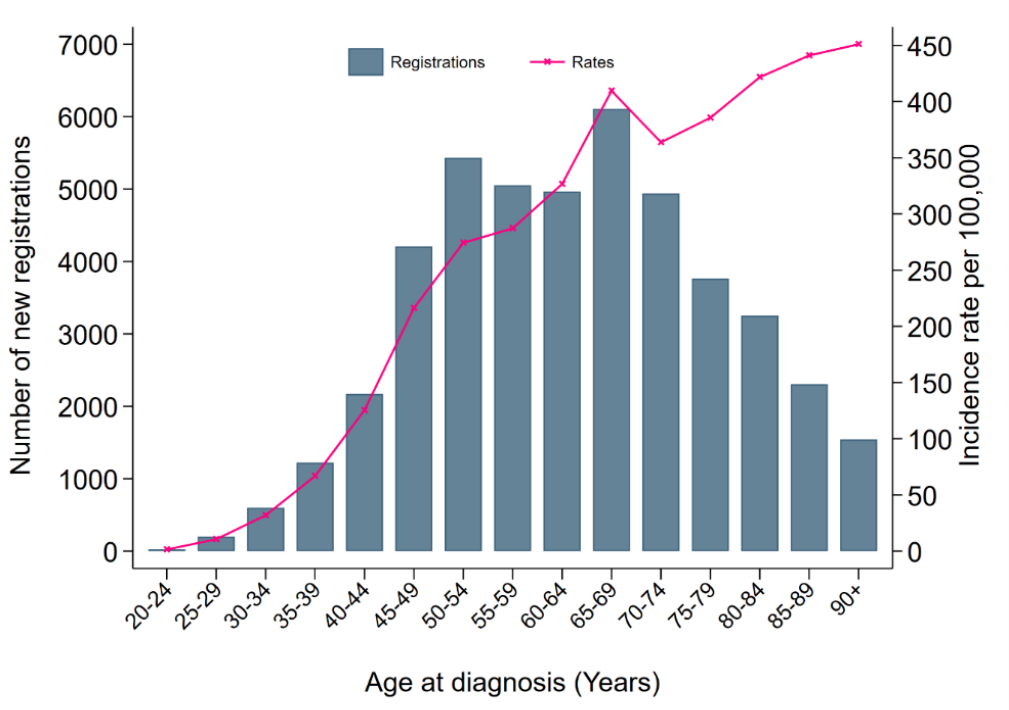
Breast cancer is categorised as early invasive breast cancer (EIBC; stage 1-3A) if the primary tumour is at least 1mm in size or the cancer has spread to the lymph nodes under the armpits but no further. EIBC is the most

commonly diagnosed breast cancer group, accounting for around 80% of women aged 50 years and over diagnosed with invasive breast cancer in England and Wales.[29] Locally advanced breast cancer (LABC; stage 3B-C) is characterised by spread of the tumour to the skin or chest wall or high involvement of local lymph nodes. Breast cancer which has spread to other parts of the body is defined as metastatic breast cancer (MBC; stage 4), also referred to as advanced breast cancer.

In addition to staging, the breast cancer cells of the primary tumour are tested to identify receptors for particular hormones or targeted drug therapy. Breast cancer cells which group in response to the hormones estrogen or progesterone are classed as hormone receptor-positive and are suitable for treatment with hormone (endocrine) therapy.[30] Breast cancer cells with higher levels of human epidermal growth factor receptor 2 (HER2) are classed as HER2-positive; they are more aggressive, may grow more quickly and are suitable for anti-HER2 therapies such as trastuzumab.[31] All of these characteristics of the primary tumour, along with if and to where it has spread in the body, patient factors (including comorbidity and fitness) and preference, inform treatment decisions for a patient.[32]

Within the United Kingdom (UK), breast cancer is the most common type of cancer among women, with around 55,000 new cases each year, and accounts for nearly a third of cancers diagnosed in women.[33,34] Among women, breast cancer is the second most common cause of cancer death, with nearly 12,000 deaths annually.[35] Incidence rates increase considerably with age, with a third of new cases occurring in women aged 70 years and over (Figure 1).[36] The ageing UK population means that the absolute number of cases in this subgroup of the population is projected to rise.[37,38]

Figure 1: Invasive breast cancer (C50), number of new cases and age-specific incidence rates per 100,000 women, UK, 2017



Note: Data used in this figure are from the Cancer Registration Statistics, England, 2017

1.2.2. Treatment for invasive breast cancer

For patients newly-diagnosed with invasive breast cancer, there are multiple treatment options with varying clinical outcomes and safety profiles.[39,40] Treatment options for invasive breast cancer are determined by the characteristics of the tumour, including the size and grade, hormone receptor and HER2 status, as well as nodal involvement and metastatic status.[41]

EIBC is potentially curable and requires treatment of the primary tumour, along with treatment of any nodal involvement as appropriate.[42] Current options include surgery, chemotherapy, radiotherapy, endocrine therapy, biological therapy.[43] For EIBC, surgical treatment is the main treatment option. Additional treatments may be provided before (neoadjuvant) or after (adjuvant) the surgical procedure. Using a combination of treatments, known as multimodal treatment, is common and includes chemotherapy around surgery, and radiotherapy, where it is indicated. There are two broad categories of chemotherapy drugs known as (i) taxanes and (ii) anthracyclines.[44,45] Within the EIBC setting a combination of chemotherapy drugs is frequently used.[46-48] Taxanes used for treating breast cancer are docetaxel (brand name Taxotere) and paclitaxel (also called Taxol). Adding a taxane drug to standard chemotherapy has shown improved survival, but is associated with an increased risk of serious adverse events (side effects/toxicity) including febrile neutropenia and neuropathy, and so an assessment of the balance of benefit and risks is required.[49,50]

LABC is the most advanced breast cancer without metastases and is associated with increased risk of the cancer returning (recurrence).[51] Treatment options for LABC are similar to those for EIBC, but in most cases, if the cancer has spread surgery is not an option.[43]

Treatment options for MBC are more restricted because the cancer has spread beyond the breast, to other areas of the body. Treatments are typically given to try and control further growth and help manage symptoms of the cancer. Surgery is not a recommended treatment option for MBC however, chemotherapy, endocrine therapy, targeted biological therapy or immunotherapy may be suitable.[52]

For HER2-positive invasive breast cancer there is the further option to use a HER2-targeting therapy, as part of treatment. The first HER2-targeting therapy available for use in routine clinical practice was trastuzumab (brand name Herceptin), which was approved by the National Institute for Health and Care Excellence (NICE) from 2002 (for MBC) and 2006 (for EIBC and LABC).[53,54] The aim of using trastuzumab is to reduce the likelihood of recurrence or of the cancer getting worse (progressing), and meta-analyses of RCTs have shown that treatment with trastuzumab (in addition to chemotherapy) considerably improves survival outcomes.[55,56] Indeed, time to recurrence or progression has been shown to be reduced by one-third among women receiving trastuzumab compared to those who were not. Considering the balance of benefits and risks however, use of trastuzumab has been found to be associated with a three to five-fold increase in risk of severe heart problems, including congestive heart failure.

1.3. National routinely collected data on patients with breast cancer

Within the UK, individuals with a confirmed diagnosis of breast cancer are registered with the national cancer registration service in each country. Most registrations are for patients diagnosed and subsequently treated within the NHS but they may also cover people whose breast cancer was only identified from the death certificate. As well as cancer registration details, data on all aspects of care and treatment for patients diagnosed with breast cancer are routinely collected in various national healthcare datasets. For patients diagnosed and treated by NHS providers, these data provide information on the process of diagnosis, staging, and treatment as well as capturing information useful to understand clinical outcomes including complications, serious adverse events (SAEs) and death. Within England, these datasets are held by the National Cancer Registration and Analysis Service (NCRAS) at NHS England (previously Public Health England and then NHS Digital).

Data on cancer diagnosis and staging, including pathology details, for patients in England are routinely collected within the Cancer Outcomes and Services Dataset (COSD).[57] This information is also used in the cancer registration process. COSD and Cancer Registry data have the capability of providing in-depth information on patients diagnosed with breast cancer and treated within routine practice.

Data on the use of systemic oncological treatments delivered within English secondary and tertiary care settings are routinely collected within the SACT dataset.[28] Submission of data to the SACT dataset is mandatory for all NHS trusts providing systemic oncological treatments. Treatments recorded within the SACT dataset include chemotherapy (standard and oral), endocrine therapy and biological therapy. The SACT dataset captures data for each treatment administration including the date, regimen, and individual drug details (dose, route of administration). As such, the SACT dataset is a unique resource with the capability of providing in-depth information on the use of systemic oncological treatment regimens within routine practice.

Along with the SACT dataset, there are multiple datasets related to hospital-based care and outcomes. In England, the Hospital Episode Statistics Admitted Patient Care (HES-APC) dataset collects hospital administrative data for all day case and overnight English NHS hospital admissions.[58] With many treatments for breast cancer administered as a day case hospital admission, along with subsequent management of any treatment-related SAEs, HES-APC has the potential to provide information on these aspects of care.

Data on medicines dispensed in a primary care setting, for cancer or otherwise, form the Primary Care Prescription Database (PCPD).[59,60] For patients with breast cancer, treatment recorded within the PCPD includes endocrine therapy. The PCPD captures information on each time treatment is dispensed including date (month and year), drug name and dose.

These datasets have national coverage of the NHS in England and so provide a unique opportunity to understand ”real-world” clinical practice, patterns of care and outcomes. The information provided within these routine healthcare datasets creates the potential for evaluation of national oncological treatment patterns and associated issues and outcomes.

1.4. Evidence of clinical effectiveness of breast cancer treatments

Clinicians and patients are supported in making treatment decisions for breast cancer by available national clinical guidelines, which are typically based on evidence generated by RCTs. The following subsections describe the causal inference framework that underpins the process of generating evidence of effectiveness and the various statistical methods by which estimates of treatment effect size can be made. Many discussions of comparative effectiveness research focus solely on these statistical methods. However, this is to ignore the preliminary work that might be required to transform the raw data items in routine datasets into the variables required for the analysis. These aspects are considered in the final section.

1.4.1. Generating evidence of effectiveness

Of primary importance in estimating clinical effectiveness is being able to confidently estimate the direct effect of a defined exposure (for example, systemic oncological treatment) on a defined outcome (for example, overall survival) rather than simply concluding there is a statistical association; that is, we are interested in the causal effect between an exposure and subsequent outcome.[61]

Within the mathematical world, we can estimate the impact of exposure on outcome for each individual (i) under the scenario of them being exposed and (ii) under the scenario of them not being exposed. We can define the exposure of interest as "A" and then we give A the value 0 if the individual is not exposed and the value 1 if the individual is exposed. If we then define the outcome of interest as "Y", we can give Y the value 0 if the individual does not experience the outcome and the value 1 if the individual does experience the outcome. We then refer to the annotation $Y_{a=0}$ as the outcome if the individual is not exposed and $Y_{a=1}$ as the outcome if the individual is exposed. Given that everything else remains the same, an individual causal effect exists if the outcome under exposure is not equal to the outcome under no exposure (or mathematically, $Y_{a=0} \neq Y_{a=1}$), that is the outcome differs according to whether the individual is exposed or not.

Clearly the individual cannot be both exposed and unexposed at the same time, and so $Y_{a=0}$ and $Y_{a=1}$ are known as counterfactual outcomes. At a population level, we calculate the proportion of individuals with outcome Y under each exposure scenario as being $\Pr[Y_a = 1]$. Following the same principle as for an individual causal effect, a population-level causal effect exists if:

$$\Pr[Y_{a=0} = 1] \neq \Pr[Y_{a=1} = 1]$$

This mathematical formulation is known as the *Causal Inference Framework*.

The framework allows a clear distinction between a causal effect and an associational effect. We note that the causal effect is estimated in the whole population of individuals observed under both levels of the exposure. In contrast, an associational effect is estimated by observing the outcome in the subset of individuals who are exposed and comparing this with the observed outcome in the subset of individuals who aren't exposed i.e., $\Pr[Y_a = 1 \mid A = 1]$. This is called a conditional probability because observing the outcome is dependent on

observing the exposure status. Conversely, the population causal effect is an unconditional, or marginal, probability. Effect measures can be calculated which also provide a measure of the strength of a causal effect where it exists, including the causal risk difference, risk ratio and odds ratio.

Study designs for evidence generation fall within the categories of experimental or observational, based on whether the exposure being studied is assigned or not.[62] In general, experimental designs allow researchers to have greater control over the research design and therefore ensure the study has adequate internal and external validity. Internal validity is the extent to which the observed causal relationship between exposure and outcome is uninfluenced by unmeasured factors, whilst external validity is the extent to which study results can be generalised to other settings.[63] Experimental studies can be randomised or non-randomised.

1.4.2. Using randomised controlled trials to generate evidence

The traditional hierarchy of evidence used within medical research often considers RCTs to be the 'gold standard' study design for determining the efficacy of an intervention such as systemic oncological treatment, with randomised studies most commonly used in generating robust evidence of causal effects.[64,65]

Design and analysis of randomised controlled trials

An RCT is a comparative, controlled experiment, carried out prospectively, whereby within the context of evaluating the effect of a treatment for a specific disease patients with the selected disease are randomly allocated to one of two or more treatment groups (referred to as treatment "arms").[66] There is always a control group, within which patients are allocated either to the current standard of care or to placebo, depending on which is most appropriate. The other group(s) receives a novel intervention (or combination of interventions) which the investigators are interested in measuring the effect of. Random allocation to each group helps ensure any factors which might influence the outcome are balanced across the groups, so that estimation of effect is not unduly influenced by these.

In some instances, it is ethical to "blind" participants to their allocated treatment group, to minimise the influence of this knowledge; knowledge which could bias measurements because of an expectation as to how the intervention will perform. There is greater need for blinding where the outcome of interest is subjectively measured and therefore might be unduly influenced by knowing the intervention assignment. In typical trials, where blinding is not used the random allocation of a patient should not be revealed (allocation concealment) until they have consented to be in the trial, to minimise selection bias. A meta-epidemiological study published in 2008 demonstrated that where trials had subjective outcomes estimates of effect were exaggerated in those situations where there was (i) inadequate or unclear allocation concealment or (ii) lack of blinding.[67] By contrast, there was little evidence of bias in trials with objective outcomes such as overall survival (OS) due to these two methodological sources of bias.

Upon randomisation each group of patients is followed up, at clearly defined intervals, until some pre-specified, measured end point. Relating this to the causal inference framework, previously described, the expected

outcome Y(A) and observed outcome Y are defined over a specified time interval. With random treatment allocation we can directly compare those individuals receiving treatment with those not. The average treatment effect (ATE) on the outcome can then be estimated as equal to the difference in expected outcomes based on treatment or not:

$$\begin{aligned} ATE &= E\big(Y(1) - \ Y(0)\big) \\ &= E\big(Y(1)\big) - \ E\big(\ Y(0)\big) \\ &= E(Y|A=1) - \ E(Y|A=0) \end{aligned}$$

With the random allocation of treatment we consider the two groups to be exchangeable, so that it is irrelevant which group got the intervention for estimating the value either of $\Pr[Y = 1 \mid A = 1] \ or \ \Pr[Y = 1 \mid A = 0]$, and the associational risk is the same as the counterfactual risk.[61]

The approach to analysis of RCTs influences the interpretation of their findings. The primary method of analysis in many RCTs is an intention to treat (ITT) analysis, with patients analysed within their randomised group irrespective of adherence to allocated treatment. This form of analysis retains the balance of prognostic factors provided by the original random allocation. With an ITT analysis, RCTs are referred to as taking a pragmatic approach, recognising that in routine clinical practice patient preference and tolerance to treatment are factors in adherence to treatment, and (an oncological) treatment may be changed based on how a patient's condition responds to it. An ITT analysis provides an estimated treatment effect largely in line with what we would expect to see overall in the population. By contrast, a per-protocol analysis considers patients in groups according to the treatment they received.

Strengths and limitations of RCTs

RCTs offer several methodological strengths. Randomisation reduces the risk of bias from confounding, wherein either an effect of treatment on the outcome is found where no treatment effect really exists, or conversely a true treatment effect is masked due to the relationship of both treatment and outcome with some other variable(s). In a well-designed RCT, the randomisation process should produce treatment arms balanced with respect to observed factors considered to influence the outcome(s) of interest. This process should also mean that unobserved factors, which might impact the outcome(s) of interest, are balanced across treatment arms. This increases our confidence that any observed difference in outcome(s) across treatment arms is solely due to the intervention offered.

While well-run RCTs have excellent internal validity, a strong criticism can be poor external validity, with limited relevance to patient groups beyond the trial setting.[68] Trial populations may differ from the general population of patients with breast cancer due to two main aspects of trial conduct: (i) trial patients are selected based on strict inclusion/exclusion criteria; (ii) enrolled patients receive more attentive follow-up than would be present in routine care, often receiving life-long care within the hospital research unit. Although these aspects aim to ensure the trial population remains homogenous and sufficiently monitored to measure any treatment benefit,

selection and care of patients in RCTs often differs from practice in routine care, meaning applying the results of RCTs to subpopulations of underrepresented patients, such as older patients, is often unclear.[69,70] Older patients are often excluded from trial protocols, based on their chronological age, or because as an older person they are more likely to present with poor functional status or with comorbid conditions which may be contraindications for the treatment under investigation.[6] Conversely, where age is not a specific protocol criterion the perception of the recruiting clinician may influence who is offered the trial.[8] Some examples of this within breast cancer are seen in relation to the age profile of patients within RCTs looking at treatments for early breast cancer: specifically (i) only eight of 29 RCTs included within the Cochrane review of taxanes for early breast cancer clearly included patients older than 70 years, for 18 RCTs patients older than 70 were explicitly not eligible, and (ii) none of the six RCTs of adjuvant trastuzumab-based treatment in the Cochrane review included patients older than 80 years, whilst patients older than 70 were not eligible for half.[50,55] Generalisability of trial results to the patient population within which treatment is ultimately to be used plays an important role in ensuring results translate into utilisation in clinical practice. A large part of this lies in ensuring the trial population is representative of the general patient population.

To this point, although we benefit from estimating a treatment's causal effect from an RCT, application of this evidence to the real world requires an understanding of treatment effectiveness, that is its benefit outside of a strict recruitment/highly monitored trial setting.[71-73]

Improving generalisability of RCT findings

Poor representation of older patients within RCTs has been a long-standing issue.[2,74] However, efforts to conduct trials among older patients with breast cancer have generally struggled to recruit participants.[75,76] The ACTION study, of adjuvant chemotherapy, reported that many eligible elderly patients who were approached for inclusion declined randomisation, largely due to not wanting to receive chemotherapy.[75] A case-control study investigating the impact of age on being offered enrolment and participating in trials reported older patients with breast cancer were offered trials less often than younger patients.[3] Age, comorbid conditions and stage were factors found to be independently predictive of whether a patient was offered a trial. They found no difference in participation by age among those offered a trial. Several systematic literature reviews looking at recruitment of older people into clinical trials, found protocol exclusions of comorbidities were commonly highlighted as barriers to recruitment of older patients, whilst the perceptions of physicians (or family) were important reasons in influencing older patients declining trial entry.[4,5,8 77] More recently, in a UK setting, the Age Gap Trial reported difficulties in the recruitment and retention of older patients even with "age-friendly study design measures" in place.[78]

With an underrepresentation of older patients or those with some level of comorbidity or frailty participating in trials, clinical guidelines typically do not provide information for clinicians on how treatment recommendations might be adapted for these patient populations. In this scenario, deciding on optimal treatment for these patients often requires clinicians to extrapolate the results of trials in younger, healthier patients to an older population with often multiple comorbidities.[14] This extrapolation potentially leads to a lack of consensus in

treatment selection resulting in inappropriate variation in the provision of care. Previous research looking at receipt of treatment among patients with breast cancer has found high variation in use by patient age, with older patients typically less likely to receive standard care.[15-18] In the case of adjuvant systemic therapy, examples have been found in relation to variation in the use of chemotherapy and HER2-targeting therapy, both in the National Audit of Breast Cancer in Older Patients (NABCOP) Annual Reports as well as in previous literature considering patient populations within and beyond the UK.[15,19,29,79,80] Notably the NABCOP found use of adjuvant chemotherapy and trastuzumab for HER2-positive EIBC decreased from 70% among women aged 50-69 years to 9% among women aged 80+, with similar patterns observed for adjuvant chemotherapy (in all women with EIBC and for ER-negative, HER2-negative, node-positive EIBC).[79] A study of women in the United States (US) treated across six community-based settings found women aged 75 years or older and those with higher comorbidity burden were less likely to receive standard treatment, including chemotherapy.[15] Another US study using routine data found less than half of women aged 66 years and older received guideline concordant care for EIBC, with lowest rates observed among the oldest age groups.[19] Similar findings are reported for the use of adjuvant radiotherapy, with rates decreasing with age.[17]

1.4.3. Using observational studies to generate evidence

Compared with RCTs, observational studies are widely considered lower ranking in providing evidence of treatment effect because of their potentially weaker internal validity. Consequently, an observational study requires a robust design and analysis. For studies looking to evaluate treatment effectiveness, accounting for potential systematic differences in the treatment groups, either in design or analysis, is crucial in minimising the bias within estimates of treatment effect. There are ways by which observational studies can be designed to emulate RCTs and provide causal estimates of treatment effects.[81-84] There has been a particular focus on statistical techniques for removing the impact of confounders. Among observational studies estimating treatment effects, propensity scores and covariate adjustment in regression analysis are among the most widely used methods to account for confounding and non-random assignment to treatment.[85-87] The principles of these and other methods are described below.

Propensity scores

First introduced in the 1980's propensity score (PS) methods aim to balance the distribution of covariates across treatment groups and involve using a PS to balance covariates predicting treatment receipt.[88] Scores are based on observed predictors for individual patients and are used to produce treatment groups largely balanced/comparable on observed variables of importance. This was an advance on traditional matching of individuals across groups with respect to shared single characteristics, which often required high numbers of control patients and a need to limit the number of covariates used to define patients as matched.[89] Balancing propensity scores across groups enables estimation of the average treatment effect in the treated group (ATT) and can provide reliable causal effect estimates assuming no unmeasured confounding.

Propensity scores can be used in several ways.[90] These include:

1. Inverse probability of treatment weighting – based on calculated scores a counterfactual sample is created, these are patients with the opposite treatment assignment but assuming the same outcome as those in the original sample. We then have two groups, one in which everyone has the exposure/treatment of interest and one in which no one is exposed.
2. Stratification on score – analysis of treatment effects within strata including patients with the same score; the average causal effect of treatment is calculated based on the weighted average of the within-strata estimates of the exposure's effect (weight for a stratum is equal to the fraction of the sample within that stratum).
3. As part of the adjustment process in carrying out usual statistical analyses – including the PS within the regression model.
4. To match individuals – there are a variety of methods proposed for this, with or without replacement[89,91]
    a. optimal matching (forms pairs based on minimising the within-pair PS difference),
    b. nearest neighbour matching (starts with a treated patient then selects a non-treated patient with the closest PS, picked at random where this applies to multiple patients), and
    c. caliper matching (which is the same as nearest neighbour matching but specifies the maximal distance, caliper distance, with the caveat that restriction on distance means not all subjects will necessarily be matched).

Adjustment in regression analysis

Adjustment of covariates in a regression analysis is a simple approach, although it can lead to increased bias in the estimate of the average treatment effect, especially where covariates are considerably imbalanced across treatment groups.[87,92] This is likely to introduce most bias where imbalanced covariates have a confounding effect and are associated with differences in the outcome of interest. For example, in a breast cancer study looking at the impact of treatment on survival, in the case of groups where the treatment group has proportionally more patients with nodal involvement, which is a factor associated with the treatment choice and is also prognostic of poor survival.

Table 2 provides some detail on the assumptions and strengths of the two statistical methods described above.

Table 2: Proposed statistical methods for analysing non-randomised/observational data to estimate the causal effect of treatment

Method:	Propensity scores (PS)	Regression with adjustment for covariates
Assumptions for comparative effectiveness research	At any value of propensity score each individual with that score has the same probability of receiving treatment (conditional on observed confounders). All confounders are observed/measured.	Adjustment for confounders. All confounders are observed/measured.
Estimation in subgroups	Where treatment assignment mechanisms differ by subgroup, subgroup-specific propensity scores should be estimated to ensure balance and valid treatment effect estimation within subgroups.	Inclusion of interaction term(s) within the regression model.
Handles missing data	Multiple imputation techniques can be used to average the PS across each imputed dataset before creating a matched cohort.[93]	Standard methods including multiple imputation techniques can be used.
Strength	If the outcome is independent of treatment assignment, conditional on measured confounders, we can view those patients receiving treatment as a random sample of all individuals with that PS. Using calculated propensity scores enables observed characteristics to be balanced, on average, across treatment groups. Robust to model misspecification, including mis-modelling of non-linearity.	Straightforward to implement. Allows for adjustment of measured confounders.
Weakness	Produces biased estimates of treatment effect in the presence of unmeasured and residual confounding. May result in a non-representative sample and loss of statistical power if PS matching is used and too many individuals are omitted. Robustness to model misspecification is unlikely to hold for applications of PS to covariate adjustment or inverse probability of treatment weighting as they use the actual value of the PS.	Produces biased estimates of treatment effect when groups are considerably unbalanced, in the presence of unmeasured and residual confounding.

Other methods

Other methods that have applications within comparative effectiveness research, but are less commonly used and so less familiar to a non-statistical audience, include instrumental variables (IV) and entropy balancing.[94,95]

Within IV analysis, estimates of treatment effect can be obtained when a pre-treatment variable (called an instrument) meets the conditions (i) it is associated with treatment (ii) it causes the outcome only through treatment and (iii) its effect on the outcome is not confounded.[96,97] An advantage of IV analysis is that it can provide consistent estimates of the average causal effect even in the presence of unmeasured confounding.[98] IV analysis however, is statistically inefficient compared to other methods, and the effective sample size is reduced when instrument(s) are defined at a higher level than patients (i.e. hospital level) which also alters the interpretation of estimates of treatment effect. An important limitation of this method when considering its use for survival analysis is that it ignores time to event and censoring.[99]

Entropy balancing is another technique to achieve balance across treatment groups with respect to important covariates.[95] Opposite to the process of creating propensity scores for weighting samples and checking balance

based on these, entropy balancing defines balance and then creates weights defined according to these criteria. It is a fairly new methodology and as such less commonly used within the applied literature.

1.4.4. Key considerations for using routine national data to generate "real-world" evidence

The UK body which makes recommendations on the use of oncological treatments, NICE, defines "real-world" data as "data relating to patient health or experience or care delivery collected outside the context of a highly controlled clinical trial" and says it "can be routinely collected during the delivery of health or social care".[100] It is this working definition that will be used when referring to real-world data within the rest of this thesis.

"Real-world" data has been successfully used to look at variation in care and compare NHS organisations within national clinical audits. For breast cancer, the NABCOP was an example of this, demonstrating how "real-world" administrative data can be used to provide information on the treatment and care provided to patients with cancer and encourage quality improvement both locally and nationally. Using routinely collected data for service evaluation and research, where these data are well completed and reflective of care, has the potential to be incredibly beneficial as (i) it does not involve the resources associated with further specific data collection and (ii) can be readily available for use, therefore providing timely information without contributing additional burden to healthcare providers or researchers. With data being collected as part of the care and management of patients, it is also less likely to come with the response bias associated with data entry where collection is specifically for evaluation purposes.

Studies using routinely collected national data have the benefit of (i) providing information on effects of treatment across patient subgroups, such as those based on age and fitness, where RCTs have proven to be infeasible and (ii) evaluating use of treatments in routine care, along with their associated safety and subsequent survival.[101] By their nature, the results of studies using routine national data would be generalisable to the "real-world" patient population. This is of importance when considering evaluation of treatments for older, less fit patient groups, most often not included in clinical trials.

When using any data, it is important to understand the primary reason for their collection and the characteristics of the information collected. Specifically, with routine data primarily collected for administrative purposes its use for service evaluation, understanding treatment utilisation and associated effectiveness, forms secondary use and as such there are a variety of potential limitations to consider.

There are several conditions which need to be satisfied when trying to estimate treatment effectiveness that have direct implications when using information derived from routine national datasets. The research presented in this thesis explored some of these key conditions as part of the work of demonstrating how existing routine data can be used for service evaluation in the breast cancer setting.

Firstly, it is important to understand the completeness of key information within the data, to identify/understand potential biases, to determine whether these are sufficiently complete for carrying out comparative effectiveness research (CER), and whether there are other data sources which might either enable

evaluation of this or contribute to improving completeness. This applies both to data on treatment as well as data on key confounders / explanatory variables. Secondly, understanding patterns of treatment utilisation are required both to help design the CER study and also to highlight what real-world patterns of care look like compared with those in RCTs. Finally, in order for a CER study to provide a robust evaluation of treatment effect, it is important for there to be sufficient uptake of the treatment of interest. For this, not only are good levels of treatment data completeness required but also a relatively steady state of use in order for a description of "real-world" utilisation to be considered reliable.

Understanding the completeness and accuracy of data on variables such as the characteristics of the tumour forms part of the diagnosis work up of all patients and so estimation of completeness is straightforward. This information forms part of the cancer registration process, for which there are validation steps taken by the cancer registration service to minimise coding errors. By contrast, information on treatment requires more investigation to understand how complete and accurate this is in within routine data sources. The SACT dataset forms the primary source of data on oncological treatment and previous work in colon cancer and lung cancer has both highlighted patient groups with lower than expected data completeness within this data source and also demonstrated how hospital administrative data can be used to identify information on chemotherapy, in those instances where the data collection in this treatment-specific data source may be incomplete.[102,103] Comparisons of information captured across data sources can be an effective way of understanding data completeness, and for oncological treatment hospital admissions data is one such data source. Within hospital admissions data standardised coding is used for each admission with details of treatments given, coded based on the Office of Population Censuses and Surveys (OPCS) procedure codes.

There may be aspects of treatment use which are either poorly captured or not directly captured within routine data. Within breast cancer, patient fitness and treatment toxicity are two such examples. For both aspects, studies have developed and validated coding frameworks using (diagnostic and procedural) codes within hospital administrative data to allow the identification of patient fitness (comorbidity and frailty) and also chemotherapy-related toxicity.[104-107]

2. Aims and Objectives

2.1. Overview

This research was designed to address the need for "real-world" evidence on the use, effectiveness and safety of systemic oncological treatments for patient populations under-represented in clinical trials.

The work used the available prospectively-collected routine national data to investigate use of systemic oncological treatments in patients with invasive breast cancer and identify groups of patients underrepresented in RCTs. In doing this, my PhD investigated the strengths and limitations of the available routinely collected treatment data, and explored ways of augmenting this information from other routine data sources. In addition to this, the research focused on the subgroups of patients with invasive breast cancer where evidence of treatment effectiveness was lacking because they were not represented in the clinical trials. In particular, older patients and those with comorbidity or frailty.

The output from this research has been written up in high quality, peer-reviewed journal publications, with further dissemination at relevant clinical conferences.

2.2. Aim

The broad aim of the thesis was to examine the value of routine healthcare data in the evaluation of systemic oncological treatments for newly-diagnosed invasive breast cancer. To fulfil this aim, the studies presented within this thesis addressed two perspectives. The first was to evaluate the completeness and concordance of treatment information within routine healthcare data, by comparing details recorded in multiple data sources. The second part of the research focused on the clinical application of observational methods to (i) evaluate the utilisation of systemic oncological treatments in clinical practice and (ii) estimate the uptake, safety and effectiveness of adjuvant trastuzumab-based treatment for HER2-positive EIBC.

The research used routine healthcare data from English Cancer Registration records linked at patient- and tumour-level to other cancer and administrative datasets including detailed information on patient and tumour factors, systemic oncological treatments and subsequent outcomes. For estimation of the effectiveness of trastuzumab-based treatment appropriate statistical methods were applied to provide estimates for the cohort overall and within patient subgroups underrepresented in RCTs, particularly older patients and those with comorbidity or poor fitness.

2.3. Objectives

The work presented in this thesis addressed the following research questions:

Understanding recording of treatment information in routine data:

1. Does information on use of cancer drug therapy (CDT) in routinely collected hospital admissions data provide information that complements the SACT dataset?
2. Can we combine information recorded in the Primary Care Prescription Database (PCPD) and within secondary care data to provide valid estimates of endocrine therapy use?

Understanding systemic oncological treatment in clinical practice:

3. How do utilisation levels of NICE-recommended systemic oncological drugs vary among women diagnosed with invasive breast cancer in England?
4. What factors are associated with the use of adjuvant trastuzumab-based treatment for HER2-positive EIBC?
5. How does the patient population receiving adjuvant trastuzumab-based treatment for HER2-positive EIBC in clinical practice compare to the trial populations, in relation to cohort characteristics and experience of severe toxicity?
6. How do survival outcomes following adjuvant trastuzumab-based treatment for HER2-positive EIBC compare across patient groups?

3. Research Design

3.1. Data sources and patient population

The work presented in this thesis made use of the routine data used within the NABCOP. Data were based on information prospectively collected by the NHS, as part of the care and support of patients with cancer. 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 were used by the NABCOP in order to reduce the burden of data collection on staff and patients. For patients in England, the data are collated, maintained and quality assured by NCRAS, which is part of NHS England (previously Public Health England and then NHS Digital during the time period of this thesis). All patients with a confirmed diagnosis of cancer have been registered since 1971, with national coverage.[108]

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. Cancer Registration records were provided, linked to multiple other data sources, via a pseudonymised patient and tumour ID provided within each. This patient and tumour ID was created and applied by NCRAS, who performed the initial data linkage of patients (and registered breast cancer tumours) across each data feed requested.

These linked data were used to understand the patient and tumour characteristics at diagnosis, the care and treatment received by patients and subsequent outcomes. This process of data extraction and release was repeated by NCRAS on an annual basis, with a full refresh of data provided each time along with an additional one year of patients diagnosed. The data feeds, broad content and structure of each dataset are detailed in Table 3.

All data provided were the raw data extracted from the CAS, but released to the NABCOP with various levels of cleaning, labelling and processing required to enable analysis. Many of the datasets were provided with information spread across multiple rows per patient/tumour, which required further processing to create manageable, “one-row-per-patient” datasets for analysis, containing the key information required. I completed this work as the NABCOP Research Fellow/Methodologist, which involved extensive statistical data management, and responsibility for preparing data for annual reports.

Table 3: Routine data sources used within the research.

Data source	Record level	Content	Provided data files used in the research & rows per record level
Cancer registry	Tumour	Data within Cancer Registry included information on all aspects of the cancer registration including information from hospital pathology systems and from the Civil Registration (death) records which contained details on date and cause of death, along with a vital status date at which patients not known to have died were considered to have been alive.	Provided as multiple data files to cover different aspects of a cancer registration. File 1: Data on patient characteristics (including age, ethnicity & calculated deprivation quintile), tumour diagnosis details (including grade, ER/PR/HER2 status, tumour size, nodal involvement), provider information, death (date & cause) & censoring information. One row per tumour ID. File 2: Data on aspects of treatment for cancer, including imaging (described as events). Multiple rows per patient & tumour ID, with one row for each aspect of the recorded event. This allowed the individual aspects to be recorded such as several imaging techniques used on the same date. File 3: Details of previous breast cancer, used to define the NABCOP cohort. Provided as one row per patient ID, with wide format used where patients had more than one previous instance of a breast cancer registration.
COSD	Tumour	Cancer Outcomes and Services Dataset (COSD) data 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. Data were provided on all aspects of diagnosis, care and treatment, including imaging, pathology, staging, demographics.	Provided as multiple data files to cover different aspects of cancer diagnosis, care, treatment and outcomes and allow for some files to have multiple rows per patient & tumour ID. Imaging file: Data on imaging modality, anatomical site & side, date, result, provider, outcome, radiological lesion size. Multiple rows per patient & tumour ID, with one row for imaging type and associated date. Stage file: Data on staging type (final pre-treatment, pathological, integrated), overall stage, TNM components, staging date. Multiple rows per patient & tumour ID, with one row for stage type and associated staging date. Treatment file: Data on treatment modality, intent, start date, surgical procedure information (date, OPCS code, discharge date), radiotherapy detail (dose, fractions). Multiple rows per patient & tumour ID, with one row for each aspect of the recorded treatment. Pathology file: Data on diagnostic pathology including grade, ER/PR/HER2 status, tumour size, nodal involvement. Recurrence file: Data on date of recurrence, type, site of metastases, care plan/key worker/palliative care specialist indicators. Multiple rows per patient & tumour ID. Non primary cancer pathway file: Data on non primary cancer diagnosis date, pathway type and recurrence type. Multiple rows per patient & tumour ID.

Table 3: Routine data sources used within the research.

Data source	Record level	Content	Provided data files used in the research & rows per record level
SACT	Tumour	Systemic Anti-Cancer Therapy (SACT) data contained information on systemic oncological treatment prescribed in a secondary or tertiary care setting. Details included dates, regime(s) and dose.	Provided as a single data file of multiple rows per patient & tumour ID – one row for each drug administered in a treatment cycle i.e. a cycle including 3 drugs would be detailed across 3 rows of information. This allowed the individual drug aspects to be recorded such as dose. Also, it allowed for dates to differ where drugs were part of the same cycle but administered on different days.
RTDS	Tumour	Radiotherapy dataset (RTDS) contained information on radiotherapy treatment including dates, prescription region and dose.	Provided as a single data file of multiple rows per patient & tumour ID – multiple rows for each radiotherapy fraction administration to allow details of fractions delivered on the same date to differ, for example with relation to the number of teletherapy fields, the region irradiated and the prescribed/administered dose. Data captured in a nested format with individual radiotherapy fractions within prescriptions, which fall within radiotherapy episodes.
HES-APC	Tumour	Hospital Episode Statistics Admitted Patient Care (HES-APC) is the administrative database of all NHS hospital admissions in England.	Provided as a single data file of multiple rows per patient & tumour ID – one row for each aspect of a hospital admission episode. This allowed for each clinical speciality involved with the admission to be recorded.
PCPD	Patient	The Primary Care Prescription Database (PCPD) contained information on prescriptions dispensed within community pharmacies. Data were provided on endocrine therapy prescriptions from April 2015 onwards. The PCPD, captures a core part of cancer treatment and became available to the NABCOP during the course of my PhD.	Details of endocrine therapy prescriptions were provided as a single data file of multiple rows per patient ID – one row for each drug prescribed. This allowed the individual drug aspects to be recorded such as dose.
CWT	Tumour	Cancer Waiting Times (CWT) data included information on the first definitive treatment received by a patient for new, progressive or recurrent cancer.	Provided as a single data file of one row per patient & tumour ID.

The following subsections provide detail on the content and validity of each of the main routine data sources used within this thesis, along with any known limitations.

3.1.1. National Cancer Registry data

Data on all aspects of the cancer registration, including information on staging and the results of diagnostic tests from hospital pathology systems, were provided within Cancer Registry data. This information is collated and processed by NCRAS who use an event-based registration model with registrations coded to international standards and a detailed quality assurance process to ensure data are as accurate as possible. For England, Cancer Registry data have national coverage with estimated 98-99% case ascertainment.[109] Cancer registration records are initiated from a COSD record highlighting a new primary cancer, and so the completeness of data items are initially dependent on COSD data completeness, with subsequent input from registry staff who can directly access NHS systems to ascertain values for missing information required to register the cancer.

3.1.2. Cancer Outcomes and Services Dataset (COSD)

Cancer service providers are required to provide information for all patients diagnosed with cancer or receiving treatment for cancer funded by the NHS in England within the COSD.[57] The data collected within COSD provides information required to support the national registration of cancers. This data collection covers all aspects of the cancer diagnosis including information on staging and the results of diagnostic tests from hospital pathology systems with information submitted by trusts, on a monthly return basis, for all patients diagnosed from 2013 onwards.

3.1.3. Systemic Anti-Cancer Therapy (SACT) data

Data collection for the SACT dataset started in April 2012, with data returns mandatory from April 2014.[28] The dataset contains longitudinal data (including drug name, dose, administration dates and administration route), recorded on prescribed systemic oncological treatments, including standard and oral chemotherapy, immunotherapy and targeted biological therapy, for NHS patients treated for cancer in England. It has whole population coverage and high case ascertainment (94% of patients reported as receiving CDT in the National Cancer Waiting Times dataset were identified in SACT data).[110] Data completeness of drug name field and administration date was excellent, reported at 100%.[28] Data in the SACT dataset describe prescribed regimens rather than dispensed and so there may be instances in which treatment was recorded in the SACT dataset but never administered to the patient.

3.1.4. Hospital Episode Statistics Admitted Patient Care (HES-APC) data

HES-APC is the administrative database of all NHS hospital admissions in England.[58] For the research presented within this thesis records were supplied by NHS Digital to NCRAS. HES-APC data cover all aspects of hospital admissions, and a HES-APC record contains information on surgical procedures (Office of Population Censuses and Surveys [OPCS] codes, date of procedure) and patient comorbidities (ICD-10 codes), the details of which are entered by hospital coders from discharge summaries. HES-APC has provided the basis of “payment by results”

since 2004/5, and has universal coverage. It has longitudinal data linkage (data for each episode of care linked at patient-level), with a low missed match rate (estimated as 4%) and low opt-out (2.3%). Limitations associated with the HES-APC database are that it only captures admitted and day case patients, there is geographical variation in match rates and it is not possible to distinguish between conditions that were present prior to admission and those that arose during an admission based on ICD-10 information (recorded within the diagnosis fields).

3.1.5. Primary Care Prescription Database (PCPD)

The PCPD was made available to the NABCOP from 2020, initially as part of a collaborative feasibility study between the audit and NCRAS to understand whether the linkage of these data to Cancer Registration records helped further understand the treatment of patients diagnosed with breast cancer. The PCPD has population coverage of all community pharmacy dispensed prescriptions within England.[59] Endocrine therapy is amongst the routinely recorded prescriptions captured within the PCPD, being one of the main treatments given for hormone receptor-positive invasive breast cancer.

3.2. Data preparation and management

3.2.1. Data extraction and provision

The NCRAS extracted all the data described in Section 3.1 for patients fulfilling the following criteria.

Inclusion criteria:

- Women
- Aged 50 years and 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 diagnosis date from 01/01/2014 onwards

Exclusion criteria:

- Instances where breast cancer was only reported on the death certificate

Eligible women diagnosed with breast cancer were identified from Cancer Registration records and linked to national electronic health records data. Only women diagnosed and treated in NHS organisations were included in analyses due to the lack of data flows from independent hospitals. Men were not included, primarily due to the low incidence meaning any conclusions about hospital performance from analyses would be unreliable.

3.2.2. Data storage

The linked datasets were received, stored, managed and analysed at the Clinical Effectiveness Unit (CEU) at the Royal College of Surgeons of England (RCSEng). The CEU is a collaborative unit between the RCSEng and LSHTM, and Professor David Cromwell is the CEU Director. The CEU has established a secure data server to hold datasets containing patient-level information. The server is protected by a firewall and intruder detection equipment that

guards the server against access from users from outside, and is physically located in one of two access-controlled server rooms within the RCSEng, incorporating environmental controls (power, temperature, fire, etc.). The data were stored in a restricted access folder on this secure server and only those with a legal basis to access the data are allowed in to this folder. Access to the RCSEng is restricted to staff, members of the College and visitors. Everybody in the building is required to wear a name badge. Visitors need to be announced in advance. There is 24-hour on-site security and the premises are controlled with electronic security barriers and CCTV. Users only have access to specific areas of the CEU server, with access controlled by Windows Server User and Group Permissions. Users of the server are obliged to use a password-protected screensaver, which is activated 5 minutes after the computer has been left idle. All data were destroyed securely according to information governance standards held by NABCOP and the CEU in October in line with the end of the data sharing agreement.

3.2.3. Preparing data for analysis

Data were imported from Excel into Stata and saved in .dta format, to be processed and prepared for analysis. This was done in several stages. All preparation of the data for this thesis was conducted using Stata (StataCorp LP, College Station, Texas USA).[111]

An initial step, upon receipt of the data files released by NCRAS, was to (i) plot the percentage of patients with a record in each data file, over time (based on their month of diagnosis within Cancer Registry data) to ensure there were no issues with data linkage over the period and (ii) check all dates were provided in a usable format.

This section describes the steps taken to clean, label and further process the data files provided, in advance of them being used for analysis.

Data were extracted and released by NCRAS in a raw format (string or numerical, depending on the column content). Initial steps in the data cleaning and labelling process were to:

1. format all dates from string format to Stata dates;
2. give an underlying name to all variables;
3. work through all categorical variables to derive numeric and labelled variables, based on categories defined with the associated data dictionary provided online – this was to ensure consistency of labelling and the associated underlying value either across datasets or across data refreshes – within this, checks were carried out to highlight any values provided which didn't fit with those expected from the associated data dictionary coding;
4. label all variables with the data source, in capitals at the start of the variable name for those raw variables and in lower case at the end of the variable name for derived variables (e.g. COSD_stage would denote the raw stage variable whilst stage_cosd would denote the processed variable);
5. save individual data files as a new Stata dataset with the suffix "_clean" to differentiate this from the original raw dataset.

The second set of steps were to process the data files to put them in a suitable format for analysis. This stage of processing constituted a large component of the work due to the way the routine data were provided with information spread across multiple data files and most files being long format, with multiple rows per patient and tumour ID. Such files, particularly those used to extract treatment information, required extensive manipulation to draw out the top level information whilst retaining sufficient detail for more in-depth analysis. For example, for chemotherapy information contained within the SACT dataset, processing covered multiple levels. These ranged from creating a basic flag for chemotherapy use (based on a known chemotherapy drug recorded in the drug name variable) with a start date (based on the earliest recorded date of administration for a chemotherapy drug), to creating variables detailing full chemotherapy regimens with a list of all the drugs involved, associated numbers of cycles, start and end dates and the weekly frequency of the cycles.

Initial work preparing data from the SACT dataset for analysis highlighted various features to navigate. These included determining which variable to use for drug details and administration dates, as well as the value of provided data on cycle number, cycle/regimen dates. For the work presented in this thesis the SACT dataset required the most processing; the work done for this, along with associated principles and implications, are presented in Table 4 below.

Table 4: Steps in SACT dataset processing.

SACT processing	Principle	Implication
Step 1 – set up code for data cleaning		
Define the drugs/regimens of interest for analysis	The SACT dataset captures information on all systemic oncological treatment prescribed, of which not all may be of interest.	Drugs (and drug types) of interest were defined and categorised into chemotherapy, her2-targeting therapy, hormone therapy, other.
Define a minimum set of variables of interest	The SACT dataset captures many aspects of treatment prescribing which may not be of use within the analysis so to maximise computing power and reduce analytical burden it is beneficial to only retain those variables to be used in analysis, within a final cleaned dataset.	Variables on administration date, drug, dose, administration route were retained as the key elements for analysis.
Write code to label categorical variables	Data are provided in raw format either as text or values with no label attached. To make sense a degree of processing is required.	The final dataset becomes usable without having to remember or know what each value of a variable refers to.
Define checks and rules to deal with missing or conflicting information:		
Administration vs cycle date	The SACT dataset includes several variables containing date information. The level of information of interest will likely determine which to use. At the level of individual drugs administration date may be the same or differ, depending on whether multiple drugs are given and if a drug is given on the first day of a cycle or not.	A useful check to decide which date variable to use for defining the first date of a therapy/regimen is whether regimen start date = earliest cycle start date = earliest date of administration. For analysis only wanting to understand the cycle level detail use administration date where this is >=7 days from cycle start date; use more accurately reflect cycles of a drug & enables collapsing

Table 4: Steps in SACT dataset processing.

SACT processing	Principle	Implication
Cycle number	The SACT dataset captures the associated cycle number for each drug. These have the potential to be useful to understand the number of cycles and what is given in each.	In practice cycle number was out of sync with dates in many instances. As such it made more sense to calculate this using administration/cycle dates.
Planned number of cycles	The SACT dataset captures the planned number of cycles for each drug. This would be useful to understand planned vs actual numbers of cycles.	In practice planned number of cycles may be inconsistent across drugs and would require understanding how this came to be recorded to understand value.
Treatment intent	Intent of each regimen of drugs can be recorded, based on the surgical setting or drugs given with curative vs palliative intent.	This would be useful to understand utilisation of drugs given with specific intents, for example in a palliative setting. The options for this have changed so it's important to be aware of the history of this variable. Also, it can be poorly recorded and sometimes inconsistent. Define treatment intent definitions for the study and rules created e.g. based on timing; based on drugs and doses. These are likely to be disease-specific
Drug details	The SACT dataset provides 3 variables with drug details recorded: Benchmark group & analysis group, provide top level mapped details on the intended regimen prescribed, which in principle give the group detail of drugs recorded within drug group; Drug group provides details of each individual drug.	Benchmark and analysis group are mapped regimen details. As such it is worth checking that the contents of this match the individual drugs recorded in drug group. For breast cancer, where there are many different multi-drug regimens this was not always the case and so drug group was used. Using drug group allows for counting of cycles for each drug and is most useful to understand sequencing or switching of drugs over time.
Step 2 – carry out data cleaning and checks		
Put all date variables in the same format	Data are provided in raw format which require processing to be useable.	Check all dates are in an expected and useable format
Check completeness of variables	The SACT dataset has many capabilities for capturing information around the provision of systemic oncological treatment and so will invariably have some variables with lower than expected completeness, particularly those referring to peripheral information.	The minimum set of variables might need to be updated to flag any with poor completeness. For example, staging information within the SACT dataset is of fair completeness but not recorded in a consistent format.
Check records are for breast cancer	The SACT dataset captures information on systemic oncological treatment for any cancer. Within the data collected is the ICD-10 code associated with treatment.	It is worth checking the drugs in the data files are all for breast cancer and removing those which are not.
Label variables and flag odd categories not in the data dictionary	The SACT dataset has been in use since 2014 and so for some variables the codes used may have changed over time.	Create new, labelled variables of expected values. Write in checks to highlight any codes not expected.

Table 4: Steps in SACT dataset processing.

SACT processing	Principle	Implication
Check the consistency of information within patient/tumour ID, regimens, cycles.	The definition of consistency will depend on the level of the variable and expected values across individual drug records. Checking consistency across expected record levels will contribute to understanding on the reliability of the data in a variable.	Decide on rules for dealing with conflicting information across records where information would be expected to be consistent e.g. diagnosis code, treatment intent.
Drop variables not of use or interest	The SACT dataset contains a lot of information across many variables. To improve analytical experience it is good practice to only retain variables which are to be used.	Save a new data file at each key step in data processing to remove the need to do everything again if the variables of interest change over the course of a project.
Drop rows with no drug information	The primary value of the SACT dataset lies in knowing the prescribed drug name. Without this information details of date, dose etc. are largely redundant.	Remove records containing no (useful) drug information: • "NOT CHEMO" – associated drugs given at the time (e.g. anti-nausea) provides no info except possibly dose • "TRIAL" – no detail of the drug • "MULTIPLE" – multiple drugs/regimens that didn't fit into the drug groups set up • "NOT MATCHED" – abbreviations used could have matched several drugs or bad spelling so not possible to identify • Records missing the drug name.
Unify drug names	Most recording of drug names within the SACT dataset is standardised and so there are no variations based on misspellings however, some drugs might have various names which are mapped and pulled through in the raw data.	Remove variations of the same drug for calculation of numbers of cycles etc. For example the chemotherapy nab-paclitaxel can be renamed paclitaxel, as the name being most commonly used.
Code up drugs of interest into types.	The SACT dataset captures drugs which fall into multiple category types. For initial analysis, it is helpful to have drugs grouped together to answer questions such as whether a particular type of treatment was used.	Drug category types for breast cancer captured in the SACT dataset cover: chemotherapy; her2-targeting therapy; endocrine therapy; bisphosphonates; immunotherapy; CDK 4/6 inhibitors The granularity of the groups created will be determined by the work being done
Check all drugs have been coded to a group.	The drugs available for use in breast cancer are regularly being updated and new types may become available.	Keep a check on any new drugs not grouped by listing those not assigned to a type to see if anything has been missed. These might form another group, be coded as “other”, or dropped if not of interest.

Further work was then done drawing together all information on recorded treatments to create a variable describing the sequence of treatment for each patient. This was done to establish ordering, to understand use of treatments around surgery, where given, which was used to define cohorts for analysis, as well as to understand sequential or concurrent chemotherapy use, alongside trastuzumab, and use of radiotherapy during systemic oncological treatment cycles.

3.2.4. Defining the NABCOP cohort

Final data processing steps were taken to draw together all information from across the multiple data files, to create final data files of patient characteristics, tumour characteristics, treatment and outcomes, and extract the NABCOP cohort. Table 5 details the criteria which were applied by the NABCOP to define the audit cohort.

Table 5: Steps taken to define the NABCOP cohort.

Criteria	Defined by	Exclude if
Breast cancer was not diagnosed at point of death	ONS date of death and date of diagnosis	ONS date of death same as date of diagnosis
Breast cancer was newly diagnosed	Flag in *Cancer Registry*	Patient has a previous diagnosis of breast cancer prior to 01/01/2014
Patient had only one breast cancer tumour	One tumour ID in *Cancer Registry* OR Laterality is bilateral	Patient has multiple tumour IDs Tumour laterality is recorded as bilateral
Patient provided by NCRAS was diagnosed in an English NHS trust	Diagnosing trust ID*	Patient has no reported place of diagnosis OR Reported place of diagnosis is outside of England OR Trust is not part of the NHS**
Patient diagnosed within an NHS trust where at least 30 women, aged 50+ years, were registered with breast cancer each year	Count of the number of patients registered for a trust each year, from the data provided by NCRAS.	Patient diagnosed within an NHS organisation with less than 30 allocated registrations of breast cancer, in women aged 50 years and over, per year
Patient was diagnosed with DCIS or invasive disease	ICD code reported as C50 or D05.1 in *Cancer Registry*	Tumour ICD code is not recorded as C50 or D05.1 (ductal carcinoma in situ)

*Note: where no trust of diagnosis was reported in Cancer Registry, the trust providing breast conserving surgery or mastectomy was used; where this was also not reported or the patient did not receive such surgery the trust where the MDT discussion occurred was used.

**Note: patients diagnosed and treated solely within the private sector were not included.

3.3. Statistical Methods

Full details of the statistical methods used for each study are outlined within the associated research paper. This section provides an overview of the main statistical methods used which were common to all papers. All statistical analysis for this thesis was conducted using Stata.[111]

3.3.1. Logistic regression modelling of clustered data

For analysis of binary outcomes, logistic regression models were used. Estimated measures within these logistic regression models were interpreted as odds ratios (ORs), whereby the OR indicates the odds of experiencing an outcome given a specified value or level of an exposure, compared with the odds of either not experiencing the exposure or a reference category of the exposure.

Models were typically multivariable, to enable identification of independent risk factors and also to adjust for potential confounders, such as factors associated with treatment decisions or other prognostic factors linked to

a clinical outcome. Details of variables used and the associated routine data source from which the information was taken are presented in Table 6.

The information presented within this thesis is for patients diagnosed and treated within NHS organisations, and so can be considered as being hierarchical or clustered. This is because patients within a single NHS organisation are more likely to be treated/cared for in a similar way than compared with patients in a different NHS organisation. As such, outcomes amongst patients diagnosed and treated within the same NHS organisation are likely to be correlated, and so standard errors estimated from conventional logistic regression models which don't account for this clustering will typically be underestimated and therefore provide incorrect estimates of variation. Underestimation of standard errors will subsequently increase the likelihood of a Type 1 error (false-positive), and so there is the increased possibility of finding evidence of an association between exposure and outcome where there is no true association. There are several methods for statistical modelling where we want to describe associations between patient-level factors and binary outcomes and also account for correlated/clustered data. Within a standard regression model, robust standard errors can be used. Using this technique will only account for the clustering within the standard errors but not within estimates of ORs. Alternatively, we can use multi-level mixed effects (MLME) models, and include a random intercept for NHS organisation. Using MLME models will account for clustering when estimating both the ORs and standard errors.[112]

Table 6: Details of the data variables used within the research and associated rules where multiple data sources provided information.

Variable	Primary data source	Rule
Age at diagnosis	Cancer Registry	Age at diagnosis as reported in Cancer Registry was never missing due to the way data were extracted and provided by NCRAS.
Ethnicity	Cancer Registry	Only one source of information on ethnicity was available for this research.
Deprivation	Cancer Registry	Index of Multiple Deprivation (IMD) quintiles were calculated by NCRAS and provided within the data.
Charlson comorbidity index (CCI)[106]	HES-APC	ICD-10 codes within the diagnosis field in HES-APC were identified for the following conditions: Myocardial infarction, Congestive cardiac failure, Peripheral vascular disease, Cerebrovascular disease, Dementia, Chronic Pulmonary disease, Rheumatological disease, Liver disease, Diabetes Mellitus, Hemiplegia or Paraplegia, Renal disease and AIDS/HIV. The condition was included in the CCI calculation if it was present within a patient's diagnosis field, at any stage, in the TWO years before the date of diagnosis. The presence of malignancy or metastatic solid tumour was not included in the score.
Secondary Care Administrative Records Frailty (SCARF) Index[107]	HES-APC	ICD-10 codes within the diagnosis field in HES-APC were identified for conditions defined in Jauhari et al. A condition was included in the SCARF Index calculation if it was present within a patient's diagnosis field, at any stage, in the TWO years before the date of diagnosis.
Year of diagnosis	Cancer Registry	Based on date of diagnosis provided within the data.
Overall stage	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with stage reported in Cancer Registry; calculated in the following order: 1. _best 2. Pathology 3. Imaging Observations which were missing or reported as "Unstageable" or "Unstated/Not staged" were updated with a 'final' value from those reported in COSD pathology; derived from stage values in (preference order): 1. stage_integrated 2. stage_pretreatment 3. stage_pathology Where stage remained missing or reported as "Unstageable" or "Unstated/Not staged" the value was calculated using the final derived 'T | N | M' variables using UICC (v6) rules. In addition Stage was calculated as 0 for those individuals with DCIS disease and no stage reported.

Table 6: Details of the data variables used within the research and associated rules where multiple data sources provided information.

Variable	Primary data source	Rule
Tumour stage	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with T stage reported within 'T_best'. Observations not reported as T1-T4 or Tis were augmented with values from (in order) 1. T_path 2. T_imaging Observations which were missing or reported as "TX" were updated with a 'final' value from those reported in COSD pathology; derived from T stage values in (preference order): 1. tstage_integrated 2. tstage_pathology 3. tstage_pretreatment
Nodal stage	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with N stage reported within 'N_best'. Observations not reported as N1-N3 were augmented with values from (in order) 1. N_path 2. N_imaging Observations which were missing or reported as "NX" were updated with a 'final' value from those reported in COSD pathology; derived from N stage values in (preference order): 1. nstage_integrated 2. nstage_pathology 3. nstage_pretreatment Any observations which were still missing or reported as NX were finally updated based on available details on the number of positive nodes (N0 = 0 nodes positive; N1 = 1-3 nodes positive; N2 = 4-9 nodes positive; N3 = 10+ nodes positive)

Table 6: Details of the data variables used within the research and associated rules where multiple data sources provided information.

Variable	Primary data source	Rule
Metastases stage	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with M stage reported within 'M_best'. Observations not reported as M0 or M1 augmented with values from (in order) 1. M_path 2. M_imaging Observations which were missing or reported as "MX" were updated with a 'final' value from those reported in COSD pathology; derived from M stage values in (preference order): 1. mstage_integrated 2. mstage_pathology 3. mstage_pretreatment Any observations remaining missing are updated based on an overall tumour stage value where known (M0 if Stage 0-3; M1 if stage 4)
HER2 status	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with HER2 status as reported in Cancer Registry. Update any observations which are missing or reported as "not performed" or "unknown" with the reported value from – COSD pathology derived from recorded values in the following specimens (in order of preference): 1. Excision 2. Biopsy 3. Cytology 4. Unknown 5. HER2ish status
ER status	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with ER status as reported in Cancer Registry. Any observations remaining missing or reported as "not performed" or "unknown" were updated using the ER score reported in the relevant primary source (negative = 0-2; positive = 3+). Further updates of any observations which remained missing or reported as "not performed" or "unknown" were with the reported value from - COSD pathology derived from recorded values* in the following specimens (in order of preference): 1. Excision 2. Biopsy 3. Cytology 4. Unknown *Note that COSD pathology values, where missing, followed the same process of augmentation with ER score where reported.

Table 6: Details of the data variables used within the research and associated rules where multiple data sources provided information.

Variable	Primary data source	Rule
PR status	Cancer Registry	Details were provided in both Cancer Registry and COSD data files. Start with PR status as reported in Cancer Registry. Any observations remaining missing or reported as "not performed" or "unknown" were updated using the PR score reported in the relevant primary source (negative = 0-2; positive = 3+). Further updates of any observations which remained missing or reported as "not performed" or "unknown" were with the reported value from - COSD pathology derived from recorded values* in the following specimens (in order of preference): 1. Excision 2. Biopsy 3. Cytology 4. Unknown *Note that COSD pathology values, where missing, followed the same process of augmentation with PR score where reported.
Invasive grade	Cancer Registry	Start with invasive grade as reported in Cancer Registry. Observations which are missing, reported as GX or incorrectly reported are augmented with – A 'final' invasive grade reported within COSD pathology. This is derived from specimen results from (in order): 1. Excision 2. Biopsy 3. Cytology 4. Unknown
Receipt & type of surgery	HES-APC	Defined based on OPCS codes recorded within the operation fields. See Appendices of Supplementary material for Research Paper 4 for detail of the codes (Appendix 5). Details on surgery/surgical procedure were also provided in both Cancer Registry and COSD treatment data files, however the information in these were not as extensive and were not used as HES-APC was considered to be the primary source of information on this aspect of treatment.
Receipt of radiotherapy	RTDS	Radiotherapy details recorded in RTDS. Detail on radiotherapy as part of cancer treatment were also provided in Cancer Registry and COSD treatment data files, however the information in these were limited and RTDS was considered to be the primary source of information on this aspect of treatment.
Receipt of chemotherapy	SACT	Recording of a chemotherapy* drug within drug group. Details on chemotherapy treatment were provided in Cancer Registry and COSD data files. Additionally it would be possible to determine use of chemotherapy treatment within HES-APC records based on OPCS codes recorded within the operation fields.

Table 6: Details of the data variables used within the research and associated rules where multiple data sources provided information.

Variable	Primary data source	Rule
Receipt of endocrine therapy	COSD, Cancer Registry, SACT, PCPD	Defined where either 1. COSD/Cancer Registry had hormone therapy recorded as a cancer treatment; 2. SACT had any of aminoglutethimide, anastrozole, buserelin, exemestane, formestane, fulvestrant, goserelin, gtx-024, letrozole, leuprorelin, megestrol, medroxyprogesterone, tamoxifen, toremifene, triptorelin recorded in drug group; 3. A record was available in PCPD.
Receipt of trastuzumab	SACT	Recording of trastuzumab within drug group. It might be possible to determine use of trastuzumab treatment within HES-APC records based on OPCS codes recorded within the operation fields.

NOTE:

For all final pathology variables created, Cancer Registry values were taken as preference for the primary source due to:

I. NCRAS data origin rules
II. Better data completion (compared to COSD)
III. Some variables were better quality based on the data range

*Chemotherapy drugs included: cabazitaxel; capecitabine; carboplatin; cisplatin; cyclophosphamide; docetaxel; doxorubicin; epirubicin; Eribulin; etoposide; fluorouracil; gemcitabine; methotrexate; mitomycin; mitoxantrone; paclitaxel; vinorelbine.

3.3.2. Time-to-event analysis

When an outcome was defined as a time to an event (including time to initiating treatment or time to death), standard survival analysis methods were used. Kaplan-Meier methods were used to estimate survival or time-to-event curves. Cox models were used to estimate hazard ratios (HRs) in many situations. Flexible parametric models were fitted to understand the presence of any non-proportional hazards violations and whether this impacted the findings.

For analysis of survival outcomes, overall survival (OS) was defined where death was recorded, regardless of the cause. Cause of death was recorded using ICD-10 codes and provided within the Cancer Registry data files, based on civil registration and mortality data feeds. Breast cancer-specific survival (BCSS) was defined where the underlying cause of death was recorded as the ICD-10 code for breast cancer (C50). Cause of death was coded as non-breast cancer where the patient was recorded as having died but the underlying cause of death was recorded as something other than breast cancer.

For estimating the effect of treatment on BCSS, competing risks analysis is most appropriate.[113] Within this analysis framework, other causes of death are accounted for as a competing risk. Where the primary outcome is death from breast cancer, but a patient dies from something other than their breast cancer, in a traditional survival analysis, the patient would be censored at this point and not considered to have experienced the event of interest/primary outcome. However, this patient no longer has a chance of dying from their breast cancer, and this is important to take into account. A competing risks analysis can be used to minimise overestimation of the risk of the primary outcome where patients are censored after experiencing a competing event. The underlying cause of death was used for competing risks analysis. The Fine and Gray competing risks regression model was used to estimate adjusted subdistribution HRs (subHRs), adjusting for multiple explanatory factors.[114]

3.4. Ethics

The research presented within this thesis ran alongside the National Audit of Breast Cancer in Older Patients (NABCOP), of which I was a member of the project team. Specifically, the research was conducted within the broader work of NABCOP and carried out under the NABCOP data permissions through my role as the NABCOP Research Fellow/Methodologist.

The NABCOP was exempt from the UK National Research Ethics Committee approval because it involved analysis of data for services. The NABCOP had Section 251 approval from the UK Ethics and Confidentiality Committee, for the use of routinely collected, patient-level data (reference number: 16/CAG/0079) for all patients, aged 50 years and over, diagnosed with breast cancer in England and Wales. This covered research in to trends and variation in numbers of patients treated, differences in characteristics, treatments given and outcomes. The NABCOP evaluated the care received by women aged 50 years and over, diagnosed in England and Wales from 2014 onwards. Work conducted within this thesis falls under the remit of the audit work and is therefore covered by pre-existing approvals. I only had access to de-identified patient-level data meaning that individual patients were not identifiable, with patients linked across the datasets provided by a pseudonymised ID. The

NABCOP had approval for the processing of data under articles 6(1)(e) and 9(2)(i) of the General Data Protection Regulation (GDPR).

This research was therefore exempt from UK National Research Ethics Committee approval. However, as part of the PhD upgrading process, research plans were reviewed by the LSHTM Ethics Committee. Ethics approval was granted on July 5th 2021, reference 22508 (Appendix 1).

3.5. Additional Outputs

Within my role as the NABCOP Research Fellow/Methodologist, I was involved in further work which has not contributed directly to this thesis, but has given me the opportunity to lead on research and audit work, provide methodological input and contribute to improvements in care for patients with breast cancer. These additional research papers, national audit reports, and presentations that I have contributed to during the course of my research are listed in the following sections.

3.5.1. Additional peer-reviewed journal publications (ordered with most recent first)

- Miller K, **Gannon MR**, Medina J, Clements K, Dodwell D, Horgan K, Park MH, Cromwell DA. Variation in Rates of Post-Mastectomy Radiotherapy Among Women with Early Invasive Breast Cancer in England and Wales: A Population-Based Cohort Study. Clin Oncol (R Coll Radiol). 2023 Sep;35(9):e549-e560. doi: 10.1016/j.clon.2023.05.016. Epub 2023 May 30. PMID: 37321887.
- Miller K, **Gannon MR**, Medina J, Clements K, Dodwell D, Horgan K, Park MH, Cromwell DA. The Association Between Survival and Receipt of Post-mastectomy Radiotherapy According to Age at Diagnosis Among Women With Early Invasive Breast Cancer: A Population-Based Cohort Study. Clin Oncol (R Coll Radiol). 2023 Apr;35(4):e265-e277. doi: 10.1016/j.clon.2023.01.008. Epub 2023 Jan 23. PMID: 36764877.
- **Gannon MR**, Dodwell D, Miller K, Horgan K, Clements K, Medina J, Kunkler I, Cromwell DA. Change in the Use of Fractionation in Radiotherapy Used for Early Breast Cancer at the Start of the COVID-19 Pandemic: A Population-Based Cohort Study of Older Women in England and Wales. Clin Oncol (R Coll Radiol). 2022 Sep;34(9):e400-e409. doi: 10.1016/j.clon.2022.05.019. Epub 2022 May 31. PMID: 35691761; PMCID: PMC9151525.
- Miller K, Kreis IA, **Gannon MR**, Medina J, Clements K, Horgan K, Dodwell D, Park MH, Cromwell DA. The association between guideline adherence, age and overall survival among women with non-metastatic breast cancer: A systematic review. Cancer Treat Rev. 2022 Mar;104:102353. doi: 10.1016/j.ctrv.2022.102353. Epub 2022 Jan 31. PMID: 35152157.
- Jauhari Y, **Gannon MR**, Dodwell D, Horgan K, Clements K, Medina J, Cromwell DA. Surgical decisions in older women with early breast cancer: patient and disease factors. Br J Surg. 2021 Mar 12;108(2):160-167. doi: 10.1093/bjs/znaa042. PMID: 33711149; PMCID: PMC7954278.
- Dodwell D, Jauhari Y, Gathani T, Cromwell D, **Gannon M**, Clements K, Horgan K. Treatment variation in early breast cancer in the UK. BMJ. 2020 Dec 1;371:m4237. doi: 10.1136/bmj.m4237. PMID: 33262116.
- Jauhari Y, Dodwell D, **Gannon MR**, Horgan K, Clements K, Medina J, Cromwell DA. The influence of age, comorbidity and frailty on treatment with surgery and systemic therapy in older women with operable triple negative breast cancer (TNBC) in England: A population-based cohort study. Eur J Surg Oncol. 2021 Feb;47(2):251-260. doi: 10.1016/j.ejso.2020.09.022. Epub 2020 Oct 2. PMID: 33268213.

- Sun L, Cromwell D, Dodwell D, Horgan K, **Gannon MR**, Medina J, Pennington M, Legood R, Dos-Santos-Silva I, Sadique Z. Costs of Early Invasive Breast Cancer in England Using National Patient-Level Data. Value Health. 2020 Oct;23(10):1316-1323. doi: 10.1016/j.jval.2020.05.013. Epub 2020 Jul 28. PMID: 33032775.
- Jauhari Y, **Gannon MR**, Dodwell D, Horgan K, Clements K, Medina J, Tsang C, Robinson T, Tang SS, Pettengell R, Cromwell DA. 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 May 5;10(5):e035395. doi: 10.1136/bmjopen-2019-035395. PMID: 32376755; PMCID: PMC7223146.
- Jauhari Y, **Gannon MR**, Dodwell D, Horgan K, Tsang C, Clements K, Medina J, Tang S, Pettengell R, Cromwell DA. Addressing frailty in patients with breast cancer: A review of the literature. Eur J Surg Oncol. 2020 Jan;46(1):24-32. doi: 10.1016/j.ejso.2019.08.011. Epub 2019 Aug 12. PMID: 31439357.
- Jauhari Y, **Gannon MR**, Tsang C, Horgan K, Dodwell D, Clements K, Medina J, Tang S, Pettengell R, Cromwell DA. Surgery and adjuvant radiotherapy for unilateral ductal carcinoma in situ (DCIS) in women aged over 70 years: A population based cohort study. Eur J Surg Oncol. 2019 Aug;45(8):1378-1387. doi: 10.1016/j.ejso.2019.02.034. Epub 2019 Mar 7. PMID: 30878169.

3.5.2. National Audit of Breast Cancer in Older Patients Annual Reports

- National Audit of Breast Cancer in Older Patients. NABCOP 2018 Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2018 Public and Patients Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2019 Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2019 Public and Patients Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2020 Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2020 Public and Patients Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2021 Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2021 Public and Patients Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2022 Annual Report
- National Audit of Breast Cancer in Older Patients. NABCOP 2022 Public and Patients Annual Report

3.5.3. Academic Conference papers and national presentations (first author only)

- **Gannon M**, Jauhari Y, Horgan K, Medina J, Clements K, Cromwell D, Dodwell D. Chemotherapy utilisation in patients aged 50 years and over, diagnosed with invasive early breast cancer in England: Data from a population-based cohort. *European Journal of Surgical Oncology* 2018; DOI: 10.1016/j.ejso.2018.02.029. [Oral presentation at ABS Conference 2018]
- **Gannon M**, Jauhari Y, Medina J, Horgan K, Cromwell D, Dodwell D. P011 Use of taxane-containing adjuvant chemotherapy regimens among older women with operable early breast cancer: Data from a population based cohort within the National Audit of Breast Cancer in Older Patients (NABCOP). *Journal of Geriatric Oncology* DOI: 10.1016/S1879-4068(18)30440-5. [Poster prize presentation at SIOG Conference 2018]
- **Gannon M**, Jauhari Y, Horgan K, Medina J, Clements K, Dodwell D, Cromwell D. Variation in use of adjuvant chemotherapy and trastuzumab for human epidermal growth factor receptor 2-positive (HER2 positive) early invasive breast cancer in a population-based study of older women. [Oral presentation at British Breast Group meeting 2019]

- **Gannon MR,** Dodwell D, Jauhari Y, Horgan K, Clements K, Medina J, Cromwell DA. Adjuvant chemotherapy and trastuzumab for human epidermal growth receptor 2 positive early invasive breast cancer among older women in England: a population-based cohort study using routine data. [Oral Presentation at the 6th Symposium of Primary Breast Cancer in Older Women 2022]
- **Gannon MR**, Dodwell D, Miller K, Horgan K, Clements K, Medina J, Kunkler I, Cromwell DA. Change in the Use of Fractionation in Radiotherapy Used for Early Breast Cancer at the Start of the COVID-19 Pandemic: a Population-based Cohort Study of Older Women in England and Wales. *European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 2022; 48(5):e224-e225. DOI: 10.1016/j.ejso.2022.03.155. [Poster presentation at ABS Conference 2022]
- **Gannon MR**, Park MH, Miller K, Dodwell D, Horgan K, Clements K, Medina J, Cromwell DA. Agreement of breast cancer drug recording in routine hospital admissions & treatment data in England. *European Journal of Surgical Oncology* 2023; DOI: 10.1016/j.ejso.2023.03.028. [Oral presentation at ABS Conference 2023]

4. Understanding recording of treatment information in routine data

4.1. Research Paper 1

Title

Concordance of cancer drug therapy information derived from routinely collected hospital admissions data and the Systemic Anti-Cancer Therapy (SACT) dataset, for older women diagnosed with early invasive breast cancer in England.

Journal details

Published in Cancer Epidemiology. The online PDF version can be accessed at:
https://pubmed.ncbi.nlm.nih.gov/36774694/

Supplementary material can be found in the Appendices.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Cancer Epideimiology		
When was the work published?	13 February 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	[redacted]	
Date		

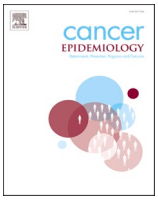
Supervisor Signature	[redacted]	
Date		

Cancer Epidemiology 83 (2023) 102337

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep



Concordance of cancer drug therapy information derived from routinely collected hospital admissions data and the Systemic Anti-Cancer Therapy (SACT) dataset, for older women diagnosed with early invasive breast cancer in England

Melissa Ruth Gannon [a,b,*], Min Hae Park [a,b], Katie Miller [a,b], David Dodwell [c], Kieran Horgan [d], Karen Clements [e], Jibby Medina [a,b], David Alan Cromwell [a,b]

[a] Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK
[b] Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK
[c] Nuffield Department of Population Health, University of Oxford, Oxford, UK
[d] Department of Breast Surgery, St James's University Hospital, Leeds, UK
[e] National Cancer Registration and Analysis Service, NHS Digital, 2nd Floor, 23 Stephenson Street, Birmingham, UK

ARTICLE INFO

Keywords:
Early breast cancer
Chemotherapy
Trastuzumab
Routine data
Hospital admissions
Older patients

ABSTRACT

Background: Evaluating uptake of oncological treatments, and subsequent outcomes, depends on data sources containing accurate and complete information about cancer drug therapy (CDT). This study aimed to evaluate the consistency of CDT information in the Hospital Episode Statistics Admitted Patient Care (HES-APC) and Systemic Anti-Cancer Therapy (SACT) datasets for early invasive breast cancer (EIBC).

Methods: The study included women (50 + years) diagnosed with EIBC in England from 2014 to 2019 who had surgery within six months of diagnosis. Concordance of CDT recorded in HES-APC (identified using OPCS codes) and SACT was evaluated at both patient-level and cycle-level. Factors associated with CDT use captured only in HES-APC were assessed using statistical models.

Results: The cohort contained 129,326 women with EIBC. Overall concordance between SACT and HES-APC on CDT use was 94 %. Concordance increased over the study period (91–96 %), and there was wide variation across NHS trusts (lowest decile of trusts had concordance≤77 %; highest decile≥99 %). Among women receiving CDT, 9 % (n = 2781/31693) of use was not captured in SACT; incompleteness was worst (18 %=47/259) among women aged 80 + and those diagnosed in 2014 (21%=1121/5401). OPCS codes in HES-APC were good at identifying patient-level and cycle-level use of trastuzumab or FEC chemotherapy (fluorouracil, epirubicin, cyclophosphamide), with 89 % and 93 % concordance with SACT respectively (patient-level agreement). Among cycles of solely oral CDT recorded in SACT, only 24 % were captured in HES-APC, compared to 71 % for intravenous/subcutaneous CDT.

Conclusions: Combining information in HES-APC and SACT provides a more complete picture of CDT treatment in women aged 50 + receiving surgery for EIBC than using either data source alone. HES-APC may have particular value in identifying CDT use among older women, those diagnosed less recently, and in NHS trusts with low SACT data returns.

1. Introduction

National guidelines for women diagnosed with early invasive breast cancer (EIBC) recommend the use of chemotherapy (in addition to surgery), along with targeted therapies where tumour and patient characteristics suggest those treatments would improve survival outcomes. For women diagnosed with human epidermal growth receptor 2 (HER2)-positive EIBC, the targeted therapy trastuzumab is recommended in combination with chemotherapy [1]. The evidence underlying such recommendations is primarily from clinical trials in relatively

* Correspondence to: London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, UK.
E-mail address: melissa.gannon@lshtm.ac.uk (M.R. Gannon).

https://doi.org/10.1016/j.canep.2023.102337
Received 7 September 2022; Received in revised form 6 January 2023; Accepted 5 February 2023
Available online 13 February 2023

fit, selected age cohorts which may be limited in their generalisability to older "real world" patient populations [2,3]. Consequently, it is desirable to be able to evaluate the uptake of oncological treatments, and subsequent outcomes, using national data sources to understand the risks and benefits of treatment outside of a trial setting [4]. Such evaluations depend on the data sources containing accurate and complete information about cancer drug therapy (CDT).

Patient-level data on aspects of breast cancer care are routinely collected in hospitals and mandatorily submitted to national organisations, as part of the care and support of patients with cancer. The Systemic Anti-Cancer Therapy (SACT) dataset collects patient and tumour-level data on CDT (such as chemotherapy and targeted therapy) delivered within secondary and tertiary care settings [5]. Previous publications using the SACT dataset have highlighted incomplete data capture and hospital-level variation in data returns and quality [5,6]. Studies of patients with lung cancer and colon cancer have compared chemotherapy recorded in the SACT dataset with information in the Hospital Episode Statistics Admitted Patient Care (HES-APC) dataset [7,8]. These identified that the recording of chemotherapy cycles in the SACT dataset was incomplete, with additional cycles identified in HES-APC and differences in data capture according to patient age and fitness, indicating that both data sources should be used to derive information about chemotherapy. For breast cancer, most CDT treatment is delivered as day case admissions in the secondary care setting, therefore HES-APC may provide an additional data source for identifying CDT use [9].

Improvement in cancer treatment outcomes requires the translation of recommendations on optimal treatment into delivery of those drugs to patients but there is evidence of considerable variation in this practice [10,11]. One aspect of the verification process of what happens in routine care is to examine complete, reliable information on CDT prescription at a patient-level. This has traditionally been a difficult task. The introduction of SACT in England greatly improved the quantity and quality of CDT information available nationally but there remain some gaps. NHS trusts with lower levels of SACT data returns require targeted approaches supported by data derived from other sources to ensure poor data returns do not mask deficiencies in care. Similarly higher levels of variation in cancer care including receipt of CDT are reported for older patients [10,12,13]. SACT alone does not currently meet all these data needs.

The aim of this study was to evaluate the consistency of CDT information recorded within SACT and HES-APC for a cohort of women aged 50 years and over newly-diagnosed with EIBC in England from 2014 to 2019. The rationale for the study was to determine the value of HES-APC in identifying CDT use, and whether it could provide information that complements the SACT dataset.

2. Materials and methods

2.1. Data and study population

This population-based cohort study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP). Linked, pseudonymised patient records were provided for all women aged 50 years and over, with a BC diagnosis recorded in Cancer Registry data, diagnosed and treated within a NHS trust in England, between 1 January 2014 and 31 December 2019. Cancer Registration records were linked at tumour-level to the Cancer Outcomes and Services Dataset (COSD) for details of patient and tumour characteristics, to HES-APC data, and to SACT records with an IBC International Classification of Diseases, 10th revision (ICD-10) diagnosis code. The Cancer Registration dataset was used to define the study cohort, while the HES-APC and SACT datasets provided information on the use of systemic therapy. The study cohort was defined as all women newly-diagnosed with EIBC (stage 1–3a), who had surgery within six months of diagnosis. For analysis looking at the capture of trastuzumab we identified a subgroup of women with HER2-positive EIBC for whom this targeted treatment is recommended.

2.2. Socio-demographic and clinico-pathological variables

Data on the following patient and tumour characteristics were taken from the Cancer Registry and COSD datasets: age at diagnosis (years), ethnicity, overall stage (1–3a), tumour stage (T1, T2, T3), nodal stage (N0, N +), HER2/ER status (positive or negative), tumour grade (G1, G2, G3).

Deprivation was measured using the Index of Multiple Deprivation (IMD) 2019 rank which was derived from the patient's postcode at diagnosis. The IMD rank was assigned to national quintiles of deprivation, from most (group 1) to least (group 5) deprived.

Comorbidity burden (0, 1, 2 +) was defined using the Royal College of Surgeons of England Charlson Comorbidity Index [14]. This Index counts the presence of specific chronic medical conditions (excluding malignancy), identified using ICD-10 diagnosis codes within patient HES-APC records for a period of two years prior to diagnosis.

Patient fitness (fit; mild-moderate frailty; severe frailty) was defined using the Secondary Care Administrative Records Frailty (SCARF) index [15]. This describes frailty in relation to 32 different symptoms, signs, diseases and disabilities (referred to as deficits), identified using ICD-10 diagnosis codes within patient HES-APC records for a period of two years prior to diagnosis.

2.3. Measuring use of cancer drug therapy (CDT)

NICE guidelines on chemotherapy for EIBC recommend taxane or anthracycline-containing regimens with the exact regimen decision decided locally [1]. All chemotherapy regimens were therefore considered eligible for this study, with clinical guidance used to identify chemotherapy drugs recorded in the drug name field in SACT. Records of HER2-targeted therapy (mostly trastuzumab) were also included, as it is predominantly used in conjunction with chemotherapy. CDT use was counted where the first recorded administration date was prior, or within four months after, date of surgery.

2.3.1. CDT data sources – The Systemic Anti-Cancer Therapy (SACT) dataset

Data collection for the SACT dataset started in April 2012, with data returns mandatory from April 2014. The dataset contains longitudinal data (including drug name, dose, administration dates, administration route), recorded on prescribed systemic anti-cancer therapies, including chemotherapy and targeted biological therapy, for NHS patients treated for cancer in England. It has whole population coverage and high case ascertainment (94 % of patients reported as receiving CDT in the National Cancer Waiting Times dataset were identified in SACT data) [16]. Data completeness of drug name and administration date is excellent, reported at 100 % [5]. The study used linked SACT data for drugs with an administration date from 1 January 2014 up to 31 March 2021.

2.3.2. CDT data sources - Hospital Episode Statistics Admitted Patient Care (HES-APC) dataset

HES-APC is an administrative dataset of all NHS hospital admissions in England. Coverage is almost universal (opt-out rate=2.3 %), and individual treatments are attributed to the same patient using an anonymised identifier (estimated missed match rate=4 %) [9]. Data on inpatient and day-case chemotherapy administrations are captured via clinical coding, primarily through pre-specified Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (OPCS) codes, or alternatively through ICD-10 codes which specify use of chemotherapy at the admission (see Appendix 1). The study used the National Tariff Chemotherapy Regimens List guidance on the OPCS procurement and delivery codes for identifying chemotherapy admissions in the linked HES-APC data with admission dates from 1 January 2014 up to 31 March 2021 [17,18]. Guidance on recording of combinations of regimens in HES notes that "in cases where a combination of regimens is prescribed and these are administered at the same

outpatient or day case attendance then the procurement code (X70, X71) and the corresponding delivery code (X72, X73) for each regimen must be assigned." Only records with an associated IBC ICD-10 diagnosis code (C50) recorded were included in the analysis.

2.4. Statistical analysis

2.4.1. Agreement between data sources

Contingency tables were used to explore patient-level agreement between SACT and HES-APC with respect to record of CDT. Concordance was defined as the percentage of women with agreement about CDT use (Yes/No) in both data sources. These were calculated for the overall cohort and within patient subgroups of age, comorbidity burden and year of diagnosis, identified in previous publications to have incomplete capture of CDT use in SACT [6–8]. Weighted kappa statistics were used to describe the strength of agreement between data sources, accounting for the degree of disagreement, and assess whether it was beyond that expected by chance alone. Kappa has a maximum of 1 (perfect agreement), and values higher than 0.80 were considered to demonstrate very good agreement.

Where patients had CDT recorded in both SACT and HES-APC, the percentage with agreement on the date of first cycle was calculated. Differences in agreement over time and by age were visually explored using bar charts. Funnel plots were used to assess variation in concordance (overall and only in women with a record of CDT) by NHS trust of diagnosis.

2.4.2. Identifying factors associated with additional treatment capture within HES-APC

Patient and tumour characteristics, year of diagnosis, type of surgery, use of radiotherapy and the CDT setting of first recorded cycle were described for patients with a record of CDT captured in either dataset. Factors were selected from previous publications that found completeness of capture of CDT use in SACT to vary between patient subgroups (age, fitness, year of diagnosis) or because they might be associated with the setting of CDT administration [6–8]. Ethnicity and deprivation were considered because of reported differences in cancer treatment according to these factors [19,20]. For each factor, multilevel mixed-effects logistic regression models were used to statistically assess the likelihood that HES-APC captures CDT use not recorded in SACT, accounting for the clustering of patients within an NHS trust. NHS trust was included as a random intercept, which estimates differences in the baseline percentage of women with a record in HES-APC only between trusts that are not explained by the factors in the model.

2.4.3. Agreement on number of cycles

Comparison of the number of CDT cycles recorded within each dataset was compared overall using overlapping bar charts. A patient-level comparison was conducted, within each CDT setting (neoadjuvant or adjuvant), among patients with CDT recorded in both SACT and HES-APC; agreement between the number of cycles recorded for a patient in each dataset was evaluated using a line of best fit from a Bland-Altman analysis [21]. Records within less than six days of each other were counted as being part of the same cycle. Neoadjuvant use was defined as all cycles with an administration date prior to date of surgery. Adjuvant cycles were counted from the first cycle within four months after surgery, up to the last cycle before a treatment break of more than three months or where no more cycles were recorded.

2.4.4. Agreement of treatment regimens

HES-APC does not directly record drug regimen names and so agreement was considered according to SACT-defined drug regimen. CDT records in SACT with an administration date that matched a CDT admission date recorded in HES-APC were used to identify the drug names most frequently recorded in SACT. The corresponding OPCS codes recorded in HES-APC were compared with expected tariff codes. OPCS codes associated with each of the most frequently recorded drug regimens in SACT (trastuzumab; paclitaxel; FEC (fluorouracil, epirubicin, cyclophosphamide); EC; docetaxel) were used to flag each within HES-APC and cross-tabulated with drug details in SACT to understand what percentage of patients had matching drug regimens recorded in HES-APC.

All data preparation and statistical analyses were conducted using Stata version 17.0.

3. Results

3.1. Recording of CDT use in SACT or HES-APC, overall and among patient subgroups

The linked dataset contained 129,326 women aged 50 years and over diagnosed with EIBC in England from 1 January 2014 to 31 December 2019 who had surgery within six months of diagnosis (Fig. A1). Among these, 25 % (n = 31,693) had a record of neoadjuvant or adjuvant CDT in either SACT or HES-APC. The recording of CDT use among women with different characteristics exhibited expected patterns (Table 1). Notably, recorded CDT use was highest for women with larger tumours, nodal involvement, grade 3, ER-negative or HER2-positive disease and following a mastectomy. The percentage of women with a record of CDT decreased with age at diagnosis and among women with more comorbidities and a greater level of frailty.

3.2. Concordance in the recording of CDT use in SACT or HES-APC

Overall concordance between the two datasets was 94 % (Table 2). Among women with CDT recorded in SACT, 81 % also had CDT recorded in HES-APC, whilst among women with CDT recorded in HES-APC, 89 % also had CDT recorded in SACT. Agreement between datasets was very good overall (weighted kappa=0.81) and varied little by age or comorbidity burden. Agreement improved slightly over time (Fig. 1).

There was variation in overall concordance by NHS trust (lowest 10 %=77 %; highest 10 %=99 %) with 14 % (n = 16/117) of trusts having less than 80 % concordance (Fig. 2). Of these trusts, 14 had no records of CDT in HES-APC.

Among women with CDT recorded in either dataset, there was variation in the percentage of CDT captured only in HES-APC across NHS trusts. For six NHS trusts high percentages of CDT captured solely in HES-APC were due to low rates of CDT in SACT (<60 %; Fig. A2).

3.3. Factors associated with additional CDT capture within HES-APC

Of the 31,693 women who had CDT use recorded, 9 % (n = 2781) had CDT use captured only in HES-APC (i.e. not captured in SACT). Women with CDT use captured only in HES-APC were more likely to be older at diagnosis, with one in five (18 %; n = 47) women aged 80 + who received CDT not captured in SACT (Table 1). Additionally, women diagnosed in 2014 were more likely to have CDT use captured only in HES-APC. Having ER-positive EIBC, HER2-negative EIBC, Grade 1 EIBC and not having adjuvant radiotherapy were also statistically associated with CDT use captured only in HES-APC, even after adjustment for each other and other factors. CDT use being captured only in HES-APC was unrelated to the underlying rate of CDT use when looking across patient subgroups.

CDT use captured only in HES-APC decreased from 21 % among women diagnosed in 2014 to 2 % among women diagnosed in 2019; this pattern was seen regardless of age. However, of women aged 80 + diagnosed in 2019 and with CDT use recorded, 9 % was captured only in HES-APC.

3.4. Information on CDT cycles

Among 23,493 women with CDT use recorded in both datasets, 88%

Table 1
Breakdown of the recording of cancer drug therapy (CDT) use in SACT or HES-APC by characteristic, among women receiving surgery for early invasive breast cancer.

Characteristic		Number of patients (column %)		Patients with CDT use recorded (row %)				
				Captured in either dataset		Of which N/% were in HES-APC alone		Adjusted p-value*
Overall		129,326		31,693	24.5 %	2781	8.8 %	
Age at diagnosis	50–59 years	41,251	31.9 %	14,926	36.2 %	1209	8.1 %	<0.0001
	60–69 years	45,802	35.4 %	11,667	25.5 %	1052	9.0 %	
	70–79 years	30,317	23.4 %	4841	16.0 %	473	9.8 %	
	80 + years	11,956	9.2 %	259	2.2 %	47	18.1 %	
Year of diagnosis	2014	21,083	16.3 %	5401	25.6 %	1121	20.8 %	<0.0001
	2015	21,612	16.7 %	5537	25.6 %	704	12.7 %	
	2016	21,661	16.7 %	5662	26.1 %	436	7.7 %	
	2017	21,369	16.5 %	5286	24.7 %	302	5.7 %	
	2018	22,274	17.2 %	5046	22.7 %	128	2.5 %	
	2019	21,327	16.5 %	4761	22.3 %	90	1.9 %	
IMD 2019	1 - Most	18,524	14.3 %	4914	26.5 %	492	10.0 %	0.2378
	2	22,434	17.3 %	5658	25.2 %	530	9.4 %	
	3	27,204	21.0 %	6714	24.7 %	599	8.9 %	
	4	29,898	23.1 %	7208	24.1 %	630	8.7 %	
	5 - least	31,266	24.2 %	7199	23.0 %	530	7.4 %	
Ethnicity	White	114,184	88.3 %	28,056	24.6 %	2518	9.0 %	0.6324
	Mixed	509	0.4 %	145	28.5 %	10	6.9%	
	Asian or Asian British	3847	3.0 %	1095	28.5 %	87	7.9 %	
	Black or Black British	1839	1.4 %	662	36.0 %	47	7.1 %	
	Other Ethnic Group	1597	1.2 %	435	27.2 %	25	5.7 %	
	Unknown	7350	5.7 %	1300	17.7 %	94	7.2 %	
ER status	Positive	104,265	80.6 %	20,350	19.5 %	1884	9.3 %	<0.0001
	Negative	14,897	11.5 %	8854	59.4 %	740	8.4 %	
	Unknown	10,164	7.9 %	2489	24.5 %	157	6.3 %	
HER2 status	Positive	22,327	17.3 %	10,630	47.6 %	836	7.9 %	<0.0001
	Negative	91,854	71.0 %	18,058	19.7 %	1703	9.4 %	
	Unknown	15,145	11.7 %	3005	19.8 %	242	8.1 %	
Invasive grade	G1	23,338	18.0 %	843	3.6 %	107	12.7 %	<0.0001
	G2	71,154	55.0 %	12,611	17.7 %	1236	9.8 %	
	G3	33,507	25.9 %	18,042	53.8 %	1423	7.9 %	
	Unknown	1327	1.0 %	197	14.8 %	15	7.6 %	
Tumour stage	T1	79, 904	61.8 %	13,016	16.3 %	1212	9.3 %	0.5409
	T2	44,237	34.2 %	16,169	36.6 %	1403	8.7 %	
	T3	5076	3.9 %	2466	48.6 %	164	6.7 %	
	Unknown	109	0.1 %	42	38.5 %	2	4.8 %	
Nodal stage	N0	95,368	73.7 %	15,941	16.7 %	1392	8.7 %	0.1945
	N +	33,652	26.0 %	15,725	46.7 %	1386	8.8 %	
	Unknown	306	0.2 %	27	8.8 %	3	11.1 %	
Charlson score	0	112,379	86.9 %	28,917	25.7 %	2506	8.7 %	0.1595
	1	11,074	8.6 %	2194	19.8 %	221	10.1 %	
	2 +	4377	3.4 %	563	12.9 %	54	9.6 %	
	Unknown	1496	1.2 %	19	1.3 %	0	0.0 %	
SCARF index	Fit	103,641	80.1 %	27,192	26.2 %	2339	8.6 %	0.1594
	Mild-Moderate	21,819	16.9 %	4271	19.6 %	419	9.8 %	
	Severe	2370	1.8 %	211	8.9 %	23	10.9 %	
	Unknown	1496	1.2 %	19	1.3 %	0	0.0 %	
Primary surgery	BCS	97,359	75.3 %	21,009	21.6 %	1849	8.8 %	<0.0001
	Mastectomy	31,967	24.7 %	10,684	33.4 %	932	8.7 %	
Adjuvant radiotherapy	No	30,284	23.4 %	4623	15.3 %	523	11.3 %	<0.0001
	Yes	99,042	76.6 %	27,070	27.3 %	2258	8.3 %	

Key: SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data; ER = estrogen receptor; HER2 = human epidermal growth receptor 2; SCARF = Secondary Care Administrative Records Frailty; BCS = breast-conserving surgery.
*grouped p-value from multilevel mixed-effects logistic regression models including all factors in the table; outcome is CDT use in SACT (with or without HES-APC) vs HES-APC only.

(n = 20,591) had the same first recorded cycle date; this percentage was largely comparable by age at diagnosis (88 % 50–69 years; 88 % 70–79 years; 91 % 80 + years) but had increased over time for all ages (Fig. 1). 8 % of first cycle dates were earlier within HES-APC whilst 4 % were earlier within SACT. The percentage with dates recorded in HES-APC first was lowest among women aged 80 + years (4 %). 98 % of first cycles with the same date in both datasets included CDT given intravenously.

Among 21,763 women with adjuvant CDT use captured in both SACT and HES-APC, 58 % had the same number of cycles reported; where neoadjuvant CDT use was captured in both sources, 68 % (n = 1917/2807) had the same number of cycles reported. Agreement between the numbers of cycles did not vary systematically according to the number of cycles (Fig. 3). Among women with CDT use recorded in both datasets, the distribution of number of cycles was similar (Fig. A3). Women with records in HES-APC only were more likely to have just one or six cycles recorded than women with records in SACT (Fig. A4).

Although the majority of CDT use recorded in SACT was delivered intravenously or subcutaneously, the route of treatment administration (oral vs intravenous/subcutaneous) appeared to differ according to whether CDT cycles were captured only in SACT or in both data sources. Among 2138 cycles recorded in SACT with only an oral CDT administered, 24 % were also captured in HES-APC. This was higher among women aged 70 + (32 %). Conversely of 298,879 cycles recorded in SACT where an intravenous/subcutaneous CDT was administered 71 % had a matching record in HES-APC. This decreased with increasing age, being 59 % among women aged 80 + . Overall there was more frequent recording of oral agents in SACT among older women.

Table 2
Agreement of the recording of cancer drug therapy (CDT) use, as identified in SACT or HES-APC, among women receiving surgery for early invasive breast cancer, broken down by age, comorbidity score and year of diagnosis.

	CDT recorded in SACT		CDT not recorded in SACT		Concordance (%)	Weighted Kappa (95% CI)
Characteristic	No. of patients	% with CDT in HES-APC	No. of patients	% without CDT in HES-APC		
All women	28,912	81.3 %	100,414	97.2 %	93.7 %	0.811 (0.807–0.815)
Age groups: 50–59 years	13,717	80.8 %	27,534	95.6 %	90.7 %	0.785 (0.778–0.791)
60–69 years	10 615	81.8 %	35,187	97.0 %	93.5 %	0.812 (0.805–0.818)
70–79 years	4368	81.3 %	25,949	98.2 %	95.8 %	0.822 (0.813–0.831)
80 + years	212	77.4 %	11,744	99.6 %	99.2 %	0.771 (0.727–0.816)
Charlson score* : 0	26,411	81.4 %	85,968	97.1 %	93.4 %	0.810 (0.806–0.814)
1	1973	80.7 %	9101	97.6 %	94.6 %	0.808 (0.794–0.823)
2+	509	80.9 %	3868	98.6 %	96.6 %	0.826 (0.799–0.853)
Year of diagnosis: 2014	4280	81.7 %	16,803	93.3 %	91.0 %	0.729 (0.717–0.740)
2015	4833	79.9 %	16,779	95.8 %	92.3 %	0.772 (0.762–0.783)
2016	5226	80.8 %	16,435	97.4 %	93.4 %	0.811 (0.802–0.821)
2017	4984	81.5 %	16,385	98.2 %	94.3 %	0.833 (0.824–0.842)
2018	4918	81.4 %	17,356	99.3 %	95.3 %	0.856 (0.847–0.864)
2019	4671	82.3 %	16,656	99.5 s %	95.7 %	0.867 (0.859–0.875)

Key: SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data.

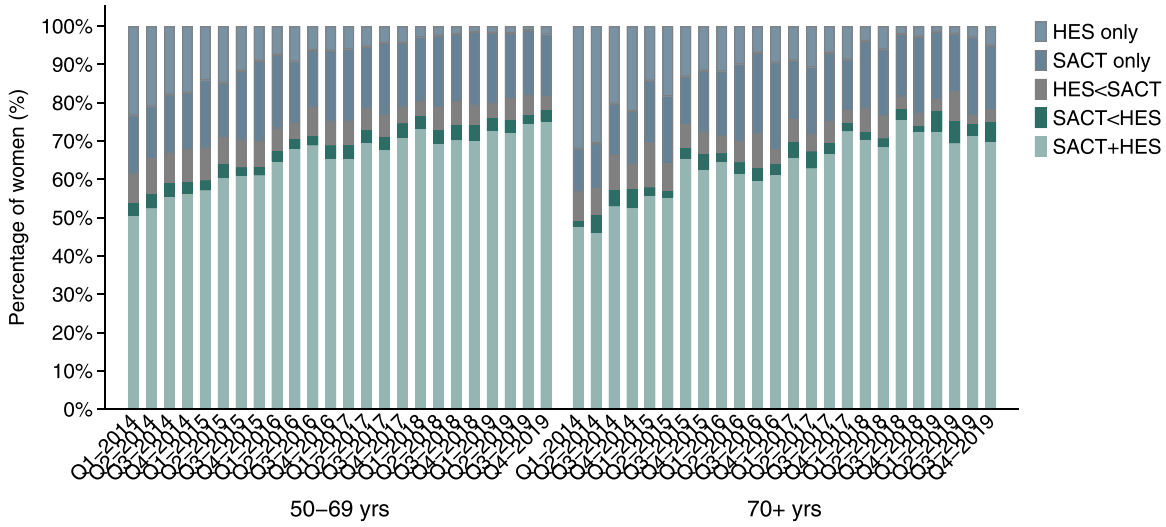


Fig. 1. Percentage with CDT details recorded in SACT or HES-APC (and agreement on first cycle date), among women receiving surgery for early invasive breast cancer, by age and date of diagnosis. **Key**: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data. SACT+HES = CDT recorded in both SACT & HES-APC; first dates match. SACT<HES = CDT recorded in both SACT & HES-APC; first date in SACT before first date in HES-APC. HES<SACT = CDT recorded in both SACT & HES-APC; first date in HES-APC before first date in SACT. SACT only = CDT recorded in SACT but not HES-APC. HES only = CDT recorded in HES-APC but not SACT.

3.5. Information on drug regimen within HES-APC

3.5.1. Cycle-level agreement of drug regimen information

Among 214,481 CDT administrations recorded in both datasets, trastuzumab, FEC/EC, paclitaxel and docetaxel were the most frequent drug regimens recorded in SACT, accounting for 83 % of administrations. Table A3 presents the OPCS procurement and delivery codes expected for each of these treatments.

In relation to the regimens specified within SACT, 88 % of trastuzumab administrations (for HER2-positive EIBC) and 94 % of FEC/EC administrations had the expected OPCS codes recorded in HES-APC (Table A3). Starting with the expected OPCS codes recorded in HES-APC, 99 % of administrations matched to a trastuzumab cycle in SACT and 92 % matched to a FEC/EC cycle in SACT. These were mostly where the SACT data identified that the drug (or combination) was given on its own. It was not possible to distinguish between administrations of FEC or EC (or its individual drug components) using just OPCS codes in HES-APC.

58 % of paclitaxel administrations and 79 % of docetaxel administrations recorded in SACT had the expected OPCS codes in HES-APC (Table A3). Starting with the expected OPCS codes in HES-APC, paclitaxel was recorded in 71 % of matched administrations in SACT and docetaxel was recorded in only 26 % of matched administrations in SACT. In the case of docetaxel just under half of these administrations included at least one other drug recorded in SACT.

3.5.2. Patient-level agreement of drug regimen information

Among 23,493 women with treatment recorded in both SACT and HES-APC, comparisons of drug regimen recorded in SACT and OPCS codes in HES-APC, to identify patient-level use of treatment, found concordance was highest for trastuzumab-based, FEC/EC-based and paclitaxel-based treatment, whilst the kappa statistics demonstrated only very good agreement for FEC/EC-based treatment (Table 3).

4. Discussion

This population-based study used linked patient-level data to compare the consistency with which CDT treatment was recorded within SACT and hospital admissions data, among more than 129,000 women (aged 50 + years) diagnosed with EIBC in England from 2014 to 2019

Fig. 2. Funnel plot showing the percentage concordance between SACT and HES-APC, among women receiving surgery for early invasive breast cancer, by diagnosing NHS trust. **Key:** CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data.

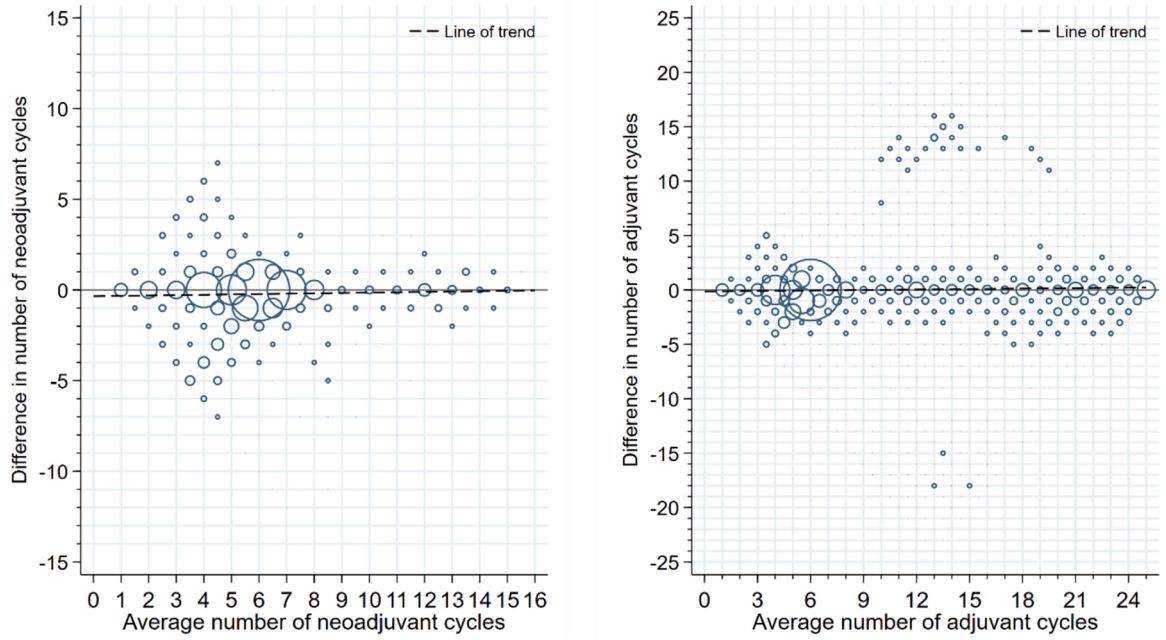


Fig. 3. Weighted scatter plot of agreement between the average numbers of CDT cycles recorded in SACT and HES-APC, and the difference in number of cycles recorded in each source, among women receiving CDT for early invasive breast cancer, by CDT setting. Key: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data. Note: the size of each data point represents the % of women in the cohort with that combination of average cycles and difference in number. Line of trend from Bland-Altman analysis.

and receiving surgery. We found that, for fact of delivery of CDT, overall agreement between SACT and HES-APC was high at 94 %. However, nearly one in ten women with a record of neoadjuvant or adjuvant CDT for EIBC were missed when just SACT data were used, with 9 % captured only in HES-APC. Although numbers were small, this increased to one in five among women aged 80 + . A potential explanation for this may be because they are more likely to have drugs delivered outside an oncological setting, which has been shown to result in poorer SACT recording [5]. As well as differences in the recording of CDT use across data sources by age, differences were observed by year of diagnosis and type of CDT administration, with comparatively poor capture of oral CDT.

SACT data returns were mandatory from April 2014, with full compliance reported from July 2014. However, lower than expected data returns for some NHS trusts may still be an issue. Analysis of HES-APC data identified an additional 20.8 % of patients diagnosed in 2014 who had received CDT, and HES-APC also identified CDT use among those NHS trusts with no/lower than expected SACT returns. These findings demonstrate the value of using both SACT and HES-APC data particularly when looking at CDT use in these patient groups and situations.

We found further differences across data sources in the recording of cycles, in coding of drug regimens and by drug administration route. Patients with CDT use only recorded in HES-APC were more likely to have just one cycle recorded, than those patients with treatment

Table 3
Agreement of drug regimen details identified either in SACT or with expected OPCS codes in HES-APC, among women receiving surgery for early invasive breast cancer with CDT use recorded in both datasets.

SACT-defined drug regimen	CDT regimen in SACT		CDT regimen not in SACT		Concordance (%)	Weighted Kappa (95% CI)
	No. of patients	% with expected OPCS codes in HES-APC	No. of patients	% without expected OPCS codes in HES-APC		
Trastuzumab *(HER2 + pts only)*	5203	92.0 %	606	63.9 %	89.0 %	0.487 (0.453–0.521)
FEC/EC	17,854	96.5%	5639	81.7%	92.9 %	0.801 (0.791–0.810)
Paclitaxel	4418	88.8 %	19,075	86.1 %	86.6 %	0.631 (0.620–0.643)
Docetaxel	8423	87.3 %	15,070	36.3 %	54.6 %	0.193 (0.184–0.202)

Key: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data; OPCS = Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures; HER2 + = human epidermal growth receptor 2 positive; pts = patients; FEC = fluorouracil, epirubicin, cyclophosphamide; EC = epirubicin & cyclophosphamide.

recorded in both SACT and HES-APC or in SACT alone. Coding of drug regimens in HES-APC is guided by standardised rules and it was difficult to accurately distinguish between different drug regimens in HES-APC, particularly where a taxane (paclitaxel or docetaxel) was given. CDT treatment for breast cancer includes multiple drug options which can be given in combination with other drugs, and so this is likely to contribute to a poor ability of HES-APC data to distinguish between drug regimens in some cases. We found use of the targeted biological therapy trastuzumab was more easily identified, with 99% of administrations with the expected OPCS codes in HES-APC matching a trastuzumab cycle in SACT. Comparison of drug administration route found that recording for drug regimens with solely oral administration was poorly identified in HES-APC, which is likely to be explained by some patients not being admitted for such drug regimens.

Comparison with other published studies evaluating agreement across the same routine data sources in different cancers, highlighted consistency of findings in relation to improvements in agreement between SACT and HES-APC over time and older patients being more likely to have CDT use captured only in HES-APC [7,8]. For this study, in breast cancer, we found higher concordance between SACT and HES-APC than reported by a previous study in colon cancer [8]. Several studies carried out in the United States comparing records across registry and claims data also found the combination of sources to be of value in identifying treatment use, noting registry data were more likely to have incomplete capture for older patients [22–25].

This study has a number of strengths. Firstly, it includes all women aged 50 years and over with a registered diagnosis of EIBC in England from 2014–2019. Secondly, both SACT and HES-APC data were available up to March 2021 giving at least 15 months of follow-up from initial diagnosis.

There are some limitations. Firstly, data were only available for women aged 50 years and older, and so it was not possible to look at the recording of CDT across these datasets in younger women (<50 years) with EIBC. Secondly, this study restricted analysis to women receiving surgery for EIBC and so it is unknown whether our findings apply to other settings and other stage groups. Finally, CDT recorded in HES outpatient data was not considered in this analysis. This may have identified further patients recorded as being treated in an outpatient setting.

The work presented in this publication was undertaken in order to inform the audit of breast cancer care in England (previously carried out as part of the NABCOP). The findings are likely to be of importance for others using routine data to look at CDT use in women with early invasive breast cancer. There are several implications for other users of the data sources which are important to highlight, but their relevance will depend on the aims of the data analysis. As a primary source of information on CDT use in routine care, the SACT dataset collects information beyond administration date and drug regimen, including drug dose, performance status through treatment, a clinical trial flag and reasons for regimen modification. This information is not collected within HES-APC and so it is important to highlight that the value of HES-APC lies in identifying additional CDT use not recorded in SACT rather than providing the full detail of this CDT use. The combination of data from HES-APC and SACT is important to understand the use of CDT in routine hospital care, particularly in those scenarios highlighted at the beginning of the discussion, when comparing across patient groups or where CDT administration is not solely oral.

5. Conclusions

Combining data from HES-APC with SACT in this cohort provided a more complete picture of the use of CDT treatment in women receiving surgery for EIBC, even among women diagnosed more recently. HES-APC may have particular value in identifying CDT use among older women, those diagnosed less recently and in NHS trusts where SACT data returns may be lower than expected. However, its value is limited for identifying oral CDT use. Rationalisation of routine cancer data collection within and between countries and across different health care systems is an important objective to simplify future analyses of care and outcomes with the objective of improving population health. The historic and current use of different systems to contemporaneously record the same treatment intervention presents complexity, but is an inevitable consideration for the analyses of care delivered in the past and does allow for improved data completeness and an opportunity for quality assurance. Current efforts should continue to improve SACT completeness, but the addition of HES-APC is currently necessary to provide a more complete picture on the use of CDT and is particularly helpful in the assessment and analysis of variation in care of patient subgroups and where an individual trust SACT return is deficient. At the core of service evaluation is understanding what happens in practice and the accurate capture of data is crucial to ensure services have confidence in evaluation findings to support local quality improvement and the delivery of better care to patients.

Ethics approval and consent to participate

This study was exempt from NHS Research Ethics Committee approval because it involved analysis of pseudonymised linked data collated for the purpose of service evaluation as part of the National Audit of Breast Cancer in Older Patients.

Consent for publication

Not applicable.

Funding

This study was undertaken as part of the work by the National Audit of Breast Cancer in Older Patients (NABCOP). The NABCOP is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit and Patient Outcomes Programme, and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). Neither the commissioner nor the funders had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. The authors had full independence from the HQIP. The aim of the NABCOP is to evaluate the care of older women with breast cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these women. More information can be found at www.nabcop.org.uk.

DD also receives funding from Cancer Research UK (grant C8225/A21133). Cancer Research UK had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

CRediT authorship contribution statement

Guarantor of integrity of the entire study: **MRG**. Study concepts and design: **MRG, MHP, KM, DD, KH, KC, JM, DAC**. Literature research: **MRG**. Clinical studies: N/A. Experimental studies / data analysis: N/A. Data acquisition: **MRG, KC, JM**. Statistical analysis: **MRG**. Manuscript preparation: MRG. Manuscript editing: **MRG, MHP, KM, DD, KH, KC, JM, DAC**. All authors were involved in data interpretation, critical appraisal of the draft manuscript, and gave final approval on the submitted version.

Declaration of Competing Interest

None.

Data availability

This work uses data that has been provided by patients and collected by the NHS as part of their care and support. The data for England are collated, maintained and quality assured by the National Disease Registration Service (NDRS), which is part of NHS Digital. Data on English Cancer Registrations can be accessed via the NHS Digital Data Access Request Service (DARS) https://digital.nhs.uk/services/data-access-request-service-dars#national-disease-registration-service-ndrs-.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.canep.2023.102337.

References

[1] NICE, Early and locally advanced breast cancer: diagnosis and management.2018. www.nice.org.uk/guidance/ng101.
[2] E.B. Ludmir, W. Mainwaring, T.A. Lin, A.B. Miller, A. Jethanandani, A.F. Espinoza, et al., Factors associated with age disparities among cancer clinical trial participants, JAMA Oncol. (2019).
[3] L.F. Hutchins, J.M. Unger, J.J. Crowley, C.A. Coltman Jr., K.S. Albain, Underrepresentation of patients 65 years of age or older in cancer-treatment trials, N. Engl. J. Med. 341 (27) (1999) 2061–2067.
[4] D. Dodwell, R. Shakir, Assessing new drugs in advanced cancer: beyond randomised evidence, Clin. Oncol. (R. Coll. Radiol.) 33 (4) (2021) e201-e2.
[5] C.J. Bright, S. Lawton, S. Benson, M. Bomb, D. Dodwell, K.E. Henson, et al., Data resource profile: the systemic anti-cancer therapy (SACT) dataset, Int J. Epidemiol. 49 (1) (2020), 15-l.
[6] M. Wallington, E.B. Saxon, M. Bomb, R. Smittenaar, M. Wickenden, S. McPhail, et al., 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study, Lancet Oncol. 17 (9) (2016) 1203–1216.
[7] L. McDonald, C. Sammon, R. Carroll, A., C.A.S. Harish, D. Tyas, et al., Consistency of recording of chemotherapy cycles in the National Cancer Registration and Analysis Service Systemic Anti-Cancer Therapy database and the Hospital Episode Statistics Admitted Patient Care database, Future Oncol. 16 (3) (2020), 4455-60.
[8] J.M. Boyle, A. Kuryba, M.S. Braun, A. Aggarwal, J. van der Meulen, T.E. Cowling, et al., Validity of chemotherapy information derived from routinely collected healthcare data: a national cohort study of colon cancer patients, Cancer Epidemiol. 73 (2021), 101971.
[9] A. Herbert, L. Wijlaars, A. Zylbersztejn, D. Cromwell, P. Hardelid, Data resource profile: hospital episode statistics admitted patient care (HES APC). Int J. Epidemiol. 46 (4) (2017), 1093-i.
[10] D. Dodwell, Y. Jauhari, T. Gathani, D. Cromwell, M. Gannon, K. Clements, et al., Treatment variation in early breast cancer in the UK, BMJ 371 (2020) m4237.
[11] M.R. Gannon, D. Dodwell, Y. Jauhari, K. Horgan, K. Clements, J. Medina, et al., Initiation of adjuvant chemotherapy and trastuzumab for human epidermal growth receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England, J. Geriatr. Oncol. 11 (5) (2020), 836-42.
[12] M.G.M. Derks, E. Bastiaannet, M. Kiderlen, D.E. Hilling, P.G. Boelens, P.M. Walsh, et al., Variation in treatment and survival of older patients with non-metastatic breast cancer in five European countries: a population-based cohort study from the EURECCA Breast Cancer Group, Br. J. Cancer 119 (1) (2018) 121–129.
[13] National audit of Breast Cancer in Older Patients (NABCOP). NABCOP , 2022. Annual Report2022 12/05/2022. https://www.nabcop.org.uk/reports/nabcop-2022-annual-report/.
[14] J.N. Armitage, J.H. van der Meulen, Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score, Br. J. Surg. 97 (5) (2010) 772–781.
[15] Y. Jauhari, M.R. Gannon, D. Dodwell, K. Horgan, K. Clements, J. Medina, 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 10 (5) (2020), e035395.
[16] National Cancer Registration and Analysis Service. Matching SACT to Cancer Waiting Times Data (2014). http://www.ncin.org.uk/publications/data_briefings/sact_cwt.
[17] Health and Social Care Information Centre. Chemotherapy Regimens Clinical Coding Standards and Guidance OPCS-4 (2014). http://www.nwisinformationstandards.wales.nhs.uk/sitesplus/documents/299/ChemoRegClinCodStanGuidOPCS-4v1.0.pdf.
[18] Health and Social Care Information Centre. Chemotherapy Regimens Clinical Coding Standards and Guidance OPCS-4 April 2017 (2017).
[19] K.E. Henson, A. Fry, G. Lyratzopoulos, M. Peake, K.J. Roberts, S. McPhail, Sociodemographic variation in the use of chemotherapy and radiotherapy in patients with stage IV lung, oesophageal, stomach and pancreatic cancer: evidence from population-based data in England during 2013-2014, Br. J. Cancer 118 (10) (2018), 1382-90.
[20] All-Party Parliamentary Group on Breast Cancer, A Mixed Picture: an Inquiry into Geographical Inequalities and Breast Cancer (2018). https://breastcancernow.org/sites/default/files/appgbc_a_mixed_picture.pdf.
[21] J.M. Bland, D.G. Altman, Statistical methods for assessing agreement between two methods of clinical measurement, Lancet 1 (8476) (1986) 307–310.
[22] C.J. Bradley, R. Liang, J. Jasem, R.C. Lindrooth, L.M. Sabik, M.C. Perraillon, Cancer treatment data in central cancer registries: when are supplemental data needed? Cancer Inf. 21 (2022), 11769351221112457.
[23] G.S. Cooper, Z. Yuan, K.C. Stange, L.K. Dennis, S.B. Amini, A.A. Rimm, Agreement of Medicare claims and tumor registry data for assessment of cancer-related treatment, Med. Care 38 (4) (2000) 411–421.
[24] A.M. Noone, J.L. Lund, A. Mariotto, K. Cronin, T. McNeel, D. Deapen, et al., Comparison of SEER treatment data with medicare claims, Med. Care 54 (9) (2016) e55–e64.
[25] J.L. Warren, L.C. Harlan, A. Fahey, B.A. Virnig, J.L. Freeman, C.N. Klabunde, et al., Utility of the SEER-Medicare data to identify chemotherapy use, Med. Care 40 (8 Suppl) (2002) 55–61.

4.2. Research Paper 2

Title

Completeness of endocrine therapy information in the Primary Care Prescription Database (PCPD) and secondary care treatment datasets: a national population-based cohort study using routine healthcare data.

Journal details

Published in Cancer Epidemiology. The online PDF version can be accessed at:
https://pubmed.ncbi.nlm.nih.gov/37473577/

Supplementary material can be found in the Appendices.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Cancer Epideimiology		
When was the work published?	18 July 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	[redacted]
Date	

Cancer Epidemiology 86 (2023) 102423

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Completeness of endocrine therapy information in the Primary Care Prescription Database (PCPD) and secondary care treatment datasets: A national population-based cohort study using routine healthcare data

Melissa Ruth Gannon [a,b,*], David Dodwell [c], Katie Miller [a,b], Kieran Horgan [d], Karen Clements [e], Jibby Medina [a,b], Min Hae Park [a,b], David Alan Cromwell [a,b]

[a] Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK
[b] Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK
[c] Nuffield Department of Population Health, University of Oxford, Oxford, UK
[d] Department of Breast Surgery, St James's University Hospital, Leeds, UK
[e] National Cancer Registration and Analysis Service, NHS Digital, 2nd Floor, 23 Stephenson Street, Birmingham, UK

ARTICLE INFO

Keywords:
Invasive breast cancer
Endocrine therapy
Primary care
Secondary care
Older patients

ABSTRACT

Background: Endocrine therapy (ET) is a widely used treatment for breast cancer. In the UK, use is typically initiated in secondary care, with subsequent treatment in primary care. Evaluating use of ET depends on data sources containing accurate and complete information. This study aimed to evaluate the completeness and consistency of ET recorded in primary and secondary care data (SCD) and determine the value of combining data sources in describing use of ET.

Methods: This cohort study included women (50 + years) diagnosed with hormone receptor-positive invasive breast cancer in England, April-2015 to December-2019. Concordance of ET recorded in SCD and the Primary Care Prescription Database (PCPD) was evaluated. Factors associated with recording of ET in each setting were assessed using statistical models.

Results: Overall 110,529 women were included. 94% had ET recorded in either SCD or PCPD. ET captured in SCD varied from 3% (in Systemic Anti-Cancer Therapy data) to 52% (in the Cancer Outcomes and Services Dataset; COSD). By contrast, 93% of patients had an ET prescription in PCPD. Among patients with ET recorded, this was not captured in COSD for 45%. Capture in COSD was lowest for younger women, those with no comorbidity/frailty, with lower stage or HER2-positive disease, or with other treatments recorded. Overall concordance between COSD and PCPD was 57%, but varied substantially across NHS trusts (lowest decile≤28%; highest decile≥86%). Among women with ET recorded in both settings, the earliest record was in COSD for 97%; 59% of initial ET prescriptions recorded in COSD were not captured in PCPD. Combining PCPD and COSD data enabled estimation of ET duration.

Conclusions: PCPD is vital for understanding the use of ET within this population. Completeness of SCD could be improved by ensuring information on first ET prescription is recorded. PCPD (linked to SCD) is a valuable resource for examining patterns of care for patients with cancer, including treatment duration and adherence.

1. Introduction

National guidelines recommend treatment with surgery and adjuvant endocrine therapy (ET; for 5 years) for estrogen receptor (ER)-positive early invasive or locally advanced breast cancer (EIBC/LABC) and first-line treatment with ET for the majority of patients with ER-positive metastatic breast cancer (MBC) [1,2]. Guidelines also advise that women newly-diagnosed with breast cancer are given the option of primary ET (PET) instead of both surgery and ET, when anticipated life expectancy is short [3]. Despite ET being widely prescribed, little is known about patterns of use among women with breast cancer in England, nor how well such patterns are captured in national cancer data.

In the UK, although ET for ER-positive breast cancer is initiated in secondary care, subsequent treatment is often prescribed within primary

* Correspondence to: London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH UK.
E-mail address: melissa.gannon@lshtm.ac.uk (M.R. Gannon).

https://doi.org/10.1016/j.canep.2023.102423

Available online 18 July 2023

care. Patient-level data on oncological treatments, including ET, delivered in secondary care are collected in multiple national healthcare datasets, including cancer registration datasets and the Systemic Anti-Cancer Therapy (SACT) database. Within primary care, the Primary Care Prescription Database (PCPD) captures all prescriptions dispensed by community pharmacies in England, with population coverage [4]. There is no gold standard data source on ET, however as ET can be delivered in secondary and primary care settings, comparison of information in secondary care datasets and PCPD could highlight where completeness in either dataset is lower than expected. Given the typical ET prescribing patterns, combining data from both settings has the potential to provide a more complete picture of ET use.

Previous publications describing levels of ET use in England used only cancer registry or routine secondary care data and none included data from primary care [5–10]. To date only two journal publications have used PCPD data, using four-months of data, linked to cancer registry data, as access has previously been limited [4,11]. They highlight the value of PCPD in understanding the patient pathway, but neither study compared ET recording across PCPD and secondary care data.

This study aimed to evaluate the completeness and consistency of ET information recorded in PCPD and secondary care data, and determine the value of combining information from these data sources in describing use of ET, for women aged 50 + years newly-diagnosed with invasive breast cancer in England from April-2015 to December-2019. We hypothesised that each dataset would capture different aspects of ET use, and that using both data sources would provide a more comprehensive picture of treatment. We also aimed to look at factors associated with recording of ET in secondary care data to understand which patient groups might be underrepresented if only secondary care data are used to ascertain use of ET.

The study was designed to evaluate agreement between different care settings on recording of ET and identify the benefits and limitations of each data source when attempting to define use of ET (Yes/No), date of ET initiation, type and duration.

The study is reported according to the RECORD extension to STROBE guidelines for observational studies using routinely collected data [12].

2. Methods

2.1. Data sources

This population-based cohort study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP). Full details of the NABCOP can be accessed via www.nabcop.org.uk. Pseudonymised Cancer Registration records were provided for women aged 50 + years, newly-diagnosed with breast cancer within an NHS trust in England, between 1-January-2014 and 31-December-2019 [13]. These data were provided by the National Cancer Registration and Analysis Service (NCRAS), who linked registration records at tumour-level to (i) the Cancer Outcomes and Services Dataset (COSD); (ii) SACT records from January-2014 to October-2021; (iii) Cancer Waiting Times (CWT) treatment data records from January-2014 to December-2020; (iv) national Radiotherapy Dataset (RTDS) records from January-2014 to April-2021; and (v) Hospital Episode Statistics Admitted Patient Care (HES-APC) data from two years prior to a patient's date of diagnosis up to March-2021 [14,15]. Registration records were linked at patient-level to PCPD records of ET prescriptions dispensed from April-2015 to March-2021.

2.1.1. Defining and identifying ET in each data source

Information on ET for breast cancer is collected in several secondary care datasets and PCPD. The data items collected within each varies (Table A1) but each collects an instance of ET and date.

COSD collects data on treatment modality (e.g. surgery, chemotherapy, radiotherapy, ET) and start date. CWT treatment data contain only first treatment received. Cancer Registry data include basic information on receipt of ET, date, drug name (91% missing), and are predominantly collated from COSD data at the point of cancer registration. SACT contains longitudinal data (including drug name, dose, administration dates), on prescribed SACT, including ET [15,16].

PCPD has population coverage for drug therapies dispensed in community pharmacies within England [4,11]. PCPD includes longitudinal data on Virtual Medicinal Product (VMP) name, dose, month and year dispensed.

We compiled a list of ET used for treatment of hormone receptor-positive invasive breast cancer in England, based on clinical input, to identify ET in data sources where drug/VMP names are completed (SACT, PCPD). This list was used by NCRAS to extract and provide information on ET prescriptions in PCPD from the VMP name field, whilst we used the list to identify ET in SACT from the drug name field. The list of ET identified in SACT and PCPD is provided in Table A2. For all other data sources, only fact of ET and date were available.

2.2. Study population

The study included women aged 50 + years diagnosed with hormone receptor-positive invasive breast cancer (Stage 1–4) from 1-April-2015–31-December-2019. Breast cancer was classified as hormone receptor-positive if either ER or progesterone receptor status were reported as positive. The following exclusion criteria were applied: (i) women with ductal carcinoma in situ or unknown stage, (ii) diagnosis date prior to April-2015, (iii) unknown, borderline or negative hormone receptor status, (iv) death date prior to the first PCPD prescription date. The start of follow-up was the date of breast cancer diagnosis. The end of follow-up was the earliest of date of death or March-2021.

2.3. Study outcome

The study outcome was record of ET in each dataset and concordance of ET records across datasets. As ET is recommended as part of initial treatment following a diagnosis of breast cancer, the study looked at ET within 12 months of diagnosis. For EIBC/LABC, type of ET (neoadjuvant, adjuvant, PET) was defined within each dataset based on month of earliest ET record in that dataset and month of surgery (where applicable) to understand the information provided in each setting as ET can be initiated at different points in relation to surgery. Type of ET was classified as: neoadjuvant if earliest ET record was within eight months prior to surgery; adjuvant if earliest ET record was within 12 months of diagnosis but after surgery; and PET where use was recorded with no record of surgery or where surgery was recorded more than eight months after earliest ET record. For PCPD if the earliest record of ET was in the same month as surgery it was not possible to know if ET started before or after surgery and so this was classified as "same month as surgery".

2.4. Study exposure variables

The following variables were used to describe the cohort and examine which factors were associated with differences in ET recording across data sources. Age at diagnosis (years), disease stage (EIBC=1–3a; LABC=3b/c; MBC=4), tumour stage (T1–4), nodal stage (N0, N +), HER2 status (positive, borderline, negative), tumour grade (G1–3) and deprivation. Deprivation was measured using the Index of Multiple Deprivation 2019 rank, where each patient is allocated to national quintiles of deprivation (from most deprived = 1 to least deprived = 5) based on their postcode at diagnosis. This allocation was done by NCRAS and only the calculated deprivation quintile provided within the data. All variables were provided within Cancer Registry/COSD.

Comorbidity and patient fitness were also calculated. ICD-10 codes recorded in HES-APC within two years prior to diagnosis were used to determine comorbidity burden (0, 1, 2 +; defined according to the Royal College of Surgeons' of England Charlson Comorbidity Index) and frailty

(fit, mild-moderate frailty, severe frailty; defined using the Secondary Care Administrative Records Frailty index) [17,18].

Primary surgery was defined as either breast-conserving surgery (BCS) or mastectomy occurring within six months of diagnosis, identified within HES-APC. Use of radiotherapy was identified based on records within the RTDS. Use of chemotherapy was identified from SACT.

2.5. Statistical analysis

All data preparation and statistical analyses were conducted using Stata version 17.0 [19].

Descriptive statistics (numbers, percentages) summarise the characteristics of women with ET recorded overall and in each dataset. Bar charts display recorded ET use by age and disease stage. Kaplan-Meier plots were used to show time to earliest ET record in each dataset, by disease stage and type of ET (for EIBC/LABC). Median time from diagnosis to earliest ET record was calculated based on months, due to PCPD not containing information on day.

Contingency tables were used to explore patient-level agreement between data settings with respect to record of ET.

	Women with ET recorded in PCPD	Women with no ET recorded in PCPD
Women with ET recorded in secondary care data	A	B
Women with no ET recorded in secondary care data	C	D

Concordance was defined as the percentage of women with agreement about ET (Yes/No) in PCPD and secondary care data [(A+D)/(A+B+C+D)]. Funnel plots were used to assess variation in concordance by NHS trust. For this analysis patients were grouped based on their NHS trust of diagnosis. NHS trust of (a) surgery or (b) MDT was used where no trust of diagnosis was recorded. We also calculated weighted kappa-statistics to describe the strength of agreement between recording of ET in PCPD and secondary care data, and assess whether agreement was beyond that expected by chance alone. The kappa-statistic has a maximum of 1 and values < 0.21 are considered to show poor agreement.

Patient and tumour characteristics, record of surgery (and type), radiotherapy, and chemotherapy were described for women with ET captured in either secondary care data or PCPD. Additionally the percentage of women who had ET recorded only in PCPD was described for each patient factor to understand poor recording in the secondary care data, with higher percentages indicating poor recording in the secondary care data whilst a lower percentage indicated better recording in secondary care data. Multilevel mixed-effects logistic regression models were used to assess the likelihood that these factors were associated with poor recording in secondary care data (outcome: ET recorded in secondary care data vs in PCPD only), accounting for the clustering of patients within NHS trust. NHS trust was included as a random intercept, which estimates differences in the baseline percentage of women with a record in PCPD only between trusts, not explained by factors in the model. Missing information for patient/tumour factors was included as "unknown" categories to contribute to understanding the likelihood of ET recording only in PCPD, with complete case analysis conducted as a sensitivity analysis. Methods accounting for the clustered nature of the data were used to account for patients diagnosed within the same NHS trust being likely to be treated more similarly than those in other NHS trusts.

For patients with ET recorded in both secondary care data and PCPD, timing (month and year) of earliest ET record was compared to establish sequencing of ET recording. A patient-level comparison of month of ET initiation across datasets was conducted for each type of ET (neoadjuvant, same month as surgery, adjuvant, PET), among patients with EIBC/LABC. Differences by age were visually explored using bar charts.

The percentage of women with multiple ET records in each dataset, along with information on which dataset picked up first record, were investigated to describe patterns of ET initiation and duration captured using only one dataset or by combining information across datasets. To understand whether combining data across settings provides estimates of ET duration in line with clinical expectation, treatment duration was calculated only among women diagnosed from April-December 2015 and alive at 1-April-2021, who were alive long enough to have received 5 years (60 months) of ET.

3. Results

Of 138,973 women aged 50 + years diagnosed with invasive breast cancer in England from 1-April-2015–31-December-2019, 80% (N = 110,529) had hormone receptor-positive disease (Figure A1). These 110,529 patients formed the main study cohort. Median follow-up was 50 months (interquartile range: 36–64).

3.1. ET recorded in SCD

ET was recorded within 12 months of diagnosis in secondary care data for 53% (n = 58,296/110,529) of all women in the cohort. The recording of ET varied greatly across datasets: SACT 3%, CWT 17%, Cancer Registry 52%, COSD 52%. There was considerable overlap of ET recording, with ET recorded in CWT and SACT being a subset of ET recorded in Cancer Registry and COSD, regardless of age (Figure A2). As COSD had the highest and most complete capture of ET in secondary care data, only this dataset was used for the remainder of the analysis. Only 210 women (0.2%) had first ET recorded more than 12 months after diagnosis.

ET recording in COSD was associated with age (Fig. 1a), increasing from 45% among women aged 50–69 years to 76% among women aged 80 + years, and disease stage (EIBC 52%; LABC 60%; MBC 64%).

Among women with EIBC/LABC with ET recorded, type of ET was classified as neoadjuvant for 11%; this percentage was similar across age groups (11% 50–69 years; 14% 70–79 years; 10% 80 + years). Overall, 67% of ET initiations were classified as adjuvant while 22% were PET. The percentage of patients with ET initiation classified as adjuvant decreased with older age (84% 50–69 years; 68% 70–79 years; 26% 80 + years), whereas PET increased with age (5% 50–69 years; 18% 70–79 years; 64% 80 + years). Type of ET also differed by disease stage (Fig. 1a), with higher rates of PET for LABC (51%).

3.2. ET recorded in PCPD

ET was recorded within 12 months of diagnosis in PCPD for 93% (n = 103,076/110,529) of all women in the cohort. Recording was largely consistent across ages but was lower for MBC (94% EIBC; 91% LABC; 79% MBC-Fig. 1b). Only 1% (n = 1089) had ET recorded more than 12 months after diagnosis.

Among women with EIBC/LABC with ET recorded, type of ET was classified as neoadjuvant for 6%, with low rates for all ages (4% 50–69 years; 8% 70–79 years; 7% 80 + years). ET was classified as adjuvant (after the month of surgery) in 76% of women, but this decreased with age (87% 50–69 years; 75% 70–79 years; 36% 80 + years); a further 5% of ET prescriptions were dispensed in the same month as surgery (similar for all ages). 13% of ET was PET, and this increased with age (4% 50–69 years; 11% 70–79 years; 52% 80 + years). Type of ET also differed by disease stage (Fig. 1b), with higher rates of PET for LABC (35%).

3.3. Comparing ET recorded in secondary care data or PCPD

Among the full cohort of women, 94% (n = 104,389) had ET recorded in either COSD or PCPD; this percentage was high regardless of age or other characteristics (Fig. 2/Table A3). There were 56,531 women with ET recorded in both COSD and PCPD, while 6140 had no ET

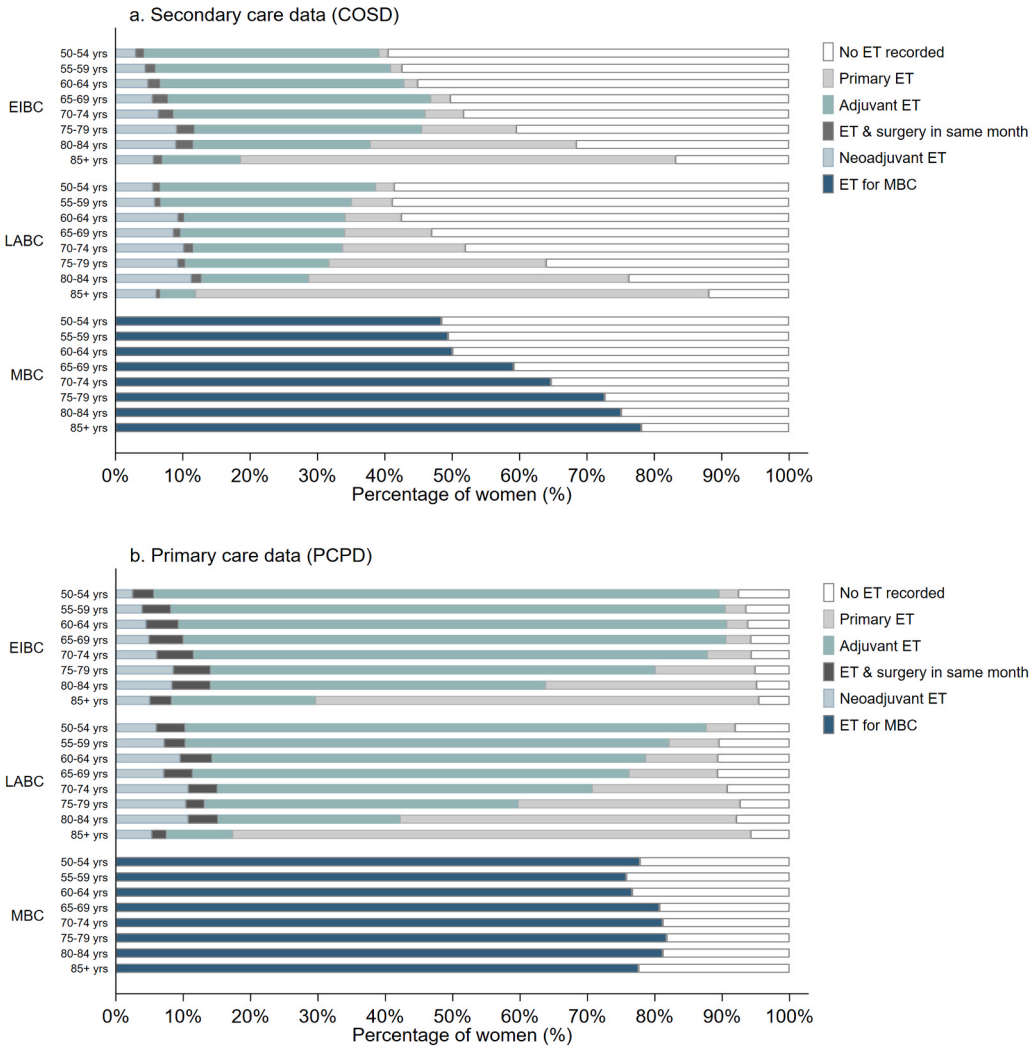


Fig. 1. Type of endocrine therapy (ET) recorded in each data source, by age at diagnosis and disease stage, among all women with hormone receptor-positive invasive breast cancer. Key: COSD = Cancer Outcomes and Services Dataset; PCPD = Primary Care Prescription Database; EIBC = Early invasive breast cancer; LABC = Locally advanced breast cancer; MBC = Metastatic breast cancer.

recorded in either COSD or PCPD. This produced an overall level of concordance of 57% (weighted kappa=0.10).

Of 104,389 women with ET recorded, the percentage with records in COSD alone was negligible (n = 1313; 1.3%). In contrast, 45% (n = 46,545) had ET recorded only in PCPD. Within Fig. 2 and Table A3, the percentage of women with ET recorded only in PCPD (45%) and the percentage with ET recorded in either COSD or PCPD (94%) are compared; patterns of recorded ET across different factors can be seen, in relation to the percentage only in PCPD as well as differences across variable levels. Women with ET in PCPD alone (and not recorded in COSD) were more likely to be younger, have lower comorbidity burden/frailty, less advanced disease, HER2-positive disease, increasing grade, or a record of other treatment (surgery, radiotherapy, chemotherapy). Differences by patient factors in the likelihood of ET being captured only in PCPD remained statistically meaningful after adjustment for other factors. Excluding women with "unknown" information for factors in the model did not change this finding.

Patterns of agreement varied substantially by NHS trust. Overall concordance was 28% for the 10% of NHS trusts with lowest concordance values, and 86% among the 10% with highest values. Out of 117 NHS trusts covered by the analysis, 54% (n = 63) had less than 60% concordance (Fig. 3).

3.4. Comparison of time to first ET record across secondary care data and PCPD

Among 56,531 women with ET recorded in both COSD and PCPD, the earliest record was in COSD for 59% (n = 33,634). By contrast, the earliest record was in PCPD for 3% (n = 1608). For 38% (n = 21,289) the earliest ET record was in the same calendar month in COSD and PCPD.

Median time from diagnosis to ET recorded in COSD was around 2 months, with shorter times for more advanced breast cancer (<1 month). Median time from diagnosis to ET recorded in PCPD was typically longer, at around 3 months (Figure A3).

3.5. Comparison of type of ET for EIBC/LABC across secondary care data and PCPD

Among the full cohort of 106,028 women with EIBC/LABC, the main reason for lack of agreement was the larger percentage of COSD records missing information on ET (Table A4). Among patients with ET recorded in PCPD, COSD records missed the largest percentage of information for ET classified as adjuvant (54%; n = 41,129/76,094) or in the same month as surgery (45%; n = 2177/4882). However, restricting to records where both PCPD and COSD contained ET information, we found

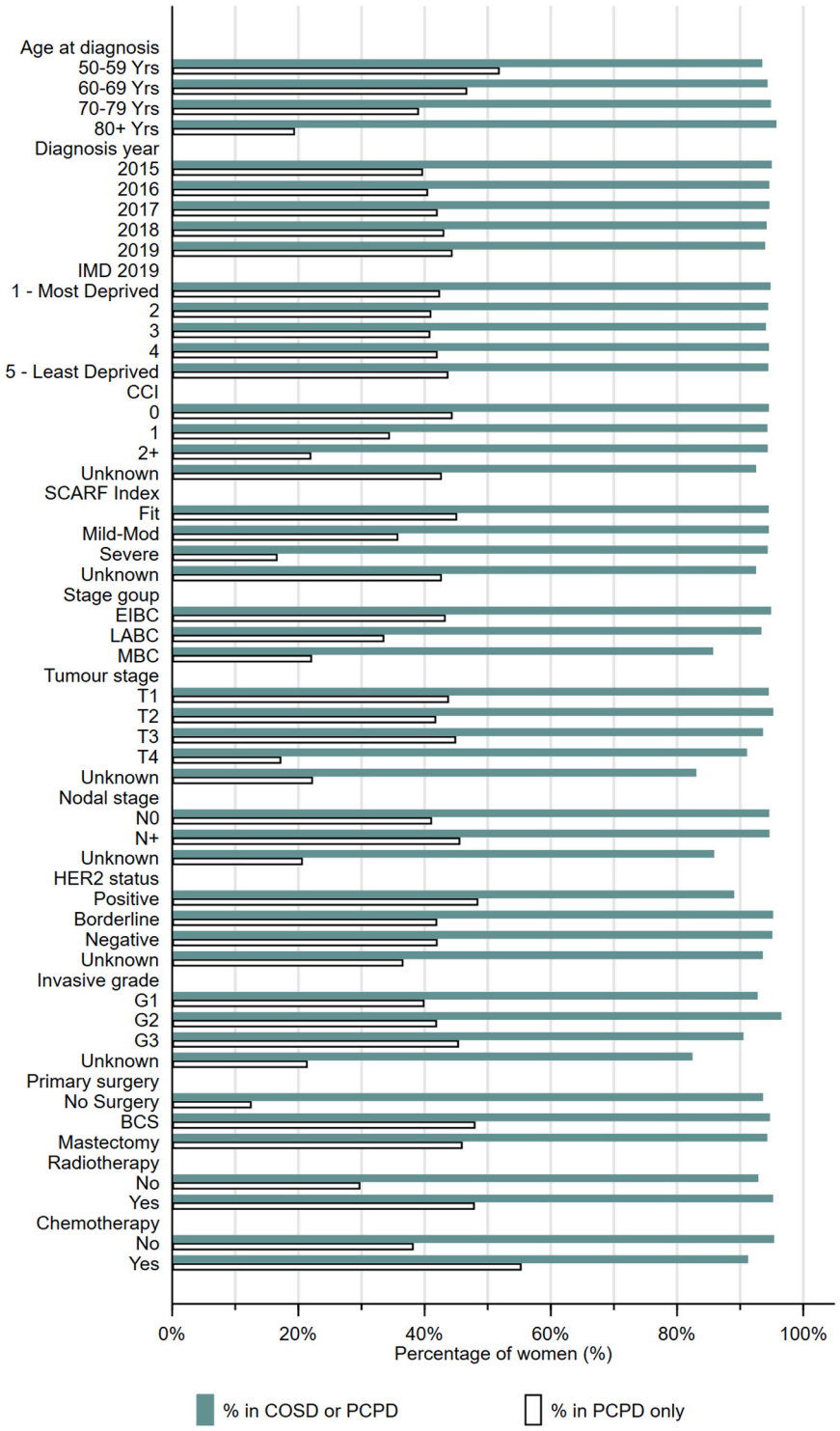


Fig. 2. Recording of endocrine therapy across COSD and PCPD, by patient subgroups, among all women with hormone receptor-positive invasive breast cancer. Key: COSD = Cancer Outcomes and Services Dataset; PCPD = Primary Care Prescription Database; IMD = Index of Multiple Deprivation; CCI = Charlson Comorbidity Index; SCARF = Secondary Care Administrative Records Frailty; Mild-Mod = Mild to moderate; EIBC = Early invasive breast cancer; LABC = Locally advanced breast cancer; MBC = Metastatic breast cancer; HER2 = Human epidermal growth receptor 2; BCS = Breast-conserving surgery.

that of records classified as adjuvant in either source 97% were classified as adjuvant in both COSD and PCPD, whilst for records classified as PET in either source 99% were classified as PET in both COSD and PCPD. This was lower where ET was classified as neoadjuvant (72% of records agreed on type where classified as neoadjuvant in either source) or in the same month as surgery (55%).

The percentages of women receiving each type of ET varied with age however the overall level of "disagreement" by age did not differ greatly when both datasets contained information on ET (Fig. 4). The dominant source of "disagreement" in type of ET was missing information within COSD, and because the percentage of missing data decreased with increasing age the overall level of agreement increased with increasing age.

3.6. Patterns of ET duration

Information on ET within secondary care data was found to typically document start of treatment only, with 3% (n = 1691) of patients with ET recorded in COSD having more than one instance recorded. In contrast, if a patient had information on ET in PCPD, 99% had more than

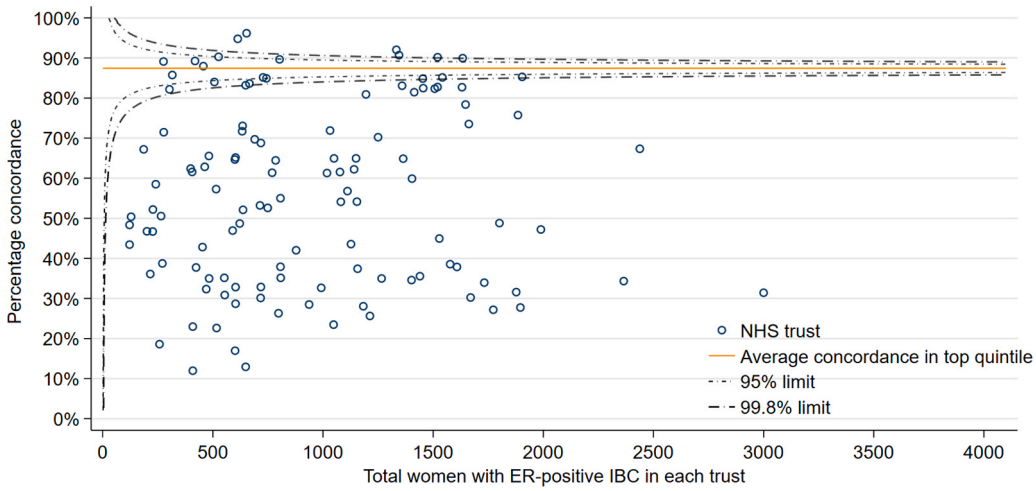


Fig. 3. Funnel plot showing the percentage concordance in recording of endocrine therapy between COSD and PCPD, by diagnosing NHS trust, among all women with hormone receptor-positive invasive breast cancer.

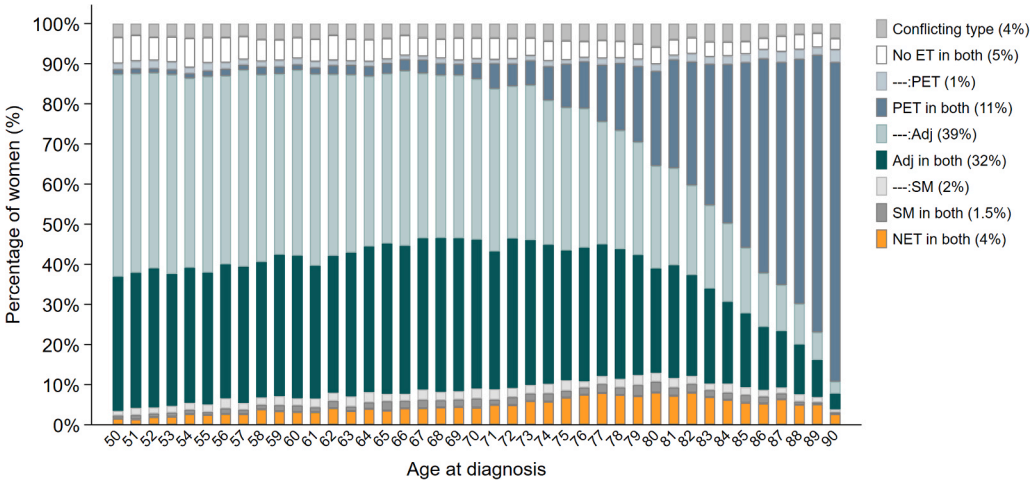


Fig. 4. Type of endocrine therapy (ET) for hormone receptor-positive EIBC/LABC, based on timing of the earliest ET record in secondary care data (COSD) or PCPD, by age at diagnosis, among all women. **Key:** ET = endocrine therapy; EIBC = early invasive breast cancer; LABC = locally advanced breast cancer; NET = neoadjuvant endocrine therapy; SM = Same month as surgery; Adj = adjuvant endocrine therapy; PET = primary endocrine therapy; No ET in both = no endocrine therapy recorded in COSD or PCPD. Colour labels: The first listed type of ET is based on earliest ET record in COSD and the second is based on earliest ET record in PCPD (e.g. "NET:ADJ" = ET before surgery in COSD but ET after surgery in PCPD). "Conflicting type" is where type of ET differs in COSD and PCPD or ET is only recorded in COSD/PCPD; all are categories with < 1%: PET in COSD only = 0.2%; Adj in COSD only = 0.7%; NET in PCPD only = 0.7%; PET in COSD, NET in PCPD = 0.1%; Adj in COSD, NET in PCPD = 0.2%.

one dispensed prescription captured.

Comparing earliest ET dates among patients with information in both COSD and PCPD confirmed COSD tends to capture ET initiation, with 97% (n = 54,923/56,531) of first recorded prescriptions captured in COSD. Specifically, the earliest month of ET was in COSD only for 59% (n = 33,634), in both COSD and PCPD for 38% (n = 21,289), and in PCPD alone for 3% (n = 1608).

A final analysis focused on women diagnosed in 2015, who could have received 5 years (60 months) of ET, to evaluate whether use of PCPD alone or combining data from across settings provides estimates of ET duration in line with clinical expectations. Among women with ET recorded in PCPD (n = 13,109), median ET duration was 60 months (interquartile range: 56–62), with little difference by disease stage (EIBC/LABC/MBC). Including available information from COSD records, median duration remained the same, whilst the interquartile range increased by one month (57–63 months).

4. Discussion

This population-based study used linked patient-level data to understand the completeness and consistency of ET recording within secondary care data and the PCPD, among 110,529 women (aged 50 + years) diagnosed with hormone receptor-positive invasive breast cancer in England from April-2015 to December-2019.

Among the various secondary care datasets that capture ET, COSD had the highest level of ascertainment. Nearly all patients (94%) had ET recorded in either COSD or PCPD. This is in line with clinical expectation, as some patients may choose not to have ET even if it was recommended. However, overall agreement on ET across datasets was poor. This was primarily due to ET not being recorded in COSD for nearly half of women (45%) with ET identified in PCPD. The predominance of care and initial treatment of patients with cancer is carried out within secondary care, with the decision to use ET being made in this setting. From this perspective we would expect any patient with ET captured in PCPD to also have a record in COSD, however this was not the case for 45%, suggesting that COSD data are incomplete on this aspect of treatment. We found 1.3% of women had ET recorded only in COSD. It is unclear why this might be the case but some potential reasons could be that their ET was always provided within secondary care, whilst some may have died whilst in hospital, may not have tolerated ET or may have discontinued for another reason.

Where ET was recorded in both datasets we found that COSD

demonstrated value in providing earliest ET date for around two-thirds (59%) of patients. This was not captured in PCPD records, which limits its utility for estimating time to starting ET. Conversely the lack of information in COSD beyond ET initiation means the most accurate figures on ET duration or adherence would be derived when combining PCPD and COSD because of the tendency for the earliest records to be captured only in COSD.

Comparison of ET records across care settings served to highlight that information from PCPD is vital in identifying and describing ET in this population. Although secondary care data were most complete for those patient groups where ET is given without surgery, such as for older patients, PCPD data are still required to look at aspects such as duration and adherence [20]. Our study demonstrates the benefit of PCPD data in highlighting the poor recording of ET in secondary care data and providing more accurate estimates of use across patient groups. Further use of the linked PCPD data would add to the research published on ET initiation and adherence in different patient groups, which have used various data sources including prescribing data and patient-reported adherence [20–23].

Where ET was recorded in both datasets we found that around one-third (38%) of patients had the earliest ET record in the same month in COSD and PCPD, whilst 3% had the earliest ET record in PCPD. This might reflect that ET was prescribed for a month, so that in the case of these 38% the two data sources are capturing the same prescription. Although some hospital prescriptions can only be collected from the hospital pharmacy, others can be collected from a community setting and so we would see the ET recorded in PCPD as well as being captured in COSD as part of the patient's cancer treatment. It is unclear why for 3% of patients with ET recorded in both COSD and PCPD the earliest record was in PCPD.

This study found ET was less likely to be recorded in COSD for younger women, women with no comorbidity/frailty, less advanced disease, HER2-positive disease, and patient groups with other treatments recorded (surgery, radiotherapy, chemotherapy). All of these groups implicitly or explicitly include patients who are more likely to have surgery or other treatments first. During the initial multidisciplinary team meeting, where treatment decisions are made, those patients who proceed immediately to surgery or neoadjuvant treatment might be less likely to have ET recorded as this is typically commenced after these treatments are completed. These findings are similar to those reported by a study carried out in the United States comparing the recording of ET within Surveillance, Epidemiology, and End Results (SEER) breast cancer data and that in pharmacy dispensing [24]. They reported higher recording of ET in pharmacy data along with differences by age, disease stage and use of chemotherapy.

The PCPD represents a valuable source of national coverage on information pertaining to other cancer drugs not well recorded in secondary care data, including ET in other cancers and bisphosphonates. Improved understanding of adherence to these drugs is important to evaluate treatment-related morbidity and outcomes. Use of PCPD, as demonstrated in this study, will support studies of treatment duration, compliance, adherence to guidelines, cost-effectiveness (whether cheaper or more expensive drugs are being routinely used) and allow a level of quality assurance between secondary care data and PCPD where data collection is duplicated. There are further potential applications beyond cancer drugs, with classes of drugs such as hormone replacement therapy (HRT) or supportive drugs (e.g. analgesics, anti-emetics, steroids) largely prescribed and dispensed within primary care. In these settings, PCPD could be used to improve our understanding of relationships between HRT drug use and breast cancer incidence, or to provide national estimates on the use of supportive drugs and allow an improved appreciation of comorbidity in patients with cancer [25–30].

This study has many strengths. Firstly, it provides robust estimates on the consistency of ET recording across COSD and PCPD by including all women aged 50 + years diagnosed with hormone receptor-positive invasive breast cancer in England from April-2015 to December-2019, which provides a large national patient cohort. Secondly, this study was able to explore the comparability of ET recorded across PCPD and COSD and highlight those factors associated with poor recording in COSD by using linked patient-level data from multiple datasets covering all aspects of patient care. Finally, with prescriptions data collected via electronic prescribing systems, estimates of ET prescribing from PCPD will most accurately reflect practice where ET is dispensed within primary care.

There are some limitations. Firstly, with data only available for women aged 50 + years it was not possible to look at ET recording across datasets for younger women. A previous publication including pre-menopausal women however demonstrated PCPD provided ET estimates in line with clinical expectation [11]. Secondly, as PCPD only captures prescriptions dispensed in the community and secondary care data is largely incomplete, with significant bias in recording, it is unclear at what point PCPD starts capturing ET use. Furthermore, type of ET for EIBC/LABC was classified based on dates of ET in either dataset and date of surgery, rather than clinical treatment intent (as this was not routinely available). Therefore definitions may vary, and where PCPD does not capture the first ET dose, the type of ET may be misclassified where ET was started in secondary care prior to surgery. However the purpose of these classifications was to compare where ET was picked up in each dataset, in relation to surgery where this was part of primary treatment. Finally, as PCPD does not include day, calculations of time are approximate. There are some further limitations relating to datasets. Drug indication is not captured in PCPD, as it is not recorded on prescriptions, therefore data from PCPD may include instances of ET prescribed for something other than breast cancer (e.g. for fertility treatment or other gynaecological diseases, although use in these circumstances is rare). Linkage to primary care information captured in the Clinical Practice Research Datalink or The Health Improvement Network may be useful in determining drug indication, however population coverage is typically low (13% and <10% respectively) [31–33]. Additionally, as non e-prescribed hospital prescriptions are not captured in PCPD, and won't be recorded in SACT, ET dispensed via this route is not reflected in this analysis. Numbers relating to this means of dispensing however are likely small so should have negligible impact.

The work presented in this publication was undertaken in order to inform the audit of breast cancer care in England (previously carried out as part of the NABCOP). The findings are likely to be important for others using routine data to look at ET in women with hormone receptor-positive invasive breast cancer, but will depend on the aims of the data analysis and the patient groups of interest. The combination of data from COSD and PCPD is important to understand the use of ET in routine care, particularly in those patient groups where ET capture is poorer than might be expected.

5. Conclusions

Secondary care data alone should not be used to evaluate ET for hormone receptor-positive invasive breast cancer as they provide an inaccurate representation of use, particularly among younger women in this cohort of women aged 50 + years, those with comorbidity/frailty, less advanced disease, HER2-positive disease or other treatments recorded. PCPD data are currently vital for identifying comprehensive information on use of ET within this population, and describing attributes such as duration and adherence. Completeness of secondary care data should be addressed to improve the available information on first ET prescription and to allow it to be a useful source to estimate use of ET. Our findings suggest PCPD (linked to secondary care data) is a valuable resource for examining patterns of care.

Ethical approval

This study was exempt from NHS Research Ethics Committee approval because it involved analysis of pseudonymised linked data

collated for the purpose of service evaluation as part of the National Audit of Breast Cancer in Older Patients.

Consent for publication

Not applicable.

Funding

This study was undertaken as part of the work by the National Audit of Breast Cancer in Older Patients (NABCOP). The NABCOP was commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit and Patient Outcomes Programme, and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). Neither the commissioner nor the funders had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. The authors had full independence from the HQIP. The aim of the NABCOP was to evaluate the care of older women with breast cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these women. More information can be found at www.nabcop.org.uk. DD also receives funding from Cancer Research UK (grant C8225/A21133). Cancer Research UK had no involvement in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. KC declares funding received for Breast Cancer Research Manager role within NHS England as part of the Cancer Grand Challenges PRECISION team, which is funded by Cancer Research UK and the Dutch Cancer Society (C7827/A27366); the grant is not related to her work on the National Audit of Breast Cancer in Older Patients or this paper.

CRediT authorship contribution statement

guarantor of integrity of the entire study: MRG: study concepts and design: MRG, DD, KM, KH, KC, JM, MHP, DAC: literature research: MRG: clinical studies: N/A: experimental studies / data analysis: N/A: data acquisition: MRG, KC, JM: statistical analysis: MRG: manuscript preparation: MRG: manuscript editing: MRG, DD, KM, KH, KC, JM, MHP, DAC. All authors were involved in data interpretation, critical appraisal of the draft manuscript, and gave final approval on the submitted version.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: KH declares being the Chair of the Endonet Trial Steering Committee.

Data Availability

This work uses data that has been provided by patients and collected by the NHS as part of their care and support. The data for England are collated, maintained and quality assured by the National Disease Registration Service (NDRS), which is part of NHS Digital. Data on English Cancer Registrations can be accessed via the NHS Digital Data Access Request Service (DARS) https://digital.nhs.uk/services/data-access-request-service-dars#national-disease-registration-service-ndrs-

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.canep.2023.102423.

References

[1] NICE. Early and locally advanced breast cancer: diagnosis and management.2018. Available from: www.nice.org.uk/guidance/ng101.
[2] NICE. Advanced breast cancer: diagnosis and treatment 2009. Available from: https://www.nice.org.uk/guidance/cg81.
[3] L. Biganzoli, N.M.L. Battisti, H. Wildiers, A. McCartney, G. Colloca, I.H. Kunkler, et al., Updated recommendations regarding the management of older patients with breast cancer: a joint paper from the European Society of Breast Cancer Specialists (EUSOMA) and the International Society of Geriatric Oncology (SIOG), Lancet Oncol. 22 (7) (2021) e327–e340.
[4] K.E. Henson, R. Brock, B. Shand, V.H. Coupland, L. Elliss-Brookes, G. Lyratzopoulos, et al., Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England, BMJ Open 8 (7) (2018), e020980.
[5] E. Bastiaannet, J. Charman, T.B. Johannesen, S. Schrodi, S. Siesling, L. van Eycken, et al., A European, observational study of endocrine therapy administration in patients with an initial diagnosis of hormone receptor-positive advanced breast cancer, Clin. Breast Cancer 18 (4) (2018) e613–e619.
[6] R.H. Jack, E.A. Davies, H. Moller, Breast cancer incidence, stage, treatment and survival in ethnic groups in South East England, Br. J. Cancer 100 (3) (2009) 545–550.
[7] D. Tataru, D. Robinson, H. Moller, E. Davies, Trends in the treatment of breast cancer in Southeast England following the introduction of national guidelines, J. Public Health 28 (3) (2006) 215–217.
[8] R. Sainsbury, L. Rider, A. Smith, MacAdam A. Does it matter where you live? Treatment variation for breast cancer in Yorkshire. The Yorkshire Breast Cancer Group, Br. J. Cancer 71 (6) (1995) 1275–1278.
[9] S.E. Ward, P.D. Richards, J.L. Morgan, G.R. Holmes, J.W. Broggio, K. Collins, et al., Omission of surgery in older women with early breast cancer has an adverse impact on breast cancer-specific survival, Br. J. Surg. 105 (11) (2018) 1454–1463.
[10] P. Monaghan, L. Murray, M. Donnelly, P. McCarron, R. Spence, A. Gavin, Breast cancer services–a population-based study of service reorganization, J. Public Health 27 (2) (2005) 171–175.
[11] G. Emanuel, K.E. Henson, J. Broggio, J. Charman, K. Horgan, D. Dodwell, et al., Endocrine therapy in the years following a diagnosis of breast cancer: a proof of concept study using the primary care prescription database linked to cancer registration data, Cancer Epidemiol. 61 (2019) 185–189.
[12] E.I. Benchimol, L. Smeeth, A. Guttmann, K. Harron, D. Moher, I. Petersen, et al., The reporting of studies conducted using observational routinely-collected health Data (RECORD) statement, PLoS Med 12 (10) (2015), e1001885.
[13] K.E. Henson, L. Elliss-Brookes, V.H. Coupland, E. Payne, S. Vernon, B. Rous, et al., Data resource profile: national cancer registration dataset in England, Int J. Epidemiol. 49 (1) (2020), 16-h.
[14] A. Herbert, L. Wijlaars, A. Zylbersztejn, D. Cromwell, P. Hardelid, Data resource profile: hospital episode statistics admitted patient care (HES APC), Int J. Epidemiol. 46 (4) (2017), 1093-i.
[15] C.J. Bright, S. Lawton, S. Benson, M. Bomb, D. Dodwell, K.E. Henson, et al., Data resource profile: the systemic anti-cancer therapy (SACT) dataset, Int J. Epidemiol. 49 (1) (2020), 15-l.
[16] National Cancer Registration and Analysis Service. Matching SACT to Cancer Waiting Times Data. 2014 [Available from: http://www.ncin.org.uk/publications/data_briefings/sact_cwt.
[17] J.N. Armitage, J.H. van der Meulen, Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score, Br. J. Surg. 97 (5) (2010) 772–781.
[18] Y. Jauhari, M.R. Gannon, D. Dodwell, K. Horgan, K. Clements, J. Medina, 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 10 (5) (2020), e035395.
[19] StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC2021.
[20] S. McGuinness, L. Hughes, R. Moss-Morris, M. Hunter, S. Norton, Z. Moon, Adherence to adjuvant endocrine therapy among White British and ethnic minority breast cancer survivors in the United Kingdom, Eur. J. Cancer Care (2022), e13722.
[21] B. Makubate, P.T. Donnan, J.A. Dewar, A.M. Thompson, C. McCowan, Cohort study of adherence to adjuvant endocrine therapy, breast cancer recurrence and mortality, Br. J. Cancer 108 (7) (2013) 1515–1524.
[22] B. Robinson, B. Dijkstra, V. Davey, S. Tomlinson, C. Frampton, Adherence to adjuvant endocrine therapy in christchurch women with early breast cancer, Clin. Oncol. R. Coll. Radio. 30 (1) (2018) e9–e15.
[23] E.J.A. Bowles, C. Ramin, D.S.M. Buist, H.S. Feigelson, S. Weinmann, L.H.S. Veiga, et al., Endocrine therapy initiation among women with stage I-III invasive, hormone receptor-positive breast cancer from 2001-2016, Breast Cancer Res. Treat. 193 (1) (2022), 203-16.
[24] C.B. Haas, E.J.A. Bowles, J.M. Lee, J. Specht, D.S.M. Buist, Accuracy of tumor registry versus pharmacy dispensings for breast cancer adjuvant endocrine therapy, Cancer Causes Control 33 (9) (2022), 1145-53.
[25] Collaborative Group on Hormonal Factors in Breast C, Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence, Lancet 394 (10204) (2019), 1159-68.
[26] N. Akter, E. Kulinskaya, N. Steel, I. Bakbergenuly, The effect of hormone replacement therapy on the survival of UK women: a retrospective cohort study 1984-2017, BJOG 129 (6) (2022) 994–1003.

[27] J. Busby, K. Mills, S.D. Zhang, F.G. Liberante, C.R. Cardwell, Selective serotonin reuptake inhibitor use and breast cancer survival: a population-based cohort study, Breast Cancer Res 20 (1) (2018) 4.
[28] B.M. Hicks, L.J. Murray, C. Hughes, C.R. Cardwell, Post-diagnostic oral bisphosphonate use and colorectal cancer mortality: a population-based cohort study within the UK Clinical Practice Research Datalink, Br. J. Cancer 113 (1) (2015) 123–126.
[29] Y. Vinogradova, C. Coupland, J. Hippisley-Cox, Exposure to bisphosphonates and risk of common non-gastrointestinal cancers: series of nested case-control studies using two primary-care databases, Br. J. Cancer 109 (3) (2013) 795–806.
[30] R.A. Wiseman, Breast cancer: critical data analysis concludes that estrogens are not the cause, however lifestyle changes can alter risk rapidly, J. Clin. Epidemiol. 57 (8) (2004) 766–772.
[31] A. Wolf, D. Dedman, J. Campbell, H. Booth, D. Lunn, J. Chapman, et al., Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum, Int J. Epidemiol. 48 (6) (2019), 1740-g.
[32] E. Herrett, A.M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, et al., Data resource profile: clinical practice research datalink (CPRD), Int J. Epidemiol. 44 (3) (2015) 827–836.
[33] B.T. Blak, M. Thompson, H. Dattani, A. Bourke, Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates, Inf. Prim. Care 19 (4) (2011) 251–255.

5. Understanding systemic oncological treatment in clinical practice.

5.1. Research Paper 3

Title

Evidence into practice: A national cohort study of NICE-recommended oncological drug therapy utilisation among women diagnosed with invasive breast cancer in England.

Journal details

Published in British Journal of Cancer. The online PDF version can be accessed at: https://pubmed.ncbi.nlm.nih.gov/37741900/

Supplementary material can be found in the Appendices.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	British Journal of Cancer		
When was the work published?	23 September 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	[redacted]
Date	

British Journal of Cancer www.nature.com/bjc

ARTICLE OPEN

Check for updates

Evidence into practice: a national cohort study of NICE-recommended oncological drug therapy utilisation among women diagnosed with invasive breast cancer in England

Melissa Ruth Gannon[1,2]✉, David Dodwell[3], Ajay Aggarwal[1,4], Min Hae Park[1,2], Katie Miller[1,2], Kieran Horgan[5], Karen Clements[6], Jibby Medina[1,2] and David Alan Cromwell[1,2]

BACKGROUND: Multiple drug treatments are approved for invasive breast cancer (IBC). We investigated uptake of NICE-recommended oncological drugs and variation by age, comorbidity burden and geographical region.
METHODS: Women (aged 50+ years) diagnosed with IBC from 2014 to 2019, were identified from England Cancer Registry data and drug utilisation from Systemic Anti-Cancer Therapy data. Interrupted time series analysis assessed national-level changes in drug use after publication of NICE recommendations. Regression models analysed variation in use.
RESULTS: This national cohort included 168,449 women. Use of drugs recommended for first-line treatment varied, from 26.6% for CDK 4/6 inhibitors to 63.8% for HER2-targeting therapies. Utilisation of drugs with a NICE recommendation published between 2014 and 2019, increased among patients diagnosed around the time of publication, except in the case of pertuzumab for metastatic breast cancer (MBC) which was previously accessible via the Cancer Drugs Fund (though use of pertuzumab for MBC increased from 34.1% to 75.0% across the study period). Use of trastuzumab and neoadjuvant/adjuvant pertuzumab varied by geographical region. Use was low for ribociclib (2.2%), abemaciclib (2.3%) and for drugs recommended beyond the first-line setting. For all drugs, use after NICE recommendation varied by age at diagnosis and increased as stage increased.
CONCLUSIONS: Use of NICE-recommended drugs for IBC in routine care is variable, with lowest use among women aged 70+ years. Improving access to effective treatments is an important step in improving outcomes.

British Journal of Cancer; https://doi.org/10.1038/s41416-023-02439-z

BACKGROUND

The National Institute for Health and Care Excellence (NICE), established in 1999, is responsible for providing evidence-based guidance to support commissioning decisions within the National Health Service (NHS) in England [1]. NICE conducts health technology assessments as part of a technology appraisal (TA) process. The TA process assesses the clinical and cost-effectiveness of drugs submitted for approval and determines which drugs should be funded for use in the routine care of patients, including those with cancer [2, 3]. Based on the findings of this TA process, NICE publishes guidance on the recommended use of the appraised treatment. Drugs recommended by NICE are routinely commissioned by the NHS and should form part of the treatment options for patients.

Over the past two decades, multiple treatments for invasive breast cancer (IBC) have been recommended by NICE. However, there is limited information on the translation of these recommendations into use of the drugs among eligible patients, particularly at a national level [4–9]. Improving the outcomes of cancer treatment requires the translation of national recommendations on optimal treatment into delivery of those drugs to patients. Understanding the utilisation of such treatments in the patient population they were intended for is a vital first step in understanding this process.

This study aimed to investigate the utilisation of NICE-recommended oncological drugs for IBC in routine care, among women aged 50 years and over diagnosed with IBC in England from 2014 to 2019. We also describe the extent to which utilisation varied by age at diagnosis, comorbidity burden and geographical region.

METHODS

Data sources and study population

This population-based, national cohort study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP).

Identification of NICE-recommended drugs. NICE Technology Appraisal Guidance (TAGs) published from January 2002 (the first TAG published for IBC) to 31 December 2019 were reviewed on 17 March 2022. We identified drugs recommended by NICE for use in routine care. Full details of the

[1]Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK. [2]Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK. [3]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [4]Department of Oncology, Guys Cancer Centre, Guy's & St Thomas' NHS Trust, London, UK. [5]Department of Breast Surgery, St James's University Hospital, Leeds, UK. [6]NHS England, 5th Floor, 23 Stephenson Street, Birmingham, UK.
✉email: melissa.gannon@lshtm.ac.uk

Received: 9 November 2022 Revised: 7 September 2023 Accepted: 12 September 2023
Published online: 23 September 2023

process of identifying NICE-recommended drugs can be found in Appendix 1.

Routine data. Data on all aspects of diagnosis and care for patients with cancer are routinely collected by the NHS as part of their care and support. This study used pseudonymised cancer registration patient records for all women aged 50 years and over diagnosed with breast cancer in England between January 2014 and December 2019. Data were linked at tumour-level to the Cancer Outcomes and Services Dataset (COSD), which provided information on patient and tumour characteristics, and Hospital Episode Statistics Admitted Patient Care (HES-APC) data for surgical and comorbidity details. Records were linked at patient-level to data in the Primary Care Prescription Database, for information on endocrine therapy use [10]. Linkage of tumour-level records to Systemic Anti-Cancer Therapy (SACT) data provided information on prescribed oncological drugs [11]. SACT data were available for drugs with an administration date from 1 January 2014 up to 28 February 2022. We used SACT data to flag use of oncological drugs recommended by NICE, based on a record of either the drug's brand or generic name in the drug name field. For analysis involving trastuzumab, Herceptin and other trastuzumab biosimilars (herzuma; ontruzant; trazimera) were included [12]; trastuzumab-emtansine was considered separately because it had its own NICE TAG.

Patient inclusion/exclusion. Women were included in the study if they had a registered diagnosis of IBC (ICD-10 code C50) and recorded stage 1–4. Women were excluded if they had a date of death within six months of diagnosis or a previous registration of breast cancer.

Definition of study variables

Data on the following patient demographics and tumour characteristics were taken from Cancer Registry and COSD: age at diagnosis (years), ethnicity (White, Mixed, Asian or Asian British, Black or Black British, Other Ethnic Group, Not Reported), level of deprivation, overall stage (1, 2, 3a, 3b, 3c, 4), tumour stage (T1, T2, T3, T4), nodal stage (N0, N1–3), invasive grade (G1, G2, G3), estrogen receptor (ER)/progesterone receptor (PR)/human epidermal growth receptor 2 (HER2) status (positive or negative). Deprivation was measured using the Index of Multiple Deprivation 2019 rank, derived from the patient's postcode at diagnosis, with ranks assigned to national quintiles of deprivation, from most (group 1) to least (group 5) deprived. This allocation was done by NCRAS and the calculated deprivation quintile provided within Cancer Registry data.

Stage groups were defined as early invasive breast cancer (EIBC; stage 1–3a), locally advanced breast cancer (LABC; stage 3b–c) and metastatic breast cancer (MBC; stage 4).

IBC was defined as hormone receptor-positive if either ER or PR status were recorded as positive.

Comorbidity burden (0, 1, 2+) was calculated according to the Royal College of Surgeons of England Charlson Comorbidity Index [13]. This counts the presence of specific chronic medical conditions (excluding malignancy) which are identified using ICD-10 diagnosis codes recorded in HES-APC data for episodes in the two years prior to diagnosis.

To identify the use of drugs recommended by NICE as part of first-line treatment, we counted any administration of the drug recorded in SACT either within 12 months of diagnosis or within the first treatment episode (defined by consecutive treatments with no more than an 8 month break between them). For drugs recommended by NICE for use in a surgical setting, use was defined as (i) neoadjuvant where the earliest recorded administration date was within six months of diagnosis and prior to surgery or (ii) adjuvant where the earliest recorded administration date was within six months after the date of surgery. For drugs recommended by NICE for relapse/recurrence, progression or after previous treatment, any drug administration after diagnosis was counted.

Statistical analysis

All data preparation and statistical analyses were conducted using Stata version 17.0.

Descriptive statistics were used to summarise the percentage (number) of eligible women initiating each NICE-recommended drug. Eligibility for each drug was defined based on stage group and HER2 and/or hormone receptor status as applicable. Use of each NICE-recommended drug was described overall and within patient subgroups defined by stage group, age, year of diagnosis and time from diagnosis to first record.

Multilevel mixed-effects (MLME) logistic regression models were used to analyse differences in drug utilisation across patient subgroups defined by age, comorbidity burden and geographical region, for drugs recommended as part of first-line treatment. MLME models were fitted among eligible women diagnosed after publication of the NICE TAG (referred to as "post-publication" in the results). Models were adjusted for factors associated with treatment decision-making including stage group (as applicable), tumour stage, nodal stage, HER2/hormone receptor status (as applicable) and grade. Deprivation and ethnicity were also included in the models. Missing values were included as unknown categories. MLME models accounted for clustering of patients within Government Office Regions (GORs). Each GOR was fitted as a random intercept, representing differences between GORs not explained by the factors in the model. NHS trusts were aggregated into GOR, due to relatively low levels of activity at NHS trust level, to understand variation by geographical region. MLME models were only fitted where rates of use among eligible women diagnosed after NICE TAG publication were at least 5%, to allow for robust estimates.

For drugs with a NICE TAG published during the period when patients in the study were diagnosed (January 2014 to December 2019), we used interrupted time series analysis (ITSA) to examine the impact of NICE guidance on national-level uptake. ITSA was only used for drugs recommended for use as part of first-line treatment and where rates of use among eligible women diagnosed after NICE TAG publication were at least 5%. The ITSA model allowed for rate of drug initiation to change smoothly over time, as well as abruptly following NICE TAG publication, and for changes due to seasonality [9, 14]. The ITSA model for each drug defined an "intervention" time point that began with the first full month following NICE TAG publication. The model also incorporated a short "transition" period (defined as 6 months prior to NICE TAG publication for neoadjuvant drugs and 12 months otherwise) to account for the fact that the time-series was defined based on month of diagnosis and treatment could occur months later. We used the 'itsa' command in Stata to test for statistical evidence of changes in the monthly trends in the number of patients starting treatment from one time period to the next (i.e., pre-publication, transition, post-publication) [15]. Seasonality was adjusted for by including the calendar month as an independent variable in the models. Monthly rates of drug initiations were calculated as the number of eligible women with a record of the drug starting in SACT divided by the number of eligible women diagnosed in the month, multiplied by 1000 (to give rates per 1000 women).

RESULTS

Patient cohort

Among 209,968 women aged 50 years and over diagnosed with breast cancer in England from 1 January 2014 to 31 December 2019, 80.2% ($N = 168{,}449$) had invasive disease with a recorded stage, and did not die within six months of diagnosis (Fig. A1). Of this cohort, women were predominantly diagnosed with EIBC (91.8%), whilst 4.3% had LABC and 3.9% had MBC. Nearly two-thirds were aged 50–69 years at diagnosis (61.3%), whilst 22.9% were 70–79 years and 15.9% were 80+ years.

We identified 13 TAGs which recommended oncological drugs for use in IBC (Table 1). Of these, one TAG had been updated and replaced by NICE guideline NG101 [16, 17]. An additional six published TAGs of oncological drug regimens including bevacizumab, fulvestrant, and lapatinib, were identified where NICE did not recommend treatment and so were excluded from this study (Appendix 1) [18–23].

Utilisation of drugs recommended by NICE for first-line treatment

Table 2A describes the percentage of eligible women with a record of starting each of the drugs recommended by NICE for use as part of first-line treatment. Rates of use for HER2-targeting therapies were found to be the highest, regardless of age, with 63.8% of women (all stages) with HER2-positive IBC having a record of receiving trastuzumab or pertuzumab as part of first-line treatment (63.3–64.1% EIBC/LABC; 71.8% MBC). Among women newly-diagnosed with HER2-negative, hormone receptor-positive LABC or MBC from 2018 to 2019, 26.6% received one of the three recommended CDK4/6 inhibitors (7.8%

Table 1. Overview of NICE technology appraisal guidance with a positive recommendation published for invasive breast cancer.

Drug (Brand name, Company)	Type	Summary of the indication/eligibility	NICE TAG Publication date (TA#)	Licensing dates [4]		Notes
				FDA	EMA	
Trastuzumab (Herceptin, Roche)	HER2-targeting therapy	HER2-positive metastatic breast cancer	March 2002 (TA34) [43]	1998	2000	-
		HER2-positive early breast cancer	August 2006 (TA107) [16]			Replaced by NG101 [17]
Gemcitabine (Gemzar, Eli Lilly and Company Ltd)	Chemotherapy drug	Metastatic breast cancer which has relapsed following adjuvant/neoadjuvant chemotherapy (including an anthracycline). In combination with paclitaxel.	January 2007 (TA116) [44]	2004	Pre-EMA	-
Pertuzumab (Perjeta, Roche)	HER2-targeting therapy	HER2-positive locally advanced, inflammatory or early breast cancer at high risk of recurrence. Neoadjuvant use, in combination with trastuzumab & docetaxel.	December 2016 (TA424) [45]	2012	2013	Provided with the discount agreed in the PAS.
		HER2-positive metastatic or locally recurrent unresectable breast cancer, not previously treated with anti-HER2 therapy or chemotherapy for metastatic disease. In combination with trastuzumab & docetaxel.	March 2018 (TA509) [46]			Previously on the CDF [47]. Provided within the agreed commercial access arrangement
		HER2-positive (locally advanced, inflammatory or) early breast cancer at high risk of recurrence (lymph node positive). In combination with trastuzumab & chemotherapy.	March 2019 (TA569) [48]			Provided according to the commercial agreement.
Everolimus (Afinitor, Novartis Pharmaceuticals)	mTOR kinase inhibitor	HER2-negative, hormone receptor-positive advanced breast cancer that has recurred or progressed after a non-steroidal aromatase inhibitor. In combination with exemestane	December 2016 (TA421) [49]	2012	2012	Provided with the discount agreed in the PAS.
Eribulin (Halaven, Eisai and Merck)	Chemotherapy drug	Locally advanced or metastatic breast cancer which has had 2 or more rounds of chemotherapy.	December 2016 (TA423) [50]	2010	2011	Provided with the discount agreed in the PAS.
Trastuzumab-emtansine (Kadcyla; Roche)	HER2-targeting therapy	HER2-positive, unresectable locally advanced or metastatic breast cancer previously treated with trastuzumab and a taxane, separately or in combination.	July 2017 (TA458) [51]	2013	2013	Provided with the discount agreed in the PAS.
Palbociclib (Ibrance, Pfizer)	CDK inhibitor	Hormone receptor-positive, HER2-negative locally advanced or metastatic breast cancer as initial endocrine-based therapy with an aromatase inhibitor.	December 2017 (TA495) [52]	2015	2016	Provided with the discount agreed in the PAS.
Ribociclib (Kisqali, Novartis)	CDK inhibitor	Hormone receptor-positive, HER2-negative locally advanced or metastatic breast cancer as initial endocrine-based therapy with an aromatase inhibitor.	December 2017 (TA496) [53]	2017	2017	Provided with the discount agreed in the PAS.
Abemaciclib (Verzenios, Eli Lilly),	CDK inhibitor	Hormone receptor-positive, HER2-negative locally advanced or metastatic breast cancer as initial endocrine-based therapy with an aromatase inhibitor.	February 2019 (TA563) [54]	2017	2018	Provided according to the commercial agreement.
Neratinib (Nerlynx, Pierre Fabre)	HER2-targeting therapy	Extending adjuvant treatment of hormone receptor-positive, HER2-positive early breast cancer where trastuzumab was completed within the previous year.	November 2019 (TA612) [55]	2017	2018	Provided according to the commercial agreement.

TAG technology appraisal guidance, *FDA* food and drug administration, *EMA* european medicines agency, *HER2* human epidermal growth receptor 2, *PAS* patient access scheme, *CDF* cancer drugs fund, *mTOR* mechanistic target of rapamycin, *CDK* cyclin dependent kinase, *MAA* managed access agreement.

Table 2. Frequency of drug use recorded within SACT, for oncological drugs for breast cancer recommended by NICE

A. Oncological drugs for breast cancer recommended by NICE for use as first-line treatment following initial diagnosis														
Drug (NICE TAG date)	**Indication/ eligibility**		**Use among women diagnosed 2014–2019**						**Use among women diagnosed post-publication**					**Note**
			Overall use % (n/N)	**% by year of diagnosis *(column %)***					**% overall and by age at diagnosis *(column %)***					
				2014/ 15	**2016**	**2017**	**2018**	**2019**	**Overall**	**50–59 years**	**60–69 years**	**70–79 years**	**80+ years**	**% starting < 12 m of diagnosis *(row %)***	
Trastuzumab (Mar-2002)	HER2-positive MBC		71.8% ($n = 669/932$)	60.1%	70.7%	77.3%	80.8%	82.3%	71.8%	80.4%	79.8%	67.2%	44.3%	98.8%	
Trastuzumab (Aug-2006)	HER2-positive	EIBC	63.2% ($n = 9355/14808$)	55.7%	63.5%	67.5%	67.5%	69.9%	63.2%	78.0%	70.7%	52.3%	9.1%	98.9%	
		LABC	64.1% ($n = 742/1157$)	62.6%	65.9%	63.8%	62.7%	68.1%	64.1%	83.8%	77.7%	63.3%	23.5%	98.0%	
Pertuzumab (Dec-2016)	HER2-positive	EIBC	15.0% ($n = 2049/13686$)	0.4%	5.8%	25.8%	28.7%	30.9%	28.4%	38.4%	30.1%	16.8%	1.6%	100.0%	Neoadjuvant use, in combination with trastuzumab & docetaxel. (99.8% recorded as starting chemo & trastuzumab).
		LABC	22.6% ($n = 214/947$)	4.1%	13.4%	41.4%	45.7%	38.8%	42.1%	60.3%	59.6%	28.2%	6.4%	100.0%	
Pertuzumab (Mar-2018)	HER2-positive MBC		52.0% ($n = 485/932$)	34.1%	49.7%	60.2%	61.7%	75.0%	68.4%	87.0%	77.4%	56.5%	24.2%	99.4%	Use for breast cancer not previously treated with anti-HER2 therapy or chemo, in combination with trastuzumab & docetaxel.
Pertuzumab (Mar-2019)	HER2-positive	EIBC	2.1% ($n = 221/10456$)	0.9%	0.6%	0.6%	1.4%	10.3%	11.4%	14.9%	10.7%	11.0%	1.8%	100.0%	Adjuvant use, in combination with trastuzumab & chemo. (93.9% had node-positive tumours).
		LABC	7.0% ($n = 39/560$)	3.8%	2.6%	3.9%	7.6%	28.4%	28.3%	25.0%	80.0%	15.4%	6.7%	100.0%	
Palbociclib (Dec-2017)	HER2-negative, hormone receptor-positive	LABC	2.9% ($n = 94/3261$)	0.0%	0.9%	4.3%	4.0%	8.8%	6.2%	7.0%	4.8%	9.2%	3.5%	89.6%	Use as initial endocrine-based therapy with an AI.
		MBC	17.5%; ($n = 497/2843$)	0.3%	2.2%	18.9%	42.6%	45.0%	43.8%	50.7%	56.9%	48.0%	15.7%	98.2%	
Ribociclib (Dec-2017)	HER2-negative, hormone receptor-positive	LABC	0.3% ($n = 10/3261$)	0.0%	0.2%	0.2%	1.2%	0.2%	0.7%	1.8%	0.0%	1.0%	0.0%	50.0%	Use as initial endocrine-based therapy with an AI.
		MBC	1.7% ($n = 48/2843$)	0.1%	0.0%	2.2%	3.1%	5.0%	4.1%	5.8%	5.2%	2.5%	3.1%	97.2%	

Table 2. continued

A. Oncological drugs for breast cancer recommended by NICE for use as first-line treatment following initial diagnosis

Drug (NICE TAG date)	Indication/ eligibility		Use among women diagnosed 2014–2019						Use among women diagnosed post-publication						Note
			Overall use % (n/N)	% by year of diagnosis *(column %)*					% overall and by age at diagnosis *(column %)*					% starting < 12 m of diagnosis *(row %)*	
				2014/15	2016	2017	2018	2019	Overall	50–59 years	60–69 years	70–79 years	80+ years		
Abemaciclib (Feb-2019)	HER2-negative, hormone receptor-positive	LABC	0.6% ($n = 19/3261$)	0.0%	0.4%	0.6%	1.4%	1.2%	0.9%	0.0%	2.1%	0.0%	2.0%	75.0%	Use as initial endocrine-based therapy with an AI.
		MBC	1.2% ($n = 35/2843$)	0.0%	1.2%	1.2%	1.6%	3.6%	3.8%	3.7%	3.4%	5.0%	2.6%	100.0%	

B. Oncological drugs for breast cancer recommended by NICE for use following relapse/recurrence, progression or previous treatment

Drug (NICE TAG date)	Indication/ eligibility		Use among women diagnosed 2014–2019						Use among women diagnosed post-publication						Note
			Overall use % (n/N)	% by year of first cycle, among women starting treatment *(row %)*					% by age at diagnosis *(column %)*					% starting 12 + m of diagnosis (row %)	
				2014/15	2016	2017	2018	2019 +	Overall	50–59 years	60–69 years	70–79 years	80+ years		
Gemcitabine (Jan-2007)	MBC		1.2% ($n = 70/5739$)	12.9%	15.7%	17.1%	21.4%	32.9%	1.2%	2.2%	1.5%	1.0%	0.0%	51.4%	Use for relapse following adjuvant/ neoadjuvant chemo.
Everolimus (Dec-2016)	HER2-negative, hormone receptor-positive MBC		5.2% ($n = 134/2576$)	4.5%	12.7%	17.9%	20.9%	44.0%	4.4%	6.6%	6.1%	3.4%	1.2%	85.5%	Use for recurrence or progression after a non-steroidal AI.
Eribulin (Dec-2016)	LABC		1.3% ($n = 89/6632$)	5.6%	12.4%	11.2%	15.7%	55.1%	1.0%	2.2%	1.1%	0.8%	0.0%	97.0%	Use for breast cancer which has had 2 or more rounds of chemo.
	MBC		3.3% ($n = 189/5739$)	3.7%	7.4%	12.2%	19.6%	57.1%	3.1%	6.0%	3.1%	2.9%	0.0%	95.5%	
Trastuzumab-emtansine (Jul-2017)	HER2-positive	LABC	8.4% ($n = 92/1089$)	5.4%	10.9%	20.7%	21.7%	41.3%	7.0%	10.8%	10.6%	3.5%	2.4%	76.0%	Use for unresectable breast cancer previously treated with trastuzumab + a taxane
		MBC	22.6% ($n = 191/846$)	6.8%	10.5%	14.7%	20.4%	47.6%	18.9%	23.7%	16.3%	18.6%	10.5%	77.4%	
Neratinib (Nov-2019)	HER2-positive, hormone receptor-positive EIBC		1.1% ($n = 111/9960$)	0.0%	0.0%	0.0%	1.8%	98.2%	3.5%	5.4%	3.2%	0.0%	0.0%	100.0%	Use where trastuzumab was completed within the previous year

TAG technology appraisal guidance, *HER2* human epidermal growth receptor 2, *MBC* metastatic breast cancer, *EIBC* early invasive breast cancer, *LABC* locally advanced breast cancer, *chemo* chemotherapy, *AI* aromatase inhibitor.

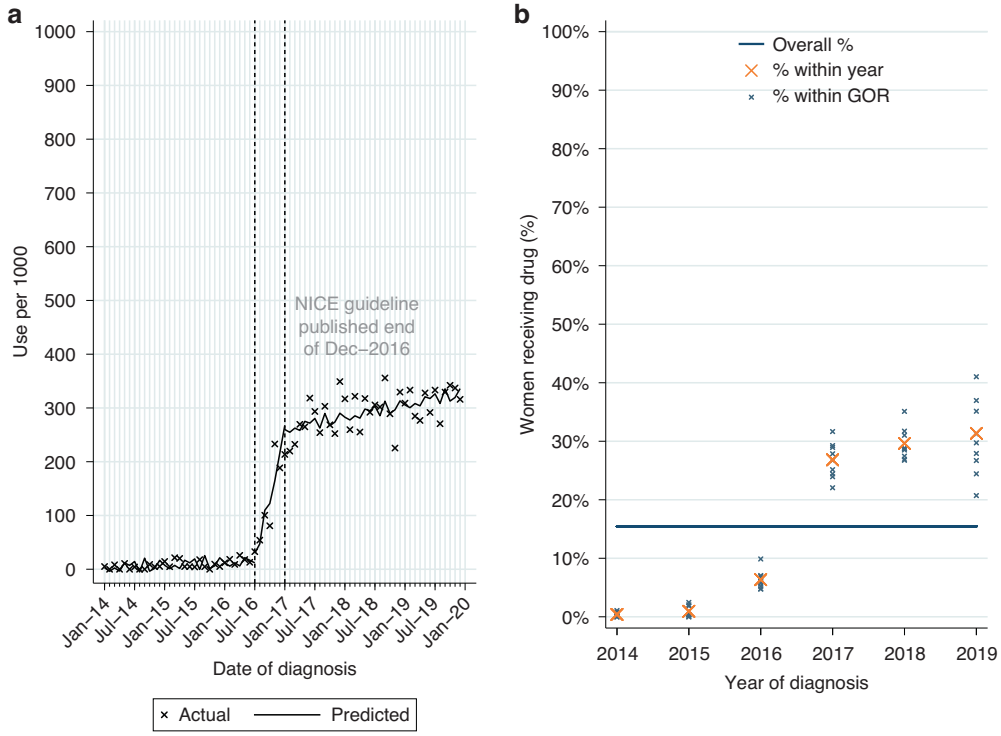


Fig. 1 Neoadjuvant pertuzumab utilisation. Neoadjuvant pertuzumab utilisation among women diagnosed with HER2-positive early invasive or locally advanced breast cancer from 2014 to 2019, and receiving surgery within 12 m of diagnosis. **a** Interrupted time series analysis of monthly neoadjuvant pertuzumab initiations. **b** Observed percentage of women initiating neoadjuvant pertuzumab, by Government Office Region (GOR) and year of diagnosis.

LABC; 49.4% MBC). Palbociclib was most frequently used, accounting for 87.2% of CDK4/6 inhibitor-based first-line treatment.

For all drugs, post-publication use varied. Use was highest among women aged 50–69 years (Table 2A) and increased with increasing stage group (if use was recommended across stage groups). The findings for each drug are presented in the following sections.

Trastuzumab. Trastuzumab, a HER2-targeting therapy, was recommended by NICE for HER2-positive MBC in 2002, and HER2-positive EIBC or LABC in 2006 (included in the 2009 NICE guideline); both recommendations were published prior to 2014, the first year for which the study had data, and so ITSA was not possible. Among eligible women 63.7% ($n = 10{,}766/16{,}897$) had a record of trastuzumab, with higher use for MBC (71.8%). Use increased over time, most prominently for EIBC and MBC (Fig. A2a). There was an increase in recorded use of trastuzumab biosimilars among women diagnosed from July 2018 onwards (Fig. A2b), forming 18.3% of trastuzumab-based treatment among eligible women diagnosed in 2019 (15.9% EIBC; 33.3% LABC; 39.2% MBC). Age and comorbidity burden were independently associated with use, and there was evidence of variation across GORs (observed range: 58.4–69.6%) (Table A1and A2), from a MLME model. Use decreased with increasing age across all GORs (Fig. A2c). Use had not increased over time among women aged 80+ years (Fig. A2d).

Pertuzumab. Pertuzumab is a HER2-targeting therapy recommended by NICE for neoadjuvant use for HER2-positive EIBC or LABC in 2016, for HER2-positive MBC in 2018, and for adjuvant use for HER2-positive EIBC or LABC in 2019; all recommendations were published during the study period and ITSA was possible. The following sub-sections describe use associated with each NICE TAG recommendation.

Neoadjuvant use for newly-diagnosed HER2-positive EIBC or LABC: Among women who had surgery for HER2-positive EIBC or LABC within 12 months of diagnosis, 15.5% ($n = 2263/14{,}633$) had a record of neoadjuvant pertuzumab. Use was higher for LABC (22.6%; Table 2A). From ITSA we found an increase in monthly initiations of neoadjuvant pertuzumab of 37.2 patients per 1000 women diagnosed from July to December 2016, with continued increase in monthly initiations of 1.9 patients per 1,000 among women diagnosed post-publication (Table A3, $p < 0.0001$; Fig. 1a). Use increased over time across all GORs (Fig. 1b).

Among eligible women diagnosed post-publication 29.2% had a record of neoadjuvant pertuzumab, with higher use for LABC (42.1%; Table 2A). Fitting a MLME model, age, comorbidity burden and GOR (observed range: 24.6–35.1%; Table A2) were all found to be independently associated with variation in use (Table A1). Use decreased as age increased across all GORs (Fig. A3a).

Use for newly-diagnosed HER2-positive MBC: Among 932 women with HER2-positive MBC at diagnosis, 52.0% had a record of pertuzumab (Table 2A). Use increased over time from 34.1% among women diagnosed in 2014–5 to 75.0% in 2019. Of women starting pertuzumab 93.0% also had a record of treatment with trastuzumab and docetaxel, in line with NICE guidance.

From ITSA, there was no evidence of a monthly increase in initiations among women diagnosed post-publication (Table A3, $p = 0.33$; Fig. 2a). Use increased over time across all GORs (Fig. 2b).

Among eligible women diagnosed post-publication 68.4% ($n = 175/256$) started pertuzumab. Due to the small number of women, it was not feasible to look further at variation in use. As pertuzumab was accessible via the CDF prior to NICE TAG publication a MLME model was fitted in patients diagnosed 2014–2019. Age and comorbidity were found to be independently associated with differences in use. There was also evidence of increasing use with increasing year of diagnosis, but no evidence of variation by GOR (Table A1).

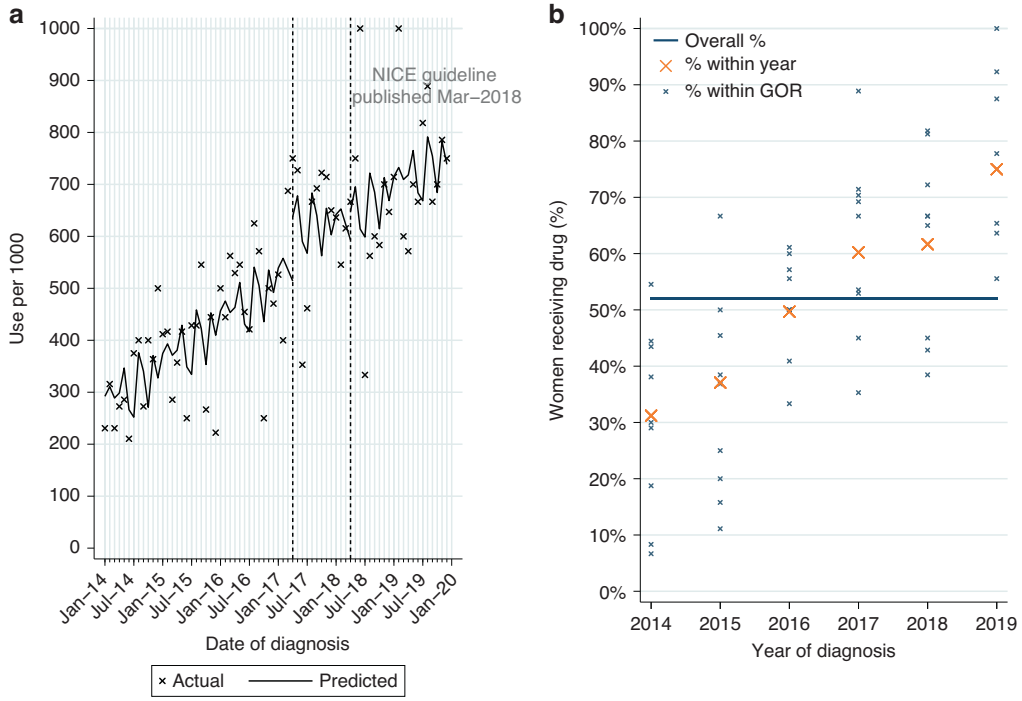


Fig. 2 Pertuzumab utilisation for metastatic breast cancer. Pertuzumab utilisation among women initially diagnosed with HER2-positive metastatic breast cancer from 2014 to 2019. **a** Interrupted time series analysis of monthly pertuzumab initiations. **b** Observed percentage of women initiating pertuzumab, by Government Office Region (GOR) and year of diagnosis.

Adjuvant use for newly-diagnosed HER2-positive EIBC or LABC: Among 11,016 women who had surgery within six months of diagnosis for HER2-positive EIBC or LABC, with no prior treatment with pertuzumab or trastuzumab, 2.4% had a record of adjuvant pertuzumab. Use was higher for LABC (7.0%; Table 2A). 93.5% also had a record of adjuvant chemotherapy and trastuzumab.

From ITSA we found monthly initiations of adjuvant pertuzumab increased by 8.1 patients per 1000, among eligible women diagnosed between April 2018 and March 2019 (Table A3, $p < 0.0001$; Fig. 3a), with an immediate increase in use among women diagnosed post-publication ($p = 0.017$). This increase was observed across all GORs (Fig. 3b). There was no evidence of continued increase in use among women diagnosed from April 2019 onwards ($p = 0.230$).

Among eligible women diagnosed from April 2019 onwards 12.1% started adjuvant pertuzumab, with higher use for LABC (28.3%; Table 2A). Fitting a MLME model, age and GOR (observed range: 3.0–22.0%; Table A2) were both independently associated with variation in use (Table A1). Low use among women aged 80+ was observed across all GORs (Fig. A3b).

Ribociclib or palbociclib. Ribociclib and palbociclib are CDK4/6 inhibitors recommended by NICE in 2017 for HER2-negative, hormone receptor-positive LABC or MBC. Among eligible women diagnosed post-publication 2.2% ($n = 44/1962$) had a record of ribociclib; 81.8% of use was for MBC. Numbers were insufficient to further investigate uptake of ribociclib.

Among women newly-diagnosed with HER2-negative, hormone receptor-positive LABC or MBC from 2014 to 2019, 9.7% ($n = 591/6104$) had a record of palbociclib, increasing to 23.2% ($n = 455/1962$) among women diagnosed post-publication. Use was highest for MBC (Table 2A). Of those starting palbociclib 93.6% ($n = 88/94$) of women with LABC and 89.9% ($n = 447/497$) of women with MBC also had a record of ever being prescribed an aromatase inhibitor. Among women with MBC, over three-quarters started pablociclib within 4 months of diagnosis.

From ITSA we found use of palbociclib increased among eligible women diagnosed during the 12 months prior to NICE TAG publication in December 2017 (monthly increase of 16.7 per 1000 patients, Table A3, $p < 0.0001$; Fig. 4a). Use continued to increase among women diagnosed from January 2018 onwards (monthly increase of 3.5 per 1000 patients, $p = 0.007$).

Among eligible women diagnosed post-publication, use varied by stage at diagnosis (6.2% LABC; 43.8% MBC; Table 2A). Fitting a MLME model, age was found to be independently associated with variation in use, but not comorbidity or GOR (Table A1). Use had increased over time across GORs (Fig. 4b) and for all age groups (Fig. A4a) but was lower among women aged 80+ years in all GORs (Fig. A4b).

Abemaciclib. Abemaciclib was recommended for use in 2019, 14 months after publication of NICE TAGs for palbociclib and ribociclib. Use among eligible women diagnosed after February 2019 was 2.3% ($n = 18/795$), with highest use for MBC. Numbers were insufficient to investigate uptake of abemaciclib further.

Uptake of drugs recommend by NICE for use following relapse/recurrence, progression or previous treatment

Table 2B describes the percentage of eligible women, among those who did not die within 12 months of diagnosis, with a record of starting each of the drugs recommended by NICE for use beyond a first-line setting or following initial treatment. For all drugs, post-publication use was highest among women aged 50–69 years and those with MBC (where use was recommended across stage groups). The findings for each drug are presented in the following sections.

Gemcitabine. Gemcitabine was recommended by NICE in January 2007 for MBC. 1.2% ($n = 70/5739$) of eligible women had a record of ever starting gemcitabine. There was no use among women aged 80+ years.

Everolimus or Eribulin. Everolimus (an mTOR kinase inhibitor for HER2-negative, hormone receptor-positive MBC) and eribulin (a

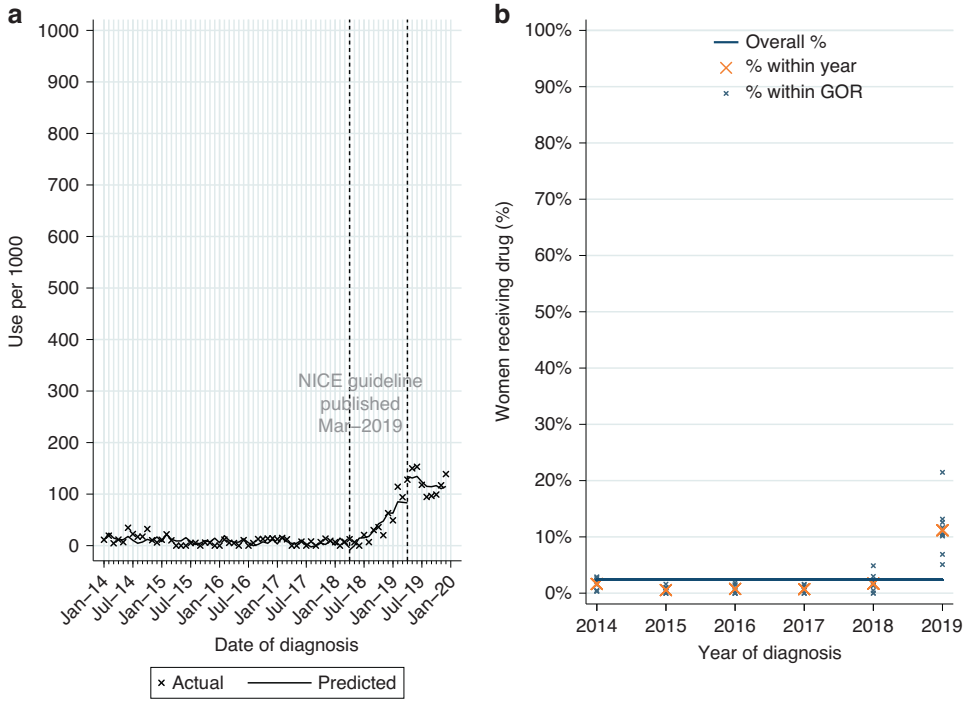


Fig. 3 Adjuvant pertuzumab utilisation. Adjuvant pertuzumab* utilisation among women diagnosed with HER2-positive early invasive or locally advanced breast cancer from 2014 to 2019, and receiving surgery within 6 m of diagnosis. **a** Interrupted time series analysis of monthly adjuvant pertuzumab* initiations within 6 m of surgery. **b** Observed percentage of women initiating adjuvant pertuzumab*, by Government Office Region (GOR) and year of diagnosis. *with no prior trastuzumab or pertuzumab received before surgery.

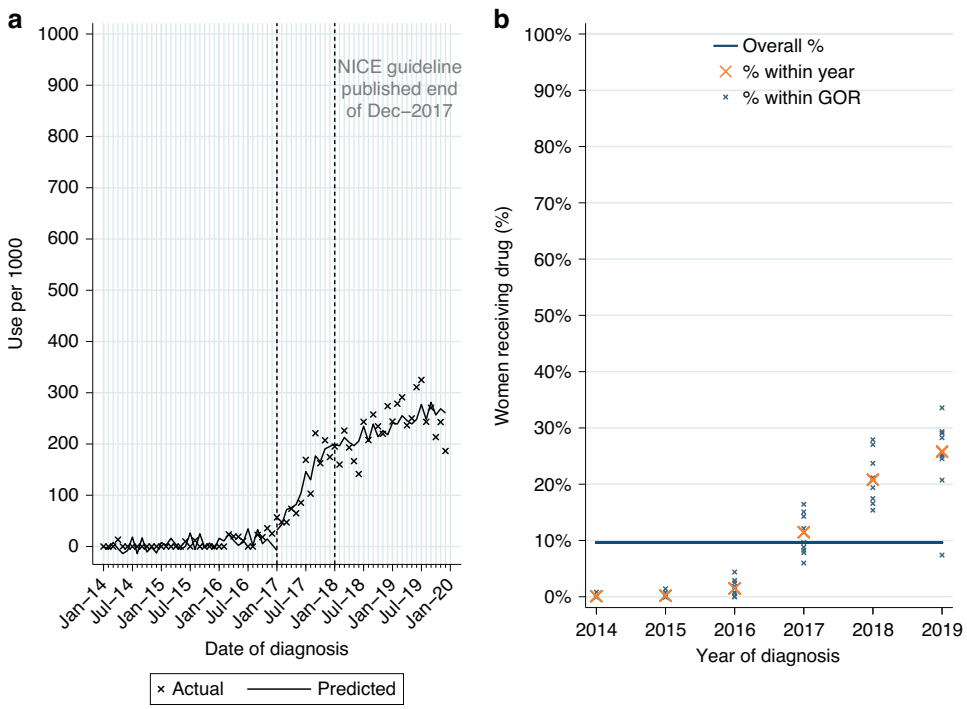


Fig. 4 Palbociclib utilisation. Palbociclib utilisation among women initially diagnosed with HER2-negative, hormone receptor-positive metastatic breast cancer from 2014 to 2019. **a** Interrupted time series analysis of monthly palbociclib initiations. **b** Observed percentage of women initiating palbociclib, by Government Office Region (GOR) and year of diagnosis.

chemotherapy for LABC or MBC) were both recommended by NICE in December 2016. Among eligible women diagnosed from 2017 to 2019, 4.4% ($n = 55/1250$) had a record of everolimus, and 2.0% ($n = 122/6117$) had a record of eribulin.

Trastuzumab-emtansine. Trastuzumab-emtansine, a HER2-targeting drug, was recommended for LABC or MBC in July 2017 for use in patients previously treated with trastuzumab and a taxane-based regimen. Among eligible women diagnosed August

2017–December 2019, 12.7% ($n = 87/684$) had a record of trastuzumab-emtansine. Use was highest for MBC (18.9%) compared with LABC (7.0%). Rates of use for MBC were highest in northern geographical regions (observed range across all regions: 5.6–33.3%).

Neratinib. Neratinib is another HER2-targeting therapy, recommended for HER2-positive, hormone receptor-positive EIBC by NICE in November 2019, for use following previous treatment. There were 114 eligible women diagnosed in December 2019, of whom 3.5% ($n = 4$) had a record of neratinib. There was no use among women aged 70+ years.

DISCUSSION

This population-based study used routinely-collected clinical data to evaluate the use of NICE-recommended drugs. Data were available for more than 160,000 women aged 50+ years diagnosed with IBC in England from 2014 to 2019.

Thirteen NICE TAGs were published between March 2002 and November 2019 where NICE made a positive recommendation. Several drug types, including HER2-targeting therapies, CDK 4/6 inhibitors and chemotherapies, were recommended for use either as part of first-line treatment or following recurrence/progression/previous treatment. Where use was recommended across stage groups, recorded use increased as stage group increased, with highest rates for MBC (where treatment was indicated). Use varied by age at diagnosis. Where numbers allowed for further analysis (trastuzumab, pertuzumab and palbociclib) there was evidence that differences by age were independent of other factors including comorbidity burden and geographical region. There are likely multiple reasons for this, including a lack of robust evidence for the efficacy and tolerability of treatments among older patients, who are under-represented in clinical trials. This might have led to a reluctance among oncologists to use new therapies for older patients. Other publications have reported reduced use of treatment in older patients irrespective of comorbidity [24, 25].

Where NICE-recommended drugs were intended for first-line treatment, utilisation among eligible women diagnosed post-publication varied, with 63.8% recorded as having HER2-targeting therapies (trastuzumab/pertuzumab), compared to 26.6% for CDK 4/6 inhibitors (ribociclib; palbociclib; abemaciclib). Highest rates of recorded drug use were for trastuzumab, a drug first introduced into clinical practice two decades ago and added to the World Health Organization Model List of Essential Medicines in 2015 [26]. A change in use to trastuzumab biosimilars for some patients may also have contributed to high use among women diagnosed in 2018 and 2019, demonstrating the continuing influence of drug development and approvals on uptake of existing approved drugs. This finding is echoed in an Italian study which reported increased use, with trastuzumab biosimilars contributing to around one-third of trastuzumab-based treatment among patients diagnosed in more recent years [27]. Additionally a review of biosimilars highlighted the value of their inclusion in trastuzumab-based treatment in increasing access to anti-HER2 therapies, particularly in relation to cost-saving [28].

Of three recommended CDK4/6 inhibitors, palbociclib was predominantly used, accounting for 87.2% of CDK4/6 inhibitor-based first-line treatment. This may be explained in part by palbociclib's existing approvals by the American Food and Drug Administration (FDA) (2015) and European Medicines Agency (EMA) (2016), whereas FDA/EMA approvals for ribociclib were in the same year as the NICE approval. Use of neoadjuvant/adjuvant pertuzumab and palbociclib, drugs recommended for first-line use by NICE between 2014 and 2019, increased among women diagnosed in the months around NICE TAG publication. The increase among women diagnosed 6–12 months prior to publication will be in part due to the timing of treatment in relation to diagnosis but may also be due to these drugs being previously approved for use by the EMA. Typically drugs with a new therapeutic indication have been approved first by the FDA in the US, followed by the EMA for use in Europe [29, 30]. Although the EMA provides market authorisation for all drugs to be used across Europe, within the UK NHS setting it is only following publication of the NICE TAG that they are usually recommended for use in routine practice [1]. There was no evidence that there was a national-level change in use following NICE recommendation of pertuzumab for MBC, however this drug was already available to patients through the NHS Cancer Drugs Fund (CDF). The CDF is another means through which oncological drugs with insufficient evidence of benefit at the point of appraisal are made available to patients [31–33]. Access via the CDF and a subsequent NICE recommendation meant there was an increase in use over the study period.

Where drugs were recommended for use beyond a first-line setting or following previous treatment, rates of recorded use were generally low. Patterns of use among women diagnosed post-publication were similar to those observed for first-line treatment, with use decreasing as age increased and highest for MBC.

There have been few previous studies looking at the translation of nationally recommended oncological drugs for IBC into routine care, with studies focusing on the safety and effectiveness of new drugs [34–40]. A study in the US identified a marked increase in the use of oral cancer drugs with no documented overall survival benefit between 2011 and 2018 [41].

There are several strengths of this study. First, it provides real-world evidence of utilisation in routine care by using routinely-collected national data available for all women aged 50 years and over with a registered diagnosis of IBC in England from 2014 to 2019. Second, linked patient-level data on drug utilisation were available up to 28 February 2022, providing good follow-up (at least 26 months). Third, it provides robust estimates of drug utilisation as the study used drug information captured in SACT. SACT is a national dataset of systemic therapy in cancer, with whole population coverage in England and 100% data completeness for the data items used in this study (drug name and administration date) [11, 42].

We are aware there are also some limitations. First, it was not possible to provide estimates of the use of NICE-recommended drugs in women aged under 50 years who had IBC, as data were only available for women aged 50 years and over. Second, data on hormone receptor status, HER2 status and stage were typically less complete as age increased. As this information is provided to NCRAS through an automatic pathology feed it is likely that lower completeness is reflective of a lack of testing of molecular markers which therefore are not available to inform treatment decisions in this group of older women. This should therefore not impact our findings on the rates of treatment use among different age groups defined according to this information. Third, as SACT data returns may be low for some NHS trusts estimates for the use of new drugs for recently diagnosed patients may be lower than in reality. Finally, it was difficult to define cohorts of eligible patients within the routine data for drugs recommended for use following progression, relapse, recurrence or previous treatment. The study cohort, representing relatively recently diagnosed cases is less likely to provide a reliable estimate of the use of these drugs in these clinical settings, and may underestimate their use in the overall population. In addition to this there was low utilisation of some drugs meaning that analysis of variation in use was not possible. This is something which would benefit from further research in the future, to identify any barriers to access.

For ribociclib and abemaciclib, the study found insufficient uptake following the NICE TAG publication to carry out ITSA and provide robust estimates of the impact of NICE guidance on utilisation. Future research should evaluate longer-term uptake of

NICE-recommended drugs and carry out ITSA to assess the impact of NICE TAG publication on use. Additionally future work to understand the extent of non-concordant use of NICE-recommended drugs would provide further insight into the drugs investigated within this study.

CONCLUSIONS

The translation of evidence from trials into routine care, beyond recommendations made in national treatment guidelines, is difficult to study but is of profound importance in efforts to improve population health. The findings of this population-based study looking at uptake of oncological drugs highlight varied utilisation of treatments recommended by NICE for IBC within the last 20 years. Additionally it highlights lower use of NICE-recommended drugs for first-line treatment as age increased, regardless of geographical region or comorbidity burden. Future work should further investigate geographical variation in access to new drugs, to identify areas of the country where routine access to new drugs is below what would be expected. Improving access to effective treatments is an important step in understanding IBC outcomes. At organisation level, NHS trusts are encouraged to perform local audit of NICE-recommended drugs to ensure patient fitness for treatment is assessed and age is not a barrier to access. Providing patients with clear information on NICE-recommended drugs may also improve engagement in decision-making where this is a contributing factor.

DATA AVAILABILITY

This work uses data that has been provided by patients and collected by the NHS as part of their care and support. The data for England are collated, maintained and quality assured by the National Disease Registration Service (NDRS), which is part of NHS England. Data on English Cancer Registrations can be accessed via the NHS Digital Data Access request Service (DARS) https://digital.nhs.uk/services/data-access-request-service-dars#national-disease-registration-service-ndrs.

CODE AVAILABILITY

All analyses were carried out using standard methods therefore the code is not provided.

REFERENCES

1. Dillon A, Landells LJ. NICE, the NHS, and Cancer Drugs. JAMA. 2018;319:767–8.
2. Anderson M, Drummond M, Taylor D, McGuire A, Carter P, Mossialos E. Promoting innovation while controlling cost: the UK's approach to health technology assessment. Health Policy. 2022;126:224–33.
3. Walker S, Palmer S, Sculpher M. The role of NICE technology appraisal in NHS rationing. Br Med Bull. 2007;81–82:51–64.
4. Leo CP, Hentschel B, Szucs TD, Leo C. FDA and EMA approvals of new breast cancer drugs-A comparative regulatory analysis. Cancers. 2020;12;437.
5. Bullement A, Podkonjak T, Robinson MJ, Benson E, Selby R, Hatswell AJ, et al. Real-world evidence use in assessments of cancer drugs by NICE. Int J Technol Assess Health Care. 2020;10:1–7.
6. Kerr AJ, Dodwell D, McGale P, Holt F, Duane F, Mannu G, et al. Adjuvant and neoadjuvant breast cancer treatments: a systematic review of their effects on mortality. Cancer Treat Rev. 2022;105:102375.
7. Sheldon TA, Cullum N, Dawson D, Lankshear A, Lowson K, Watt I, et al. What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. BMJ. 2004;329:999.
8. Fitzpatrick RW, Pate RG. Assessing the impact of NICE guidance on the prescribing of hormonal treatments of breast cancer in England. J Eval Clin Pract. 2015;21:759–61.
9. Curtis HJ, Walker AJ, Goldacre B. Impact of NICE guidance on tamoxifen prescribing in England 2011-2017: an interrupted time series analysis. Br J cancer. 2018;118:1268–75.
10. Henson KE, Brock R, Shand B, Coupland VH, Elliss-Brookes L, Lyratzopoulos G, et al. Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England. BMJ open. 2018;8:e020980.
11. Bright CJ, Lawton S, Benson S, Bomb M, Dodwell D, Henson KE, et al. Data resource profile: the Systemic Anti-Cancer Therapy (SACT) dataset. Int J Epidemiol. 2020;49:15–l.
12. Miller EM, Schwartzberg LS. Biosimilars for breast cancer: a review of HER2-targeted antibodies in the United States. Ther Adv Med Oncol. 2019;11:1758835919887044.
13. Armitage JN, van der Meulen JH. Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score. The. Br J Surg. 2010;97:772–81.
14. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol. 2017;46:348–55.
15. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. Stata J. 2015;15:480–500.
16. NICE. Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. 2006. Available from: https://www.nice.org.uk/guidance/ta107.
17. NICE. Early and locally advanced breast cancer: diagnosis and management. 2018. Available from: www.nice.org.uk/guidance/ng101.
18. NICE. Bevacizumab in combination with a taxane for the first-line treatment of metastatic breast cancer. 2011. Available from: https://www.nice.org.uk/guidance/ta214.
19. NICE. Fulvestrant for the treatment of locally advanced or metastatic breast cancer. 2011. Available from: https://www.nice.org.uk/guidance/ta239.
20. NICE. Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2. 2012. Available from: https://www.nice.org.uk/guidance/ta257.
21. NICE. Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer. 2012. Available from: https://www.nice.org.uk/guidance/ta263.
22. NICE. Fulvestrant for untreated locally advanced or metastatic oestrogen-receptor positive breast cancer. 2018. Available from: https://www.nice.org.uk/guidance/ta503.
23. NICE. Eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen. 2018. Available from: https://www.nice.org.uk/guidance/ta515.
24. Gannon MR, Dodwell D, Jauhari Y, Horgan K, Clements K, Medina J, et al. Initiation of adjuvant chemotherapy and trastuzumab for human epidermal growth receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England. J Geriatr Oncol. 2020;11:836–42.
25. Dodwell D, Jauhari Y, Gathani T, Cromwell D, Gannon M, Clements K, et al. Treatment variation in early breast cancer in the UK. BMJ. 2020;371:m4237.
26. World Health Organization. WHO model list of essential medicines: 19th list. 2015. Available from: https://www.iccp-portal.org/system/files/resources/EML2015_8-May-15.pdf.
27. Piezzo M, D'Aniello R, Avallone I, Barba B, Cianniello D, Cocco S, et al. Uptake of Trastuzumab Biosimilars for the treatment of HER2-positive breast cancer: a real-world experience from a cancer center. Pharmaceutics. 2021;13;684.
28. Blackwell K, Gligorov J, Jacobs I, Twelves C. The Global Need for a Trastuzumab Biosimilar for Patients With HER2-Positive Breast Cancer. Clin Breast Cancer. 2018;18:95–113.
29. Lythgoe MP, Desai A, Gyawali B, Savage P, Krell J, Warner JL, et al. Cancer therapy approval timings, review speed, and publication of pivotal registration trials in the US and Europe, 2010-2019. JAMA Netw Open. 2022;5:e2216183.
30. Cherla A, Naci H, Kesselheim AS, Gyawali B, Mossialos E. Assessment of coverage in England of cancer drugs qualifying for US food and drug administration accelerated approval. JAMA Intern Med. 2021;181:490–8.
31. NICE. Cancer Drugs Fund. Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/cancer-drugs-fund.
32. NHS England Cancer Drugs Fund Team. Appraisal and funding of cancer drugs from July 2016 (including the new Cancer Drugs Fund): a new deal for patients, taxpayers and industry. 2016. Available from: https://www.england.nhs.uk/wp-content/uploads/2013/04/cdf-sop.pdf.
33. Aggarwal A, Fojo T, Chamberlain C, Davis C, Sullivan R. Do patient access schemes for high-cost cancer drugs deliver value to society?-lessons from the NHS Cancer Drugs Fund. Annals of oncology : official journal of the European Society for. Med Oncol. 2017;28:1738–50.
34. Hall BJ, Bhojwani AA, Wong H, Law A, Flint H, Ahmed E, et al. Neoadjuvant Trastuzumab and Pertuzumab for Early HER2-positive breast cancer: a real world experience. Breast J. 2022;2022:7146172.
35. Fasching PA, Hartkopf AD, Gass P, Haberle L, Akpolat-Basci L, Hein A, et al. Efficacy of neoadjuvant pertuzumab in addition to chemotherapy and trastuzumab in routine clinical treatment of patients with primary breast cancer: a multicentric analysis. Breast cancer Res Treat. 2019;173:319–28.

36. Rath S, Elamarthi P, Parab P, Gulia S, Nandhana R, Mokal S, et al. Efficacy and safety of palbociclib and ribociclib in patients with estrogen and/or progesterone receptor positive, HER2 receptor negative metastatic breast cancer in routine clinical practice. PloS one. 2021;16:e0253722.
37. Patt D, Liu X, Li B, McRoy L, Layman RM, Brufsky A. Real-world treatment patterns and outcomes of Palbociclib Plus an aromatase inhibitor for metastatic breast cancer: flatiron database analysis. Clin Breast Cancer. 2022;22:601–10.
38. Lin J, McRoy L, Fisher MD, Hu N, Davis C, Mitra D, et al. Treatment patterns and clinical outcomes of palbociclib-based therapy received in US community oncology practices. Future Oncol. 2021;17:1001–11.
39. Boer K, Rubovszky G, Rokszin G, Abonyi-Toth Z, Foldesi C, Dank M. Demographic characteristics and treatment patterns among patients receiving Palbociclib for HR+/HER2- advanced breast cancer: a nationwide real-world experience. Onco Targets Ther. 2021;14:3971–81.
40. El Badri S, Tahir B, Balachandran K, Bezecny P, Britton F, Davies M, et al. Palbociclib in combination with aromatase inhibitors in patients >/= 75 years with oestrogen receptor-positive, human epidermal growth factor receptor 2 negative advanced breast cancer: a real-world multicentre UK study. Breast 2021;60:199–205.
41. Fu M, Naci H, Booth CM, Gyawali B, Cosgrove A, Toh S, et al. Real-world Use of and Spending on New Oral Targeted Cancer Drugs in the US, 2011-2018. JAMA Intern Med. 2021;181:1596–604.
42. National Cancer Registration and Analysis Service. Matching SACT to Cancer Waiting Times Data. 2014. Available from: http://www.ncin.org.uk/publications/data_briefings/sact_cwt.
43. NICE. Guidance on the use of trastuzumab for the treatment of advanced breast cancer. 2002. Available from: https://www.nice.org.uk/guidance/ta34.
44. NICE. Gemcitabine for the treatment of metastatic breast cancer. 2007. Available from: https://www.nice.org.uk/guidance/ta116.
45. NICE. Pertuzumab for the neoadjuvant treatment of HER2-positive breast cancer. 2016. Available from: https://www.nice.org.uk/guidance/ta424.
46. NICE. Pertuzumab with trastuzumab and docetaxel for treating HER2-positive breast cancer. 2018. Available from: https://www.nice.org.uk/guidance/ta509.
47. NHS England. NHS England releases update to national Cancer Drugs Fund list. 2013. Available from: https://www.england.nhs.uk/2013/05/cdf-update/.
48. NICE. Pertuzumab for adjuvant treatment of HER2-positive early stage breast cancer. 2019. Available from: https://www.nice.org.uk/guidance/ta569.
49. NICE. Everolimus with exemestane for treating advanced breast cancer after endocrine therapy. 2016. Available from: https://www.nice.org.uk/guidance/ta421.
50. NICE. Eribulin for treating locally advanced or metastatic breast cancer after 2 or more chemotherapy regimens. 2016. Available from: https://www.nice.org.uk/guidance/ta423.
51. NICE. Trastuzumab emtansine for treating HER2-positive advanced breast cancer after trastuzumab and a taxane. 2017. Available from: https://www.nice.org.uk/guidance/ta458.
52. NICE. Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. 2017. Available from: https://www.nice.org.uk/guidance/ta495.
53. NICE. Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. 2017. Available from: https://www.nice.org.uk/guidance/ta496.
54. NICE. Abemaciclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer. 2019. Available from: https://www.nice.org.uk/guidance/ta563.
55. NICE. Neratinib for extended adjuvant treatment of hormone receptor-positive, HER2-positive early stage breast cancer after adjuvant trastuzumab. 2019. Available from: https://www.nice.org.uk/guidance/ta612.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of the entire study: MRG. study concepts and design: MRG, DD, AA, MHP, KM, KH, KC, JM, DAC. literature research: MRG data acquisition: MRG, KC, JM. statistical analysis: MRG. manuscript preparation: MRG. manuscript editing: MRG, DD, AA, MHP, KM, KH, KC, JM, DAC. All authors were involved in data interpretation, critical appraisal of the draught manuscript, and gave final approval on the submitted version.

COMPETING INTERESTS

This study was undertaken as part of the work by the National Audit of Breast Cancer in Older Patients (NABCOP). The NABCOP is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit and Patient Outcomes Programme, and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). Neither the commissioner nor the funders had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. The authors had full independence from the HQIP. The aim of the NABCOP is to evaluate the care of older women with breast cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these women. More information can be found at www.nabcop.org.uk. DD also receives funding from Cancer Research UK (grant C8225/A21133). AA acknowledges funding from the NIHR Advanced Fellowship NIHR300599. KH declares being the Chair of the Endonet Trial Steering Committee. KC declares funding received for Breast Cancer Research Manager role within NHS England as part of the Cancer Grand Challenges PRECISION team, which is funded by Cancer Research UK and the Dutch Cancer Society (C7827/A27366); the grant is not related to her work on the National Audit of Breast Cancer in Older Patients or this paper. DAC declares grants/contracts from Healthcare Quality Improvement partnership; participation on the Pregnancy Outcome Prediction Study (POPS2) Trial Steering Committee; being on the Editorial Committee for the Journal of Health Services Research and Policy; being Deputy Chair on the Examination Committee for the MSc, PG Diploma and PG Cert in Public Health distance learning programme at the London School of Hygiene & Tropical Medicine. There are no other relationships or activities that could appear to have influenced the submitted work.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was exempt from NHS Research Ethics Committee approval because it involved analysis of pseudonymised linked data collated for the purpose of service evaluation as part of the National Audit of Breast Cancer in Older Patients.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02439-z.

Correspondence and requests for materials should be addressed to Melissa Ruth Gannon.

Reprints and permission information is available at http://www.nature.com/reprints

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

5.2. Research Paper 4

Title

Initiation of adjuvant chemotherapy and trastuzumab for human epidermal growth factor receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England.

Journal details

Published in Journal of Geriatric Oncology. The online PDF version can be accessed at: https://pubmed.ncbi.nlm.nih.gov/32007402/

Supplementary material can be found in the Appendices.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Journal of Geriatric Oncology		
When was the work published?	30 January 2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	
Date	

Supervisor Signature	
Date	

Journal of Geriatric Oncology 11 (2020) 836–842

Contents lists available at ScienceDirect

Journal of Geriatric Oncology

Initiation of adjuvant chemotherapy and trastuzumab for human epidermal growth receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England

Melissa Ruth Gannon [a,b,*], David Dodwell [c], Yasmin Jauhari [b,d], Kieran Horgan [e], Karen Clements [f], Jibby Medina [b], David Alan Cromwell [a,b]

[a] Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK
[b] Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK
[c] Nuffield Department of Population Health, University of Oxford, Oxford, UK
[d] St Georges Healthcare NHS Trust, London, UK
[e] Department of Breast Surgery, St James's University Hospital, Leeds, UK
[f] National Cancer Registration and Analysis Service, Public Health England, 1st Floor, 5 St Philip's Place, Birmingham, UK

ARTICLE INFO

Article history:
Received 22 October 2019
Received in revised form 10 December 2019
Accepted 6 January 2020
Available online 30 January 2020

Keywords:
HER2-positive
Early breast cancer
Adjuvant trastuzumab
Older women
Regional variation

ABSTRACT

Background: Clinical guidance on recommended treatment for older patients with breast cancer is often ambiguous, particularly in the context of comorbidities and poor functional status. Older patients, aged 70 years and over, account for a substantial proportion of women with breast cancer yet are underrepresented in randomized controlled trials. This paper investigates the initiation of adjuvant chemotherapy and trastuzumab in older patients in routine care.

Materials and methods: Women, aged 50 years and over, newly diagnosed with human epidermal growth receptor 2 (HER2)-positive early invasive breast cancer from January 2014 to December 2017 were identified from the England Cancer Registry. Chemotherapy and trastuzumab use was obtained from the Systemic Anti-Cancer Therapy (SACT) dataset. Patient and tumor characteristics influential in treatment decision-making were included in multilevel mixed-effects logistic regression models.

Results: 10% of women had HER2-positive tumors. Initiation of adjuvant chemotherapy and trastuzumab decreased with age from ≥70% among women aged 50–64 years to <15% among women aged 80+ years. Initiation varied additionally by tumor characteristics and number of comorbidities. Age remained a factor in treatment decisions despite favorable other factors, with lower use among women aged 70+ years. There was also marked variation across geographical regions.

Conclusions: In women with operable HER2-positive early invasive breast cancer, adjuvant chemotherapy plus trastuzumab was started less frequently as age increased, regardless of tumor characteristics or comorbidity burden. There was substantial variation in the proportion of women who started these treatments across the country.

1. Introduction

Adjuvant (post-surgical) chemotherapy is a well-established treatment for early breast cancer, with evidence of its efficacy derived from multiple randomized controlled trials (RCTs) and subsequent meta-analyses [1]. In the UK, the National Institute for Health and Care Excellence (NICE) recommends that adjuvant treatment decisions should be based on a balance between the risks and benefits of chemotherapy, particularly in people with comorbidities [2]. For patients with human epidermal growth receptor 2 (HER2)-positive breast cancer, the HERceptin Adjuvant (HERA) trial demonstrated a strong efficacy benefit of trastuzumab, with a subsequent Cochrane review of eight RCTs reporting a hazard ratio (HR) of 0.66 for improvement in overall survival [3,4]. As such, the European Society of Medical Oncology (ESMO) guidelines recommend treatment with chemotherapy and trastuzumab, regardless of estrogen receptor (ER) status [5]. This is echoed in the NICE guidelines and the American Society of Clinical Oncology (ASCO) guidelines as well as the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA) recommendations for the management of older patients with breast cancer [2,6–8].

* Corresponding author at: London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, London WC1H 9SH, United Kingdom.
E-mail address: melissa.gannon@lshtm.ac.uk (M.R. Gannon).

https://doi.org/10.1016/j.jgo.2020.01.005
1879-4068/

The randomized trials on which guidance is based largely underrepresent patients aged 70 and over [9,10]. Specifically, the maximum age of patients enrolled into the eight RCTs providing evidence of the efficacy of trastuzumab was 80 years at the point of randomization. This means that clinical guidelines for older patients are often less definitive than for their younger counterparts. A recent report from Cancer Research UK highlights a need for better evidence of treatment effectiveness in older patients [11]. Outside of the trial setting, age is a well-documented risk factor for receipt of non-standard treatment and there have been differences reported in the rates of access to treatments among older patients, with subsequent poorer outcomes [12–19].

>50,000 women in England are diagnosed with breast cancer each year, with one third of cases occurring in women aged 70 and over [20]. With an aging population, this number is projected to rise, increasing the need for evidence to support treatment use in older breast cancer populations [21]. In a population where trials have been initiated and often failed to recruit [22,23], an alternative approach is to use the wealth of routinely collected health data to evaluate the use of adjuvant therapies for breast cancer. Indeed a recent study in the United States used the Surveillance, Epidemiology and End Results (SEER)-Medicare linked dataset to investigate disparities in treatment provision, and reported that approximately a half of patients aged 65 years and older did not receive trastuzumab for Stage I-III HER2-positive breast cancer [24]. To our knowledge, there is no such study which has reported the prevalence of HER2-positive early breast cancer in older women and has identified predictors of adjuvant chemotherapy and trastuzumab initiation among such eligible women in a large population-based English cohort.

The aim of this study was to investigate the initiation of adjuvant chemotherapy and trastuzumab by age, in older women with HER2-positive early invasive breast cancer newly diagnosed and treated in England, to assess whether there is variation in the decision to use post-operative chemotherapy and trastuzumab across England. We did not consider the duration of adjuvant treatment, nor its impact on survival.

2. Materials and Methods

2.1. Data

This population-based cohort study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP). Data for this study are based on patient-level information collected by the National Health Service (NHS), as part of the care and support of patients with cancer. The data are collated, maintained and quality assured by the National Cancer Registration and Analysis Service (NCRAS), which is part of Public Health England (PHE). Cancer Registry data was provided, linked to the Cancer Outcomes and Services Dataset (COSD) for details of patient and tumor characteristics at diagnosis; the Hospital Episode Statistics (HES) database which captured all NHS hospital admissions, for details of surgical procedure; and the Systemic Anti-Cancer Therapy (SACT) database for information on chemotherapy and trastuzumab. The study is exempt from UK National Research Ethics Committee approval as it involved secondary analysis of an existing dataset of pseudonymized data. The NABCOP has approval for processing health care information under Section 251 (reference number: 16/CAG/0079) for all NHS patients aged 50 years and over diagnosed with breast cancer in England and Wales.

2.2. Study Population

Linked, pseudonymized patient records were provided for all women aged 50 years and over, with a new diagnosis of breast cancer between 1 January 2014 and 31 December 2017, diagnosed and treated within a NHS trust in England. For this analysis, we identified all women with HER2-positive early invasive breast cancer (ICD-10 code C50; stage 1-3A UICC TNM staging classification, 7th edition), who received primary surgery with no prior (i.e. neoadjuvant) chemotherapy or trastuzumab. Women receiving neoadjuvant chemotherapy were not included, to avoid the possible inclusion of patients with locally advanced disease and to allow a better appreciation of the initiation of chemotherapy and trastuzumab in the context of pathological variables unaffected by any previous systemic therapy. Women were classified as having HER2-positive breast cancer where the HER2 status was reported as either positive or borderline but with a positive HER2-FISH (fluorescence in situ hybridization) or equivalent test result. Primary surgery was defined as either breast conserving surgery (BCS) or mastectomy that occurred within six months of diagnosis, and was identified from Office of Population Censuses and Surveys (OPCS) procedure codes entered within the HES patient records (see Appendix Table A1). Patients were allocated to the NHS trust where they were diagnosed. If this information was unavailable, the NHS trust of (1) surgery or (2) MDT was used.

2.3. Adjuvant Chemotherapy and Trastuzumab

The initiation of adjuvant chemotherapy and trastuzumab was identified primarily from the SACT dataset. Treatment with chemotherapy and trastuzumab was defined as adjuvant where the first cycle was started within six months following primary surgery. Records of chemotherapy administration within HES and the England Cancer Registry/COSD treatment files were examined where adjuvant chemotherapy was not reported in SACT.

2.4. Explanatory Variables

The following patient and tumor characteristics were considered likely to inform treatment decisions: age at diagnosis (years), tumor stage (T1, T2, T3), nodal stage (N0, N1, N2), ER status (positive or negative), tumor grade (G1, G2, G3), social deprivation (1–5), number of comorbidities (0, 1, 2+), and performance status (0, 1, 2–4).

Social deprivation was measured using the Index of Multiple Deprivation (IMD) 2015 rank [25] which was derived from the patient's postcode at diagnosis. The IMD rank was grouped into quintiles from most (group 1) to least (group 5) income deprived.

Comorbidity burden was defined using the Royal College of Surgeon's Charlson Comorbidity Index [26]. This counts the presence of specific chronic medical conditions (excluding malignancy), identified using ICD-10 diagnosis codes within patient HES records for a period of two years prior to diagnosis [27].

2.5. Statistical Analysis

The proportion of women who started adjuvant chemotherapy and trastuzumab was calculated for the overall cohort and within patient subgroups. The statistical significance of differences between group proportions was assessed using a *t*-test or chi-square test, as appropriate.

A multilevel mixed-effects logistic regression model was developed to describe the relationship between receipt of adjuvant chemotherapy / trastuzumab and the patient factors. The model included age at diagnosis, tumor stage, nodal stage, ER status, tumor grade, social deprivation, and number of comorbidities. A spline was used to describe the relationship between treatment and age, with a knot defined at age = 70 years. For women aged 50–69 years, the spline was simply a linear term; for women aged 70+ years, the spline also included a quadratic term. The model was found to have good prognostic performance, both in terms of its discrimination (concordance (C) statistic / ROC value of 0.846) and calibration (see Appendix Fig. A2). The proportions of women with different characteristics starting adjuvant chemotherapy / trastuzumab were predicted from the model as marginal effects (achieved using the margins command in Stata).

A multilevel model was used to account for the clustering of patients within NHS trusts and geographical regions (in which NHS trusts were aggregated into Cancer Alliances). Due to the relatively low levels of activity at NHS trust level, the multilevel model was limited to geographical region, with each Alliance fitted as a random intercept [28]. These represent differences between Alliances that are not explained by the patient factors in the model.

As these variables contained few missing values, we conducted a complete case analysis as the primary analysis. We conducted a sensitivity analysis in which categories of "unknown" were created where data items had missing, unintelligible or conflicting information.

3. Results

Of 103,568 women, aged 50 years and over, diagnosed with early invasive breast cancer in England between January 2014 and December 2017, there were 10,109 (10%) with HER2-positive tumors. Prevalence of HER2-positive cancer was found to be slightly higher in younger women (those aged 50–69 years at diagnosis) at 11% compared with 8% among older women aged 70+ years. Of these, 7471 women received primary surgery within six months of diagnosis (97% received surgery within three months after diagnosis), with no prior chemotherapy or trastuzumab. A final total of 6780 women (91%) had data on the patient and tumor characteristics included in the regression models, and these complete cases formed the primary analysis. Full details of patient selection are shown in Fig. A1.

Table 1 provides detail of patient and tumor characteristics for the cohort. Over two-thirds of patients had ER-positive cancers and one-third had malignant lymph nodes. Older women tended to have larger tumors and nodal involvement. Multiple comorbidities were also more prevalent. Overall, 60% ($N = 4051$) of women were identified as having started adjuvant chemotherapy and trastuzumab (Table 2). As expected, there was greater initiation among women with higher grade tumors, and with higher T and N stage disease. Rates of treatment initiation fell as age increased, and were also lower among women with more comorbid conditions.

Table 1
Patient and tumor characteristics in women with HER2-positive, early invasive breast cancer, diagnosed in NHS trusts in England between January 2014 and December 2017 and receiving primary surgery, overall and by age at diagnosis.

Characteristic	All patients	50–69 years	70+ years	*P*-value
	N (%)	*N* (%)	*N* (%)	(chi-squared test)
	$N = 6780$	$N = 4630$	$N = 2150$	–
Stage at diagnosis				
Stage 1	3015 (44%)	2318 (50%)	697 (32%)	$p < .0001$
Stage 2	3246 (48%)	2029 (44%)	1217 (57%)	
Stage 3A	519 (8%)	283 (6%)	236 (11%)	
T stage				
T1	3630 (54%)	2776 (60%)	697 (32%)	$p < .0001$
T2	2903 (43%)	1722 (37%)	1217 (57%)	
T3	247 (4%)	132 (3%)	236 (11%)	
N stage				
N0	4613 (68%)	3263 (70%)	1350 (63%)	$p < .0001$
N1	1736 (26%)	1135 (25%)	601 (28%)	
N2	431 (6%)	232 (5%)	199 (9%)	
ER status				
Positive	4875 (72%)	3433 (74%)	1442 (67%)	$p < .0001$
Negative	1905 (28%)	1197 (26%)	708 (33%)	
Tumor grade				
G1	218 (3%)	162 (3%)	56 (3%)	$p < .0001$
G2	2703 (40%)	1936 (42%)	767 (36%)	
G3	3859 (57%)	2532 (55%)	1327 (62%)	
Charlson comorbidity index				
0	5936 (88%)	4204 (91%)	1732 (81%)	$p < .0001$
1	613 (9%)	327 (7%)	286 (13%)	
2+	231 (3%)	99 (2%)	132 (6%)	
IMD quintiles				
1 - most deprived	1025 (15%)	721 (16%)	304 (14%)	$p = .218$
2	1227 (18%)	852 (18%)	375 (17%)	
3	1376 (20%)	915 (20%)	461 (21%)	
4	1594 (24%)	1095 (24%)	499 (23%)	
5 - least deprived	1558 (23%)	1047 (23%)	511 (24%)	

Note: Percentages may not add to 100% due to rounding.

The pattern of initiation among women with different combinations of factors is described in Fig. 1. It shows that tumor characteristics (T stage and ER status) seem to play a secondary role compared with the influence of patient age, even taking account of the presence of comorbidity. In particular, a lower proportion of older women started chemotherapy and trastuzumab, even where no comorbidity was recorded.

Fig. 2 shows how the initiation of adjuvant chemotherapy and trastuzumab varied by age across the 22 geographical regions, after adjusting for the other patient factors. Among younger women (aged 50–69 years), there was considerable variation between the Alliances in the pattern of treatment. These differences continued among women aged 70–79 years before diminishing as the age at diagnosis increased further. Only among women aged 85+ years did the rates become more similar, being at a low level in all regions.

Among those women for whom chemotherapy and trastuzumab was initiated, use of anthracyclines was observed to vary by age, with use decreasing as age increased. Specifically, rates varied from 71% among women aged 50-59 yrs.; 64% among women aged 60-69 yrs.; 45% among women aged 70-79 yrs.; to 13% among women aged 80 + yrs.

The results of the sensitivity analysis were similar to those in the primary analysis. Including those women with "unknown" information for the patient and tumor characteristics in the model did not change the conclusion that age has a strong, negative association with the initiation of adjuvant chemotherapy and trastuzumab, independent of other patient factors.

4. Discussion

This study examined the initiation of adjuvant chemotherapy and trastuzumab among women with breast cancer, a therapy which clinical trials have proven to be effective, both in terms of delaying time to recurrence and lengthening overall survival. This study shows that the initiation of these therapies is high among younger women with HER2-positive early invasive breast cancer, a patient population that largely corresponds to the patients enrolled in the clinical trials. The analysis also shows lower rates of adjuvant chemotherapy and trastuzumab among the older patients. This could reflect the impact of reduced levels of patient fitness but the patterns of treatment were not wholly consistent with this interpretation. First, the initiation of adjuvant chemotherapy and trastuzumab among older women was observed to be low for those women with no comorbidity burden, and in spite of the fact that all women in the cohort were considered fit to have received surgery. Second, the initiation of such adjuvant treatment was observed to vary across geographical regions in England. This persisted after adjustment for measured patient and tumour characteristics suggesting further regional factors responsible for variation; these might include factors relating to work force, funding for adjuvant treatment and cultural differences in shared decision-making.

The findings in this study are in line with similar research in this setting, which has shown marked variation in the treatment of older women compared with younger women, both in terms of primary and adjuvant treatment [15,24,29–31]. Several studies have been conducted in the United States using the SEER-Medicare dataset to look at disparity in the use of targeted therapy for breast cancer along with other treatments such as surgery and radiotherapy. In 2013, an analysis published by Reeder-Hayes et al. found that among women with HER2-positive early breast cancer, who were aged 85+, years 15% received adjuvant trastuzumab, compared with 60% of women aged 65–74 years [24].

Table 2
Proportion of women starting adjuvant chemotherapy and trastuzumab, by baseline characteristic; odds ratios (OR) from multilevel mixed-effects (MLME) logistic regression models.

Characteristic	Number of patients	% starting chemotherapy and trastuzumab	Unadjusted OR[a] (95% CI)	Adjusted OR[a] (95% CI)	Grouped *p*-value
	6780	60%	–	–	
Age Groups – 5-year bands					
50–54 yrs	1221	76%	1.00	–	
55–59 yrs	1110	74%	0.92 (0.75–1.13)	–	
60–64 yrs	1095	70%	0.70 (0.57–0.86)	–	
65–69 yrs	1204	63%	0.50 (0.41–0.60)	–	
70–74 yrs	820	57%	0.37 (0.30–0.46)	–	
75–79 yrs	660	36%	0.14 (0.11–0.18)	–	
80–84 yrs	428	14%	0.03 (0.02–0.05)	–	
85+ yrs	242	1%	0.00 (0.00–0.01)	–	
Age (continuous)					
Age spline: 50–69 yrs	–	–	–	0.95 (0.94–0.96)	<.0001
Age spline: 70+ yrs	–	–	–	0.89 (0.84–0.95)	<.0001
Squared age spline 70+ yrs	–	–	–	0.99 (0.98–0.99)	<.0001
T stage					
T1	3630	59%	1.00	1.00	<.0001
T2	2903	60%	1.11 (1.00–1.24)	1.43 (1.25–1.63)	
T3	247	64%	1.23 (0.93–1.64)	1.89 (1.30–2.74)	
N stage					
N0	4613	57%	1.00	1.00	<.0001
N1	1736	65%	1.52 (1.34–1.72)	1.64 (1.41–1.90)	
N2	431	68%	1.63 (1.30–2.04)	2.46 (1.84–3.30)	
ER status					
Positive	4875	59%	1.00	1.00	<.0001
Negative	1905	63%	1.27 (1.13–1.42)	1.41 (1.22–1.63)	
Tumor grade					
G1	218	35%	0.23 (0.17–0.31)	0.17 (0.12–0.23)	<.0001
G2	2703	53%	0.57 (0.51–0.64)	0.49 (0.43–0.55)	
G3	3859	66%	1.00	1.00	
Charlson comorbidity index					
0	5936	62%	1.00	1.00	<.0001
1	613	46%	0.49 (0.41–0.59)	0.68 (0.55–0.85)	
2+	231	32%	0.26 (0.19–0.35)	0.34 (0.24–0.49)	
IMD quintiles					
1 - most deprived	1025	59%	1.00	1.00	.0201
2	1227	59%	1.04 (0.86–1.25)	1.10 (0.89–1.36)	
3	1376	59%	1.14 (0.95–1.37)	1.34 (1.08–1.66)	
4	1594	61%	1.23 (1.03–1.47)	1.35 (1.09–1.66)	
5 - least deprived	1558	60%	1.16 (0.97–1.39)	1.29 (1.04–1.60)	

[a] MLME model with NHS trust as the cluster level.

An earlier study in 2010 looking at variation in initial treatments received, by Schonberg et al., found that among women diagnosed with early (stage I-II) breast cancer, the effect of age on receipt of treatment was stronger than the effect of a patient having multiple comorbidities [15]. This mirrors the findings from an integrated health care system, reported by Enger et al. in 2010, highlighting women aged 80+ years were nearly six times more likely to receive non-standard treatment for Stage I-II breast cancer, when compared to women aged 65–69 years [29]. Considering studies conducted in England, an analysis of all women presenting with primary invasive breast cancer in 2007, found that after the age of 70 years women were increasingly less likely to receive surgery [30]. A more recent study, concentrating on two English cancer registry areas, found that among women aged 70–79 years with stage I-III (or unknown) breast cancer treated with surgery for each additional year of age, over the age of 70, the odds of receiving chemotherapy reduced by 24% [31].

The study has a number of strengths. It used a large, population-based sample, including women diagnosed over a period of four years. Additionally, the data related to women diagnosed within the last five years and so reflects current treatment patterns. Finally, the dataset contained sufficient patient and tumor characteristics associated with treatment decisions to produce a regression model with good discrimination and calibration.

There are various limitations of this study. Of primary concern is the completeness and accurate reporting of adjuvant treatments within the SACT database. Among women aged 50–69 years with HER2-positive early invasive breast cancer who received surgery, 76% were recorded to have received any drug treatment. There is potential for under-reporting. The SACT database does not include treatments delivered in private hospitals, although this corresponds to a small proportion of care within England. Case-ascertainment may also be incomplete from some NHS trusts. This might have lowered the recorded absolute rates of use but there is no reason to believe it would produce either the strong association with age or the extent of the regional variation. The proportion of women considered to have received adjuvant chemotherapy and trastuzumab included 803/4051 (20%) women for whom only adjuvant chemotherapy was reported in SACT, along with a further 98/4051 (2%) women for whom only adjuvant trastuzumab was reported in SACT. Including these 803 women in the surgery-only group as a sensitivity analysis made no difference to the findings. Of those 98 women with only trastuzumab details reported in SACT, investigation from other sources including HES and COSD/England Cancer Registry suggested that 82/98 (84%) of such women received adjuvant chemotherapy within six months of surgery. Additionally, it is noted that SACT provides data on prescribed therapies meaning there may be patients considered in this analysis as

Age at diagnosis	T stage	ER-positive Charlson comorbidity score 0	ER-positive Charlson comorbidity score 1	ER-positive Charlson comorbidity score 2+	ER-negative Charlson comorbidity score 0	ER-negative Charlson comorbidity score 1	ER-negative Charlson comorbidity score 2+
55 yrs	T1	73%	67%	53%	79%	73%	60%
	T2	79%	73%	60%	83%	78%	67%
	T3	83%	77%	66%	87%	82%	72%
65 yrs	T1	64%	57%	43%	71%	64%	50%
	T2	71%	64%	50%	76%	70%	57%
	T3	76%	69%	56%	81%	75%	63%
75 yrs	T1	46%	39%	26%	53%	45%	32%
	T2	54%	46%	32%	61%	53%	39%
	T3	59%	52%	38%	66%	59%	44%
85 yrs	T1	5%	4%	2%	7%	5%	3%
	T2	7%	5%	3%	10%	7%	4%
	T3	9%	7%	4%	12%	9%	5%

Fig. 1. Predicted initiation of adjuvant chemotherapy and trastuzumab, from a multilevel mixed-effects logistic regression model, across four patient and tumour characteristics. Note: Predictions based on women diagnosed between 2014 and 2017 and derived from a multilevel mixed-effects logistic regression model shown in Table 2. Higher percentages are shown in dark blue with a gradient down to light blue for lowest percentages. N stage, grade and IMD included at overall means.

having started adjuvant therapy where they did not receive it; however as this study aimed to look at variation around treatment decisions the data provided is informative.

Another concern is the potential for errors in patient and tumor characteristics within the England Cancer Registry and COSD datasets. The cancer registration service has various validation steps when compiling the national registration data and the overall effect of coding errors should be small. It is also possible that differing indications for the use of neoadjuvant therapy between trusts over time may have affected the risk profiles of patients being considered for post-operative systemic therapies.

Previous research findings have also noted the impact of unmeasured factors in receipt of treatment and that, in order to fully measure variation in treatment utilisation, the potential confounding effect of factors such as patient choice should be adjusted for [32]. The study described here was unable to include all patient factors that influence treatment decisions, such as performance status, expected tolerability of treatment, patient frailty, and preference [33–35]. Omission of these factors from the regression model would reduce its level of discrimination. However, putative differences in the prevalence of these characteristics among NHS trusts are unlikely to account for the large variation observed between regions.

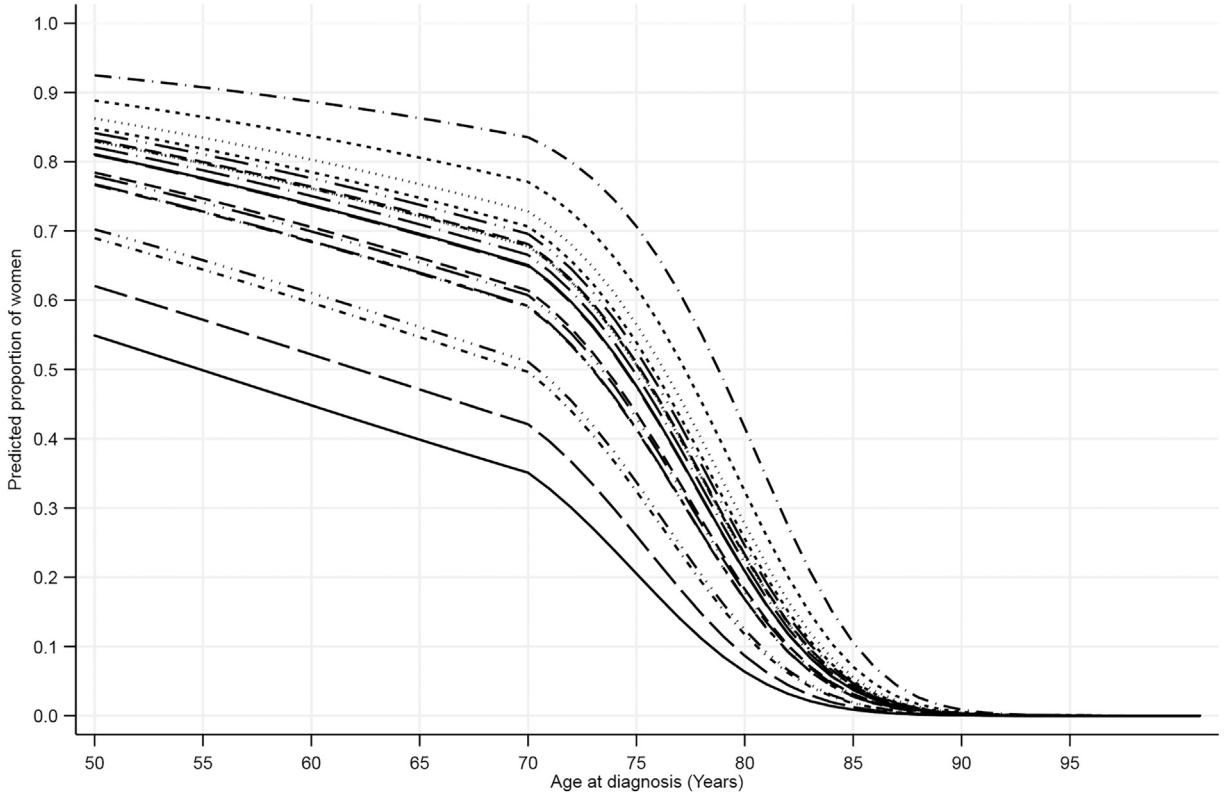


Fig. 2. Predicted initiation of adjuvant chemotherapy and trastuzumab, from a multilevel mixed-effects logistic regression model, by age at diagnosis within English geographical regions. Note: Each line in the above figure is a geographical region within England, defined based on Cancer Alliance. Predictions based on women diagnosed between 2014 and 2017 and derived from a multilevel mixed-effects logistic regression model shown in Table 2.

5. Conclusions

The findings of this study show that fewer older women with operable HER2-positive early breast cancer start the most targeted oncological systemic treatment. While the initiation of adjuvant chemotherapy was observed to vary by tumor characteristics, these factors did not seem to be the dominant reason as to why a patient did not start chemotherapy and trastuzumab. Instead, the rates were strongly associated with age at diagnosis independent of these clinical factors and the presence of comorbidities. This fact, together with the variation in the initiation of adjuvant chemotherapy across regions, suggests there is a need for breast cancer teams to review chemotherapy provision and the criteria for selecting patients. This may lead to a reduction in the unexplained variation in the initiation of adjuvant treatment in older women.

Ethics approval and consent to participate

The study is exempt from UK National Research Ethics Committee approval as it involved secondary analysis of an existing dataset of pseudonymized data. The NABCOP has approval for processing health care information under Section 251 (reference number: 16/CAG/0079) for all NHS patients aged 50 years and over diagnosed with breast cancer in England and Wales.

Consent for publication

Not applicable.

Availability of data and material

This work uses patient data that has been provided by, or derived from, patients and collected by the NHS as part of their care and support. The data are collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). Access to the data was facilitated by the PHE Office for Data Release. No additional data are available. Data on English cancer registrations can be accessed via the Office for Data Release at Public Health England. https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data.

Funding

This study was undertaken as part of the work by the National Audit of Breast Cancer in Older Patients (NABCOP). The Audit is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the *National Clinical Audit and Patient Outcomes Programme*, and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). Neither the commissioner nor the funders had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. The authors had full independence from the HQIP. The aim of the NABCOP is to evaluate the care of older women with breast cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these women. More information can be found at www.nabcop.org.uk.

Authors' contributions

M.R.G., D.D., and D.A.C were responsible for study design. M.R.G. conducted the statistical analyses and drafted the manuscript. All authors were involved in data interpretation, critical appraisal of the draft manuscript and gave final approval on the version to be published.

Declaration of Competing Interest

None

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgo.2020.01.005.

References

[1] EBCTCG. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 2005;365(9472):1687–717 [Epub 2005/05/17].
[2] NICE. Early and locally advanced breast cancer: diagnosis and management. Available from: www.nice.org.uk/guidance/ng101; 2018.
[3] Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 2005;353(16):1659–72 [Epub 2005/10/21].
[4] Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, et al. Trastuzumab containing regimens for early breast cancer. Cochrane Database Syst Rev 2012;4 CD006243. [Epub 2012/04/20].
[5] Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26(Suppl. 5) v8–30. [Epub 2015/09/01].
[6] Biganzoli L, Wildiers H, Oakman C, Marotti L, Loibl S, Kunkler I, et al. Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). Lancet Oncol 2012;13(4) e148–60. [Epub 2012/04/04].
[7] Denduluri N, Somerfield MR, Eisen A, Holloway JN, Hurria A, King TA, et al. Selection of optimal adjuvant chemotherapy regimens for human epidermal growth factor receptor 2 (HER2) – negative and adjuvant targeted therapy for HER2-positive breast cancers: an American Society of Clinical Oncology Guideline Adaptation of the Cancer Care Ontario Clinical Practice Guideline. J Clin Oncol 2016; 34(20):2416–27.
[8] Giordano SH, Temin S, Chandarlapaty S, Crews JR, Esteva FJ, Kirshner JJ, et al. Systemic therapy for patients with advanced human epidermal growth factor receptor 2-positive breast cancer: ASCO Clinical Practice Guideline Update. J Clin Oncol Off J Am Soc Clin Oncol 2018;36(26):2736–40 [Epub 2018/06/26].
[9] Hutchins LF, Unger JM, Crowley JJ, Coltman Jr CA, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. N Engl J Med 1999;341 (27):2061–7 [Epub 1999/12/30].
[10] Ludmir EB, Mainwaring W, Lin TA, Miller AB, Jethanandani A, Espinoza AF, et al. Factors associated with age disparities among cancer clinical trial participants. JAMA Oncol 2019;5(12):1769–73. https://doi.org/10.1001/jamaoncol.2019.2055.
[11] CRUK. Advancing years: treating and caring for an ageing population. Available from: www.cancerresearchuk.org/about-us/we-develop-policy/our-policy-on-access-to-cancer-treatments/treating-and-caring-for-an-ageing-population; 2018.
[12] Bouchardy C, Rapiti E, Fioretta G, Laissue P, Neyroud-Caspar I, Schafer P, et al. Undertreatment strongly decreases prognosis of breast cancer in elderly women. J Clin Oncol 2003;21(19):3580–7 [Epub 2003/08/13].
[13] Lavelle K, Todd C, Moran A, Howell A, Bundred N, Campbell M. Non-standard management of breast cancer increases with age in the UK: a population based cohort of women > or =65 years. Br J Cancer 2007;96(8):1197–203 [Epub 2007/03/28].
[14] Ring A. The influences of age and co-morbidities on treatment decisions for patients with HER2-positive early breast cancer. Crit Rev Oncol Hematol 2010;76(2):127–32 [Epub 2010/01/26].
[15] Schonberg MA, Marcantonio ER, Li D, Silliman RA, Ngo L, McCarthy EP. Breast cancer among the oldest old: tumor characteristics, treatment choices, and survival. J Clin Oncol 2010;28(12):2038–45 [Epub 2010/03/24].
[16] Foot C, Harrison T. How to improve cancer survival: Explaining England's relatively poor rates. Available from: www.kingsfund.org.uk/sites/default/files/How-to-improve-cancer-survival-Explaining-England-poor-rates-Kings-Fund-June-2011.pdf; 2011.
[17] Lavelle K, Sowerbutts AM, Bundred N, Pilling M, Degner L, Stockton C, et al. Is lack of surgery for older breast cancer patients in the UK explained by patient choice or poor health? A prospective cohort study. Br J Cancer 2014;110(3):573–83 [Epub 2013/12/03].
[18] Public Health England and Association of Breast Surgery: An audit of screen detected breast cancers for the year of screening April 2013 to March 2014. Available from. https://associationofbreastsurgery.org.uk/media/1108/nhs-bsp-abs-audit-2013-14.pdf; 2015.
[19] Mislang AR, Biganzoli L. Adjuvant systemic therapy in older breast cancer women: can we optimize the level of care? Cancers 2015;7(3):1191–214 [Epub 2015/07/08].
[20] ONS. Cancer registration statistics. England; 2017.
[21] The Lancet O. Not old, just older: considering age in cancer care. Lancet Oncol 2019; 20(7):887 [Epub 2019/07/04].
[22] Leonard R, Ballinger R, Cameron D, Ellis P, Fallowfield L, Gosney M, et al. Adjuvant chemotherapy in older women (ACTION) study - what did we learn from the pilot phase? Br J Cancer 2011;105(9):1260–6 [Epub 2011/10/13].
[23] Crivellari D, Gray KP, Dellapasqua S, Puglisi F, Ribi K, Price KN, et al. Adjuvant pegylated liposomal doxorubicin for older women with endocrine nonresponsive

breast cancer who are NOT suitable for a "standard chemotherapy regimen": the CASA randomized trial. Breast 2013;22(2):130–7 [Epub 2013/03/05].
[24] Reeder-Hayes K, Peacock Hinton S, Meng K, Carey LA, Dusetzina SB. Disparities in use of human epidermal growth hormone receptor 2-targeted therapy for early-stage breast cancer. J Clin Oncol 2016;34(17):2003–9 [Epub 2016/04/14].
[25] The English Indices of Deprivation. Statistical Release 2015. Available from: www.gov.uk/government/statistics/english-indices-of-deprivation-2015; 2015.
[26] 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(5):772–81 [Epub 2010/03/23].
[27] Preen DB, Holman CD, Spilsbury K, Semmens JB, Brameld KJ. Length of comorbidity lookback period affected regression model performance of administrative health data. J Clin Epidemiol 2006;59(9):940–6 [Epub 2006/08/10].
[28] Macmillan. Cancer Alliances: a crucial first step. Available from: www.macmillan.org.uk/documents/campaigns/canceralliancesreport.pdf; 2015.
[29] Enger SM, Thwin SS, Buist DS, Field T, Frost F, Geiger AM, et al. Breast cancer treatment of older women in integrated health care settings. J Clin Oncol 2006;24(27):4377–83 [Epub 2006/09/20].
[30] Bates T, Evans T, Lagord C, Monypenny I, Kearins O, Lawrence G. A population based study of variations in operation rates for breast cancer, of comorbidity and prognosis at diagnosis: failure to operate for early breast cancer in older women. Eur J Surg Oncol 2014;40(10):1230–6 [Epub 2014/08/02].
[31] Ward SE, Holmes GR, Ring A, Richards PD, Morgan JL, Broggio JW, et al. Adjuvant chemotherapy for breast cancer in older women: an analysis of retrospective English cancer registration data. Clin Oncol (R Coll Radiol) 2019;31(7):444–52 [Epub 2019/05/28].
[32] Lavelle K, Downing A, Thomas J, Lawrence G, Forman D, Oliver SE. Are lower rates of surgery amongst older women with breast cancer in the UK explained by co-morbidity? Br J Cancer 2012;107(7):1175–80 [Epub 2012/08/11].
[33] Mandelblatt JS, Sheppard VB, Hurria A, Kimmick G, Isaacs C, Taylor KL, et al. Breast cancer adjuvant chemotherapy decisions in older women: the role of patient preference and interactions with physicians. J Clin Oncol 2010;28(19):3146–53 [Epub 2010/06/03].
[34] Handforth C, Clegg A, Young C, Simpkins S, Seymour MT, Selby PJ, et al. The prevalence and outcomes of frailty in older cancer patients: a systematic review. Ann Oncol 2015;26(6):1091–101 [Epub 2014/11/19].
[35] Brain E, Caillet P, de Glas N, Biganzoli L, Cheng K, Lago LD, et al. HER2-targeted treatment for older patients with breast cancer: an expert position paper from the International Society of Geriatric Oncology. J Geriatr Oncol 2019 Nov;10(6):1003–13. https://doi.org/10.1016/j.jgo.2019.06.004 [Epub 2019/06/22].

5.3. Research Paper 5

Title

Treatment-related acute toxicity with adjuvant systemic treatment among patients with HER2-positive early invasive breast cancer: A national population-based cohort study.

Journal details

Published in BMJ Oncology. The online PDF version can be accessed at:
https://bmjoncology.bmj.com/content/2/1/e000081

Supplementary material can be found in the Appendices.

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed <u>for each</u> research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	BMJ Oncology		
When was the work published?	4 September 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**No**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	

Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	
Date	

Supervisor Signature	
Date	

BMJ Oncology

Treatment-related acute toxicity with adjuvant systemic treatment among patients with HER2-positive early invasive breast cancer: a national population-based cohort study

Melissa Ruth Gannon ,[1,2] David Dodwell,[3] Katie Miller,[1,2] Jibby Medina,[1,2] Karen Clements,[4] Kieran Horgan,[5] Min Hae Park,[1,2] David Alan Cromwell[1,2]

To cite: Gannon MR, Dodwell D, Miller K, *et al*. Treatment-related acute toxicity with adjuvant systemic treatment among patients with HER2-positive early invasive breast cancer: a national population-based cohort study. *BMJ Oncology* 2023;**2**:e000081. doi:10.1136/bmjonc-2023-000081

Received 25 April 2023
Accepted 16 August 2023

[1]Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK
[2]Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK
[3]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[4]National Cancer Registration and Analysis Service, NHS England, Birmingham, UK
[5]Department of Breast Surgery, St James's University Hospital, Leeds, UK

Correspondence to
Mrs Melissa R Gannon;
melissa.gannon@lshtm.ac.uk

ABSTRACT

Objective Although adjuvant trastuzumab-based treatment (TBT) improves survival for patients with HER2-positive early invasive breast cancer (EIBC), risk of toxicity grows as patient age increases. We examined use of TBT and associated severe acute toxicity event (SATE) rates to understand the real-world impact.

Methods and analysis Women (50+ years), newly diagnosed with HER2-positive EIBC in England, 2014–2019, were identified from Cancer Registry data, linked to the Systemic Anti-Cancer Therapy dataset for TBT information. SATEs were measured using hospital administrative data. Statistical models were developed to identify potential predictors of SATE.

Results Among 5087 women who received trastuzumab, with median duration 11.7 months, 47.4% (95% CI 46.0% to 48.7%) completed treatment. Women aged 70+ years made up 20.2% of patients aged 50+ who received adjuvant TBT in routine care, compared with 5% of women aged 50+ across trials. 32.8% (95% CI 31.5% to 34.1%) had a record of any SATE. 6.8% (95% CI 6.1% to 7.5%) had a cardiovascular SATE. Congestive cardiac failure rate was 0.5% (95% CI 0.3% to 0.7%). High deprivation, anthracycline use, increasing frailty were associated with increased odds of any SATE. Older age, sequential chemotherapy, history of myocardial infarction/chronic pulmonary disorder/liver disease were associated with increased odds of cardiovascular SATE. Among two-thirds of women not eligible for trial cohorts SATE rates were lower than for trial-eligible patients, explained by baseline differences in patients.

Conclusion Evidence of treatment-related SATE among patients treated in routine care is needed to inform treatment decisions and counsel older patients. This study provides information on SATE rates for adjuvant TBT, and common types, overall and by age for such discussions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Adjuvant trastuzumab-based treatment, given as a targeted therapy for HER2-positve early invasive breast cancer (EIBC), is associated with increased cardiotoxicity risk.

WHAT THIS STUDY ADDS

⇒ This study is the first comprehensive presentation of severe acute toxicity event (SATE) rates with adjuvant trastuzumab-based treatment for HER2-positive EIBC, with comparisons across patient subgroups defined by age and patient fitness. Data for 5087 women diagnosed from 2014 to 2019 highlighted one-third had a record of any SATE, while around 1 in 15 had a record of a cardiovascular SATE with increased odds as patient age increased.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Current evidence of treatment-related toxicity in older patients treated in routine care is vital in informing clinical practice by providing reference of average SATE rates, broken down by patient subgroups including age. This information is valuable for discussions around treatment decisions and helping counsel older patients on the side effects of treatment.

INTRODUCTION

Patients with HER2-positive breast cancer had a poor prognosis until the development of effective HER2-directed therapy, shown in randomised controlled trials (RCTs) to improve survival.[1] Guidelines recommend adjuvant trastuzumab in combination with surgery, chemotherapy and radiotherapy, for patients with HER2-positive early invasive breast cancer (EIBC).[2–5]

While improving survival, adjuvant trastuzumab is associated with increased cardiotoxicity risk, particularly among older patients.[6–8] Older age was reported to be associated with an increased frequency of cardiac events, notably congestive cardiac failure (CCF) but trial evidence has been limited by the under-representation of older patients.[7–9]

Patients aged 70+ years accounted for an estimated 2.5% of all patients in RCTs evaluating adjuvant trastuzumab, although at least one-third of patients diagnosed annually with BC are 70 or older. Observational studies describing toxicity, including cardiotoxicity, among patients who received trastuzumab-based treatment in the USA, Canada and Europe, have been in cohorts of patients diagnosed prior to 2010, or provided limited detail about specific toxicity events by age.[10–23] Few studies have examined the safety of trastuzumab-based treatment given to older women in routine care, and there is a gap in our understanding of the treatment-related toxicity for this population. This information is valuable in informing treatment decisions and counselling older patients on the side effects of treatment.

This study aimed to characterise the cohort of women (50+ years) with HER2-positive EIBC who received adjuvant trastuzumab-based treatment in routine care in England, and investigate treatment-related severe acute toxicity events (SATEs). The study is reported according to the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) extension to the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines.[24] Although SATEs are only one aspect of toxicity burden, they are of importance to understand in the management of patients as they are often the cause of early treatment discontinuation which is associated with poorer outcomes.[25]

MATERIALS AND METHODS

Data source

This population-based cohort study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP; full details can be accessed via www.nabcop.org.uk). NABCOP received pseudonymised Cancer Registry data for all women aged 50+ years, with a new BC diagnosis between 1 January 2014 and 31 December 2019, and treated within an English National Health Service (NHS) trust.[26] Data were linked at tumour level to data from Cancer Outcomes and Services Dataset (COSD), Hospital Episode Statistics Admitted Patient Care (HES-APC),[27] Systemic Anti-Cancer Therapy dataset (SACT),[28] national Radiotherapy Dataset (RTDS) and at patient level to the Primary Care Prescription Database (PCPD).[29]

Study population

We identified women diagnosed with HER2-positive EIBC (stages 1–3A; International Classification of Diseases, 10th revision [ICD-10] diagnosis code C50), who received surgery within 6 months of diagnosis and had a record of receiving adjuvant trastuzumab-based treatment (commenced within 4 months of primary surgery with no prior trastuzumab) in SACT. BC was classified as HER2-positive where HER2 status was reported as either positive or borderline with a positive HER2-FISH (fluorescence in situ hybridisation) or equivalent molecular test result. To aid interpretation of SATE rates among these women, we defined a comparison group of women with HER2-negative EIBC, who had surgery (with no prior chemotherapy) within 6 months of diagnosis and adjuvant chemotherapy (commenced within 4 months of surgery) recorded in SACT.

Only patients with a recorded or calculated treatment window end date prior to 1 April 2021 and complete data on patient fitness, tumour stage, nodal stage and invasive grade were included.

Study outcome

Episodes of treatment-related SATE were identified from ICD-10 diagnosis codes recorded in diagnosis fields associated with an overnight hospital admission within hospital administrative data (HES-APC) using a coding framework previously validated in patients with colon cancer (online supplemental table A1).[30] The study limited the period during which an episode of SATE could occur to the time from the start of adjuvant treatment up to 8 weeks (56 days) after the last reported cycle. The definition of SATE covered a wide range of possible events, including haematological disorders, infection, cardiovascular disorders, neutropenic sepsis and gastrointestinal disorder. We also looked specifically at cardiovascular SATEs.

For the comparison group of patients with HER2-negative EIBC, the measure of SATEs included events captured in overnight admissions that occurred from the first reported cycle of chemotherapy up to 536 days later, which corresponded to the 90th centile of the duration of trastuzumab-based treatment for HER2-positive EIBC (online supplemental figure A1). Follow-up in HES-APC data was available up to 31 March 2021 for the HER2-positive cohort and comparison group. Deaths related to admissions were defined as those events with death recorded as the discharge method in HES-APC.

Study variables

Information on trastuzumab-based treatment was extracted from SACT. Trastuzumab was identified from records which contained either trastuzumab, Herceptin or trastuzumab biosimilar in the drug name field. Trastuzumab treatment in which the first recorded cycle was after surgery, and either within 4 months of the surgery date or following chemotherapy, was defined as adjuvant. Trastuzumab frequency was defined based on time between consecutive cycles. As the recommended interval between cycles is 3 weeks, cycles with a gap of 4–6 weeks since the previous cycle were labelled ‘delayed’, while a gap of 7 weeks up to 4 months between consecutive cycles was defined as a ‘treatment break’. Cycles given after a gap of more than 4 months were considered to be a different treatment episode. Cycles with a dose of zero recorded, with no associated hospital admission for treatment, were not counted. All sequential cycles were counted, and a complete course of treatment was defined as 18 cycles, or 17 cycles with a duration of at least 51 weeks. Treatment was defined as discontinued where less

than 16 cycles were recorded. Administration route was categorised as subcutaneous where at least one cycle had this route recorded. Chemotherapy was categorised as: 'sequential' where cycles were administered prior to the first trastuzumab date, with no cycles delivered during the trastuzumab cycles (online supplemental figure A2); and 'concurrent' where any cycles were administered either on the same day as trastuzumab or between trastuzumab cycles (including chemotherapy started prior to and continued during trastuzumab). Trastuzumab and chemotherapy recorded in HES-APC was included as a sensitivity analysis to evaluate the impact of treatment not recorded in SACT, as described previously.[31 32]

Primary surgery was defined as either breast-conserving surgery (BCS) or mastectomy which occurred within 6 months of diagnosis, identified from Office of Population Censuses and Surveys procedure codes within HES-APC records. Radiotherapy use was identified based on records within RTDS that occurred during the initial treatment episode following diagnosis. Endocrine therapy use was identified from the PCPD.[29]

Data on baseline patient and tumour characteristics were taken from Cancer Registry/COSD. Hormone receptor (HR)-positive breast cancer was defined where either ER or PR status were recorded as positive. Deprivation was measured using the Index of Multiple Deprivation 2019 rank, derived from the patient's postcode at diagnosis and assigned to national quintiles of deprivation (most (1) to least (5) deprived). ICD-10 codes recorded in HES-APC within 2 years prior to diagnosis were used to determine comorbidity burden (0, 1, 2+; defined using the Royal College of Surgeons of England Charlson Comorbidity Index—CCI) and frailty (fit, mild frailty, moderate-to-severe frailty; defined using the Secondary Care Administrative Records Frailty—SCARF index).[33]

Patients with HER2-positive EIBC were categorised as 'trial eligible' (yes/no) where application of the inclusion/exclusion criteria and recruited age range of RCTs conducted in the adjuvant setting, using the routine data available, flagged them as being 'eligible' for at least one trial (online supplemental table A2).

Statistical analysis

All data preparation and statistical analyses were conducted using Stata V.17.0.

Patient and tumour characteristics of women receiving trastuzumab-based treatment are presented for the overall cohort, as well as by age.

SATE rates are presented for the overall cohort as well as by age, comorbidity burden, sequential/concurrent chemotherapy use, and HER2 status (HER2-positive/HER2-negative). Kaplan-Meier survival curves were used to visually assess differences in time to first SATE from start of adjuvant treatment/trastuzumab cycle. Multilevel mixed-effects logistic regression models were used to identify factors associated with odds of (cardiovascular) SATE and odds of treatment discontinuation. Models included baseline variables (measured at diagnosis) of age (50–59/60–69/70–79/80+), patient fitness (CCI, SCARF index), deprivation and comorbidity flags for history of myocardial infarction (MI), CCF, chronic pulmonary disease (CPD), liver disorder, renal disorder or diabetes, based on those created to calculate the CCI (online supplemental table A3). Additionally, chemotherapy details (sequential use and anthracyclines use) were included as predictors of SATE.

Further comparison of cohort characteristics and SATE rates was done across patient subgroups defined as 'trial eligible'.

RESULTS

Among 156 375 women, aged 50 years and over, diagnosed with EIBC in England between 1 January 2014 and 31 December 2019, 9.6% (n=14 936) had HER2-positive tumours. There were 11 584 of these 14 936 patients who had surgery within 6 months of diagnosis, of which 45.0% (n=5215) went on to receive adjuvant trastuzumab-based treatment (not in combination with other HER2-targetting drugs) with no trastuzumab prior to surgery (online supplemental figure A3). A total of 5087 women had a treatment window ending prior to 1 April 2021 and information on patient fitness and tumour characteristics, and constituted the analysed study cohort. There were 13 577 women diagnosed with HER2-negative tumours over the same time period who received adjuvant chemotherapy with a calculated treatment window ending prior to 1 April 2021. Median follow-up of both groups was 61.1 months (IQR: 44.4–76.1).

Cohort characteristics

Table 1 provides patient and tumour characteristics plus treatment details for the cohort of women receiving adjuvant trastuzumab-based treatment for HER2-positive EIBC. The majority (79.8%) of the cohort was aged 50–69 years, had no comorbidities (90.6%) and were fit (85.7%). Two-thirds had hormone receptor-positive cancers. The distribution of tumour characteristics differed by age, with a greater proportion of older women having higher grade disease, larger tumours and nodal involvement. Comorbidities or some level of frailty were also more prevalent in older age groups. Rates of BCS and radiotherapy were lower for women aged 80+ years. While nearly all women were recorded as receiving chemotherapy, the regimen type differed by age, with older women more likely to receive a taxane-only (paclitaxel use increased as age increased, while docetaxel use decreased).

Based on reported eligibility criteria for trials of adjuvant trastuzumab, 35.6% of the study cohort were categorised as trial eligible. Comparison with reported characteristics of patients within RCTs highlighted that women treated in routine practice include a higher per cent of older women (5% of women aged 50+ in the RCTs were aged 70+ years, compared with 20.2% in this cohort), higher percentages with HR-positive, smaller,

Table 1 Distribution of patient, tumour and treatment characteristics in women with HER2-positive, early invasive breast cancer, diagnosed in NHS trusts in England between January 2014 and December 2019 and receiving adjuvant trastuzumab-based treatment, overall and by age at diagnosis

		Total	50–59 years	60–69 years	70+ years
Total		**5087**	**2153**	**1906**	**1028**
Age group	50–59 years	2153 (42.3%)			
	60–69 years	1906 (37.5%)			
	70–79 years	941 (18.5%)			
	80+ years	87 (1.7%)			
IMD 2019	1—Most deprived	15.2%	17.5%	13.9%	12.8%
	4	16.7%	17.6%	16.4%	15.7%
	2	21.4%	20.9%	22.3%	20.7%
	3	22.8%	21.6%	23.6%	24.2%
	5—Least deprived	23.8%	22.5%	23.9%	26.6%
Charlson Comorbidity Index	0	90.6%	93.6%	89.2%	86.9%
	1	7.5%	5.3%	8.4%	10.2%
	2+	1.9%	1.0%	2.4%	2.9%
SCARF index	Fit	85.7%	90.8%	84.2%	77.7%
	Mild frailty	9.3%	6.8%	10.1%	13.1%
	Moderate-to-severe frailty	5.0%	2.5%	5.7%	9.1%
Stage grouping	1	46.5%	49.4%	51.7%	30.8%
	2	46.4%	44.3%	43.1%	57.0%
	3A	7.1%	6.3%	5.2%	12.2%
Grade of disease	G1	1.9%	2.6%	1.7%	1.0%
	G2	36.7%	36.9%	39.7%	30.7%
	G3	61.4%	60.5%	58.6%	68.3%
Tumour stage	T1	56.3%	59.7%	60.7%	41.0%
	T2	40.4%	37.3%	37.1%	52.9%
	T3	3.3%	2.9%	2.2%	6.1%
Nodal stage	N0	68.8%	68.8%	73.7%	59.7%
	N1	25.4%	26.1%	21.8%	30.5%
	N2	5.8%	5.1%	4.6%	9.7%
Positive hormone-receptor status	Yes	68.9%	72.5%	68.2%	62.5%
	No/unknown	31.1%	27.5%	31.8%	37.5%
Surgery type	BCS	68.4%	70.5%	72.1%	57.0%
	Mastectomy	31.6%	29.5%	27.9%	43.0%
Radiotherapy reported/ setting	No	20.7%	20.3%	19.1%	24.7%
	Yes—before trastuzumab	5.9%	5.2%	7.0%	5.5%
	Yes—during trastuzumab	58.1%	60.3%	59.2%	51.8%
	Yes—after trastuzumab	15.2%	14.3%	14.7%	18.0%
Hormone therapy prescribed (Y)		69.0%	72.4%	68.7%	62.4%
Chemotherapy reported (Y)		98.6%	98.9%	98.8%	97.5%
	Anthracycline* only (Y)	18.2%	16.1%	21.4%	16.4%

Continued

Table 1 Continued

		Total	50–59 years	60–69 years	70+ years
Total		**5087**	**2153**	**1906**	**1028**
	Taxane* only (Y)	41.3%	31.7%	40.7%	62.8%
	Anthracycline and taxane (Y)	39.6%	51.5%	36.8%	19.8%
	Anthracycline prior, taxane concurrent (Y)	37.2%	48.7%	34.9%	17.4%
	Docetaxel (Y)	49.5%	61.6%	45.6%	31.2%
	Paclitaxel (Y)	32.7%	22.9%	33.4%	52.1%
	Sequential chemo (Y)	21.9%	20.0%	24.8%	20.5%
	Concurrent chemo (Y)	78.1%	80.0%	75.2%	79.5%
	Concurrent anthracycline (% of concurrent chemo)	0.6%	0.8%	0.4%	0.5%
	Concurrent taxane (% of concurrent chemo)	98.9%	99.3%	98.8%	98.3%
Trastuzumab discontinued (Y)		34.4%	32.9%	34.6%	37.1%

*Anthracyclines = doxorubicin, epirubicin, mitoxantrone recorded in SACT. Taxanes = docetaxel, cabazitaxel, paclitaxel, nab-paclitaxel recorded in SACT.
BCS, breast-conserving surgery; IMD, Index of Multiple Deprivation; NHS, National Health Service; SACT, Systemic Anti-Cancer Therapy dataset; SCARF, Secondary Care Administrative Records Frailty.

node-negative tumours and undergoing BCS in contrast to mastectomy.

The cohort diagnosed with HER2-negative EIBC who received adjuvant chemotherapy had a similar distribution across most characteristics with small differences in relation to age, stage grouping, invasive grade, tumour stage, nodal involvement and use of radiotherapy (online supplemental table A4). The groups differed in the type of chemotherapy used, with more use of anthracyclines (with or without taxanes) and less use of paclitaxel in the comparison group of women with HER2-negative EIBC.

Trastuzumab treatment details

Median duration of trastuzumab (time from first to last recorded cycle) was 11.7 months (IQR=10.4–12.0), 54.1% of patients had 17–18 cycles and 95.8% (n=4873/5087) had a record of trastuzumab ever being administered subcutaneously. In patients who had more than one cycle of trastuzumab, the typical interval was 3 weekly for 96.1% (n=4802/4998), with only 64 patients (1.3%) having weekly treatment (online supplemental figure A4).

Of 5015 patients with a record of adjuvant chemotherapy, 21.9% (n=1099) commenced trastuzumab after chemotherapy (sequential), while 78.1% (n=3916) received concurrent chemotherapy, of which nearly all (98.9%) included a taxane (table 1). Concurrent taxane use was high across all age groups. Trastuzumab was started after an anthracycline and given concurrently with a taxane for 37.2% (n=1867), with highest use among younger women. Across all age groups, <1% received an anthracycline concurrently.

Severe acute toxicity events

Among women who received adjuvant trastuzumab-based treatment, 32.8% (n=1670; 95% CI 31.5% to 34.1%) had at least one SATE (table 2). The percentage of women with a cardiovascular SATE captured was 6.8% (n=346; 95% CI 6.1% to 7.5%). Rates of CCF captured were low at 0.5% (95% CI 0.3% to 0.7%). Fifteen (0.3%) women were reported to have died during an admission with a SATE.

Admissions could contain several conditions related to the SATE. Treating these separately, the most common SATEs, captured among more than 10% of patients, were haematological disorder (15.2%; most commonly neutropenia), infection (15.2%), neutropenic sepsis (13.0%), gastrointestinal disorder (10.0%).

The odds of SATE were greatest during the first 18 weeks of treatment (figure 1). Although there was no difference in SATE rates over the full treatment course according to receipt of sequential versus concurrent chemotherapy (online supplemental table A5), SATEs recorded in overnight admissions after trastuzumab started were higher where patients received concurrent rather than sequential chemotherapy (figure 2). Nearly all concurrent chemotherapy involved a taxane; anthracycline regimens were typically given sequentially. Figure 2 highlights that SATE among patients receiving sequential chemotherapy were lower during trastuzumab suggesting the majority were experienced during chemotherapy rather than trastuzumab. SATE rates did not differ according to whether patients received the full course of trastuzumab treatment or not (33.4% vs 32.3%).

Table 2 Frequency of severe acute toxicity events (SATE), overall/by individual SATE type, among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer, by age at diagnosis (ordered by most frequently recorded; only individual SATE with >5% presented)

	Total N=5087		50–59 years N=2153		60–69 years N=1906		70–79 years N=941		80+ years N=87	
Event	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Any	1670	32.8	710	33.0	625	32.8	312	33.2	23	26.4
Haematological	774	15.2	325	15.1	304	15.9	142	15.1	3	3.4
Neutropenia	714	14.0	302	14.0	287	15.1	124	13.2	1	1.1
Anaemia	103	2.0	33	1.5	38	2.0	30	3.2	2	2.3
Thrombocytopaenia	15	0.3	6	0.3	7	0.4	2	0.2	0	0.0
Infection	773	15.2	333	15.5	289	15.2	140	14.9	11	12.6
Neutropenic sepsis	659	13.0	279	13.0	267	14.0	112	11.9	1	1.1
Gastrointestinal	507	10.0	193	9.0	199	10.4	107	11.4	8	9.2
Cardiovascular	346	6.8	132	6.1	125	6.6	83	8.8	6	6.9
Arrhythmia	106	2.1	43	2.0	36	1.9	23	2.4	4	4.6
Hypertension	73	1.4	26	1.2	29	1.5	18	1.9	0	0.0
Angina	7	0.1	2	0.1	3	0.2	2	0.2	0	0.0
Congestive cardiac failure	25	0.5	2	0.1	12	0.6	9	1.0	2	2.3
Cerebrovascular	14	0.3	6	0.3	3	0.2	4	0.4	1	1.1
Other	175	3.4	65	3.0	72	3.8	37	3.9	1	1.1

Comparison with the group of women who received chemotherapy for HER2-negative EIBC, overall SATE rates captured were higher for HER2-negative EIBC (37.7%, 95% CI 36.9% to 38.5%; Online supplemental figure A5 and table A6), and cardiovascular SATE rates were also higher (8.5%, 95% CI 8.1% to 9.0%). There were no evidence rates differed after accounting for baseline differences in the two groups. A similar pattern from the first treatment cycle to the first SATE was observed for both groups (online supplemental figure A5).

Overall SATE rates were similar for patients of different ages (online supplemental figure A6). Rates among

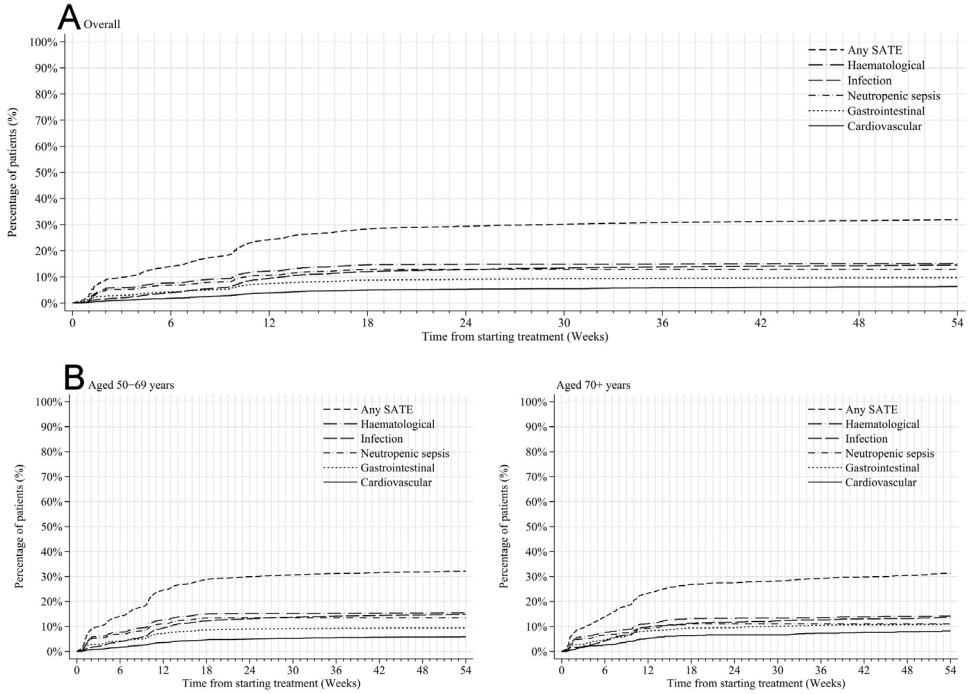


Figure 1 Time from first treatment cycle to first severe acute toxicity event, any and most frequently reported individual toxicity group, among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer, (A) overall and (B) by age at diagnosis. SATE, severe acute toxicity event.

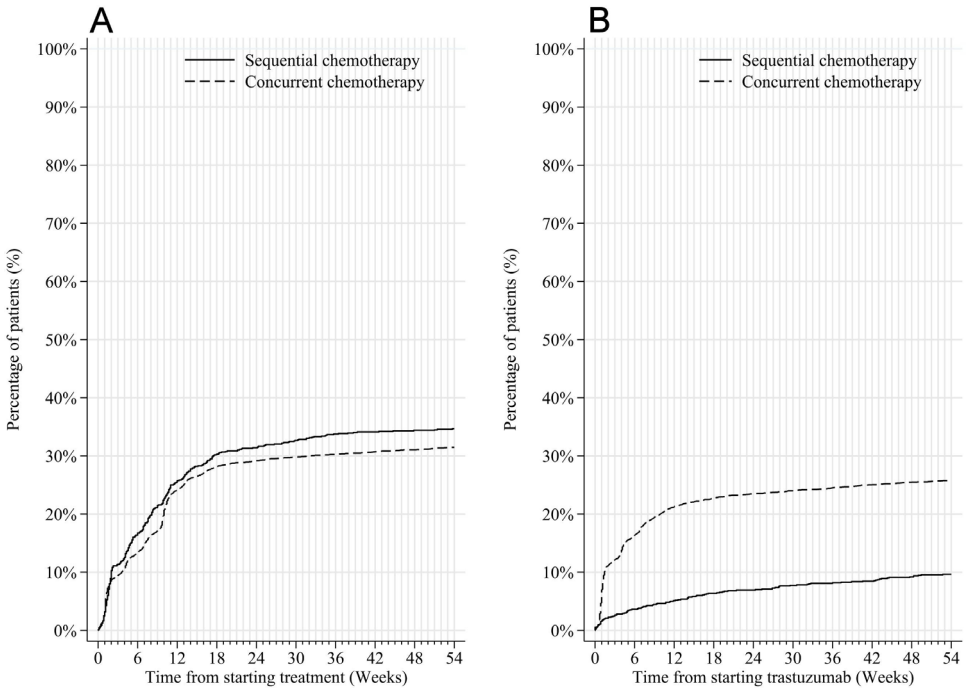


Figure 2 Time to severe acute toxicity event from (A) first treatment cycle and (B) first trastuzumab cycle, among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer, by sequential or concurrent use of chemotherapy.

women aged 80+ were 26.4% (95% CI 17.6% to 37.0%; table 2). SATE rates increased with a greater comorbidity burden, particularly in relation to the SATE due to infection, cardiovascular disorders and gastrointestinal disorders (online supplemental table A7). The rates of cardiovascular disorders rose with increasing age.

High deprivation, use of anthracyclines and greater frailty were associated with increased odds of any SATE, even after accounting for each other and other factors (online supplemental table A8). Older age, history of MI, CPD and liver disease, as well as having sequential chemotherapy, were associated with increased odds of cardiovascular SATE (online supplemental table A8).

Relationship between SATE and treatment discontinuation/delay

Of all women in the cohort, 47.4% (95% CI 46.0% to 48.7%; n=2409/5087) received the complete course of trastuzumab treatment (18 cycles or 17 cycles over at least 51 weeks). A further 18.3% (n=930) had 16 cycles or 17 cycles with less than 51 weeks duration. The remaining 34.4% (n=1748/5087) of women were defined as having discontinued treatment (online supplemental figure A4; table 1). Discontinuation was higher among those receiving sequential chemotherapy (figure 3; 41.2% compared with 31.8% among women receiving concurrent chemotherapy). Additionally, odds of discontinuation were higher among women aged 80+, those with a history of CCF and those who had a treatment break, while having a delay was associated with decreased odds.

Of 4998 patients who received more than one cycle, 69.9% (n=3494) had at least one delay/break between cycles; 63.5% of patients had at least one cycle delayed while 21.1% had at least one break between trastuzumab cycles. Although there was no evidence of a difference by age, the percentage who either discontinued treatment following a SATE (figure 3) or had a delay/break before the next cycle increased with age (online supplemental table A9). Around two-thirds of patients had no SATE captured during trastuzumab cycles but still had a delay/break in between cycles (online supplemental table A9).

Estimation of SATE rates when the cohort was limited to a 'trial-eligible population'

Applying the RCT eligibility criteria to our cohort receiving adjuvant trastuzumab-based treatment, 35.6% were defined as 'trial eligible'. Comparisons of characteristics found lower percentages of trial eligible patients among patient groups of moderate-to-severe frailty, with greater comorbidity burden, grade 1 disease, smaller, node-negative tumours, among women having BCS and by chemotherapy type (online supplemental table A10).

Higher rates of overall SATE were seen among those defined as 'trial eligible' compared with those not eligible (36.8% vs 30.6%, p<0.001). This difference was greatest among the group of patients who received concurrent chemotherapy (38.9% vs 30.0%, p<0.001). There were no evidence rates differed after accounting for baseline differences in the two groups.

DISCUSSION

This study examined the characteristics and treatment safety profile of 5087 women (50+ years) with

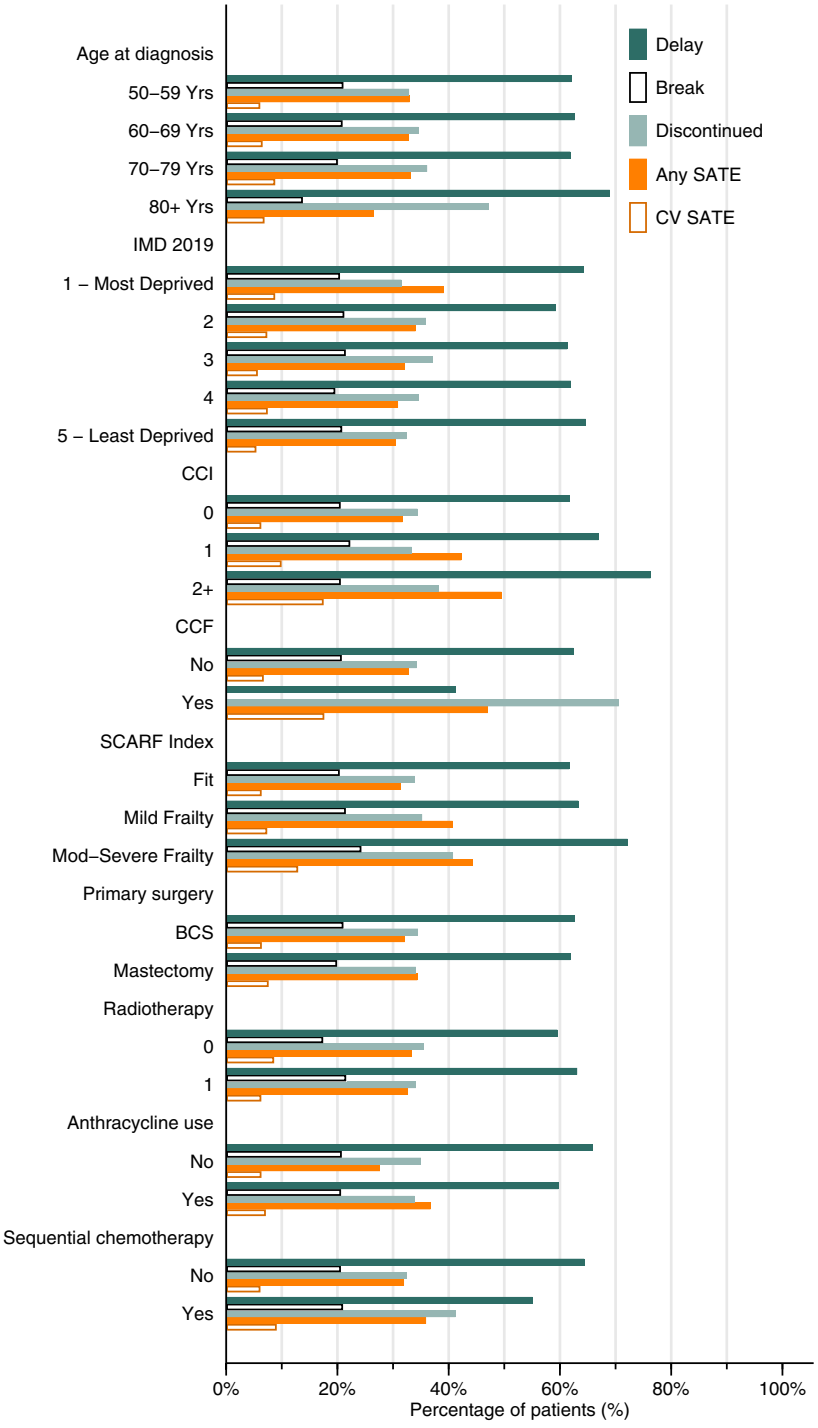


Figure 3 Trastuzumab delays, breaks, discontinuation and SATE rates, among women with HER2-positive, early invasive breast cancer, diagnosed in NHS trusts in England between January 2014 and December 2019 and receiving adjuvant trastuzumab-based treatment, by cohort characteristics. BCS, breast-conserving surgery; CCF, congestive cardiac failure; CCI, Charlson Comorbidity Index; CV, Cardiovascular; IMD, Index of Multiple Deprivation; NHS, National Health Service; SATE, severe acute toxicity event.

HER2-positive EIBC who received adjuvant trastuzumab-based treatment. Median trastuzumab duration was 11.7 months, in line with the expected duration of 12 months. One in three women (32.8%) had at least one SATE captured within an overnight hospital admission throughout the course of their adjuvant trastuzumab-based treatment. SATEs were more likely among women with frailty and those receiving anthracyclines. One in 15 women (6.8%) had a cardiovascular SATE captured, with increased rates among older women, and those

with a history of cardiovascular or liver disease. Comparison with the group of women receiving chemotherapy for HER2-negative EIBC highlighted comparable SATE rates, after accounting for differences in the cohorts, suggesting trastuzumab did not add toxicity beyond that experienced from chemotherapy.

Comparison with RCTs highlighted differences in cohort characteristics, with women (50+ years) with HER2-positive EIBC treated in routine care, nearly all receiving 3-weekly trastuzumab, including a higher percentage of older women (20.2% aged 70+, compared with 5% of patients aged 50+in the RCTs) and small, node-negative tumours.

Reviews of the four major adjuvant RCTs highlighted rates of CCF varied from 0.6% to 3.3%, with overall cardiovascular toxicity ranging from 5.7%–18.0% and higher odds as age increased.[34 35] Rates of CCF SATE estimated for our study cohort were 0.5%, with overall cardiovascular SATE rates of 6.8%.

Estimates of trastuzumab-based treatment use in this older population are in line with other studies in this setting.[22 23] Nearly all had trastuzumab administered subcutaneously, in line with more recent advances which demonstrated clinical comparability with this delivery route.[36 37] Women aged 80+ were more likely to discontinue treatment, as reported in a previous population-based study in the USA.[38] Cardiotoxicity and CCF rates reported in other population-based studies provide a mixed picture, with some studies reporting very low rates from 2.6% to 8.5%[14–17] while others report higher rates up to 29.4%.[13 18 39] This variation is likely to be due to differences in patient cohorts and data collection methods. Cardiotoxicity captured in hospital admissions data were lower than those reported across several studies, while CCF rates were similarly low: a Dutch study in a hospital setting, which defined cardiotoxicity using the same definition as the HERA trial reported 12.6% cardiotoxicity; a US study in older patients (66+ years) reported 1.2% admissions for CCF; the OHERA study reported 1.0% severe CCF and 17.5% cardiac events; a meta-analysis of studies reported 12% overall cardiotoxicity incidence; and a study across three NHS trusts reported 15.7% cardiotoxicity rates during treatment.[10 12 38 40 41] Several studies have reported age differences in cardiotoxicity, with risk increasing with age.[12–16] Additionally, sequential therapy use has been described in another population-based study to be associated with increased odds of cardiovascular SATEs.[42] Prevalent hypertension captured in the data was low; however, recent publications highlight the risk of chemotherapy-induced hypertension.[43 44]

The study has a number of strengths. It used a large, population-based sample, which included women diagnosed over a period of 6 years (2014–2019) with HES-APC data to 31 March 2021 and so reflects current treatment practice. All data used in the study were linked at patient or tumour level, so all estimates of treatment characteristics and SATE are for the same patients. The cohort in this study included more patients aged 70+ years than were included in the RCTs. The methodology used to identify systemic treatment-related SATEs incorporates toxicity recorded in all diagnosis fields within hospital administrative data and will, therefore, document both those severe events causing an overnight admission as well as symptomatic events experienced and flagged during an overnight admission but which were not the cause for the admission. Pre-existing comorbidities were also accounted for to avoid misclassification of chronic conditions as toxicity, providing more certainty that SATEs were treatment related. Finally, the dataset contained longitudinal treatment information, along with sufficient patient and tumour characteristics associated with treatment decisions to provide robust comparison with trial populations.

There are various limitations of this study. First, as estimation of treatment-related toxicity is based on an overnight stay captured in hospital admissions data, treatment-related toxicity which is either purely symptomatic/less severe or identified via purposeful clinical observation and does not result in an overnight NHS hospital admission, or is managed in a non-NHS setting will not be counted. For this study, many of the toxicities captured will, therefore, predominantly relate to more severe events, the type of toxicity a trial would classify as a serious adverse event. This will result in an under estimate of the true toxicity burden of treatment and as such whether cardiac toxicity rates are different to those observed in the RCTs, which use a different method of measurement to that applied in this study, is unclear. Furthermore, it was not possible within the routine data to know whether the SATE was considered to be a reaction to treatment (serious adverse reaction), something which could be recorded by clinicians within an RCT. Additionally, time to SATE will be an overestimate where the SATE is captured within an admission record but was not the reason for admission, as this will have developed prior to the admission. Second, the time frame for chemotherapy is typically substantially shorter than for trastuzumab-based treatment, therefore, a time window for counting treatment-related admissions among women receiving chemotherapy for HER2-negative EIBC may include admissions relating to any treatments given beyond the initial chemotherapy. Third, as this study used routine data, which were not created or collected to answer this specific research question, there may be issues such as misclassification bias, unmeasured confounding and missing data. As SACT provides data on prescribed therapies, there may be a small number of patients included within the study cohort who were prescribed trastuzumab but for whom it was never administered. Additionally, estimates of treatment cycles/duration may be higher than in practice. As this study aimed to compare the patient cohort and SATE rates to those reported in the trials, where information was presented on an intention-to-treat basis so may also include patients randomised to treatment who never received it or who did not have all treatment cycles, the data provided is informative and

complementary. SACT also has various quality assurance processes carried out before data release, details of which can be found via: https://digital.nhs.uk/ndrs/data/datasets/sact. Another concern is the potential for errors in patient and tumour characteristics within the England Cancer Registry and COSD datasets. The cancer registration service has various validation steps when compiling the national registration data and the overall effect of coding errors should therefore be minimal. We note that missing data for this cohort were small; in addition, sensitivity analysis using HES data to identify further treatment information highlighted few patients were missed from the cohort. Finally, it was not possible to perform a comprehensive comparison of the patients treated in routine practice with those recruited to the RCTs. In part, this was due to the limited reporting of baseline characteristics in the RCT publications and because many of the baseline function tests required prior to enrolment are either not routinely done outside of a trial setting or the details are unavailable within routine national data. As such the estimated number of patients fulfilling trial eligibility is likely to differ from reality. Additionally, this study did not have information on whether any patients in the study cohort were participating in RCTs, which might contribute to SATE rates being under-estimated and to some patients not being recorded as receiving 12 months of treatment.[45]

This study estimated SATE rates for patients receiving adjuvant trastuzumab-based treatment for HER2-positive EIBC in routine care. Overall rates were comparable by age, suggesting patients were well monitored, with an increase in cardiotoxicity as age increased, most likely related to an increased susceptibility to cardiovascular problems due to reduced physiological reserve. Few studies have reported overall SATEs, or individual SATEs beyond cardiotoxicity. Reporting the full safety profile of trastuzumab-based treatment is important in understanding the impact of treatment in routine care, and acknowledging that chemotherapy is part of the treatment provided. Detail of SATE is also valuable in providing information for treatment discussions between clinicians and patients. We found that SATEs were higher among women receiving anthracyclines as part of their chemotherapy treatment, which were typically given sequentially. The majority of this cohort treated in routine care received adjuvant chemotherapy, as is recommended practice, however, we note that more recent trials looking at use of trastuzumab monotherapy have suggested that it might be more appropriate for those patients who are more frail or where SATEs are a concern.[46 47] Future work looking at SATE for patients receiving trastuzumab monotherapy would be beneficial to understand the safety profile of this among patients treated in routine care. Additionally, frailty rather than increasing age was associated with increased SATEs, re-enforcing the message that age alone should not determine treatment decisions.[48]

Since trastuzumab was first approved for use in this setting in 2005, there have been several further HER2-targeting therapies licensed and approved and a move towards use of trastuzumab biosimilars.[49] Future research should characterise the cohort of patients receiving these newer treatments to understand the associated benefits and harms from their use in routine care.

In conclusion, this national cohort study found that among patients who received adjuvant trastuzumab-based treatment for HER2-positive EIBC in routine clinical practice, one-third had any SATE recorded, with frailty and use of anthracyclines associated with increased odds. Rates of cardiovascular SATE increased with increasing age and use of sequential therapy. CCF rates were low. The addition of trastuzumab to chemotherapy added little to major SATE experienced, suggesting that where chemotherapy is recommended for HER2-positive EIBC trastuzumab should also be recommended. Two-thirds of patients were estimated to not be represented in trial populations; lower SATE rates among such patients were explained by baseline differences in patients.

Contributors Guarantor of integrity of the entire study: MRG. Study concepts and design: MRG, DD, KM, KH, KC, JM, MHP and DAC. Literature research: MRG. Clinical studies: N/A. Experimental studies/data analysis: N/A. Data acquisition: MRG, KC and JM. Statistical analysis: MRG. Manuscript preparation: MRG. Manuscript editing: MRG, DD, KM, KH, KC, JM, MHP, DAC. All authors were involved in data interpretation, critical appraisal of the draft manuscript and gave final approval on the submitted version.

Funding This study was undertaken as part of the work by the National Audit of Breast Cancer in Older Patients (NABCOP). The audit is commissioned by the Healthcare Quality Improvement Partnership (HQIP) as part of the National Clinical Audit and Patient Outcomes Programme, and funded by NHS England and the Welsh Government (www.hqip.org.uk/national-programmes). Neither the commissioner nor the funders had any involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication. The authors had full independence from the HQIP. The aim of the NABCOP is to evaluate the care of older women with breast cancer in England and Wales, and support NHS providers to improve the quality of hospital care for these women. More information can be found at www.nabcop.org.uk. DD also receives funding from Cancer Research UK (grant C8225/A21133).

Disclaimer CRUK had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the article for publication.

Competing interests KH declares being the Chair of the Endonet Trial Steering Committee. KC declares funding received for Breast Cancer Research Manager role within NHS England as part of the Cancer Grand Challenges PRECISION team, which is funded by Cancer Research UK and the Dutch Cancer Society (C7827/A27366); the grant is not related to her work on the National Audit of Breast Cancer in Older Patients or this paper. DAC declares grants/contracts from Healthcare Quality Improvement partnership; participation on the Pregnancy Outcome Prediction Study (POPS2) Trial Steering Committee; being on the Editorial Committee for the Journal of Health Services Research and Policy; being Deputy Chair on the Examination Committee for the MSc, PG Diploma and PG Cert in Public Health distance learning programme at the London School of Hygiene & Tropical Medicine. There are no other relationships or activities that could appear to have influenced the submitted work.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study was exempt from NHS Research Ethics Committee approval because it involved analysis of pseudonymised linked data collated for the purpose of service evaluation as part of the National Audit of Breast Cancer in Older Patients.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. This work uses data that have been provided by patients and collected by the NHS as part of their care and support. The data for England are collated, maintained and quality assured by the National Disease Registration Service (NDRS), which is part of NHS England. Data on English Cancer Registrations can be accessed via the NHS Digital Data Access Request Service (DARS) https://digital.nhs.uk/services/data-access-request-service-dars#national-disease-registration-service-ndrs-.

ORCID iD

Melissa Ruth Gannon http://orcid.org/0000-0002-9163-0571

REFERENCES

1 Loibl S, Gianni L. Her2-positive breast cancer. *Lancet* 2017;389:2415–29.
2 Senkus E, Kyriakides S, Ohno S, *et al*. Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015;26 Suppl 5:v8–30.
3 NICE. Early and locally advanced breast cancer: diagnosis and Management.2018. n.d. Available: www.nice.org.uk/guidance/ng101
4 Biganzoli L, Wildiers H, Oakman C, *et al*. Management of elderly patients with breast cancer: updated recommendations of the International society of geriatric oncology (SIOG) and European society of breast cancer specialists (EUSOMA). *Lancet Oncol* 2012;13:e148–60.
5 Biganzoli L, Battisti NML, Wildiers H, *et al*. Updated recommendations regarding the management of older patients with breast cancer: a joint paper from the European society of breast cancer specialists (EUSOMA) and the International society of geriatric oncology (SIOG). *Lancet Oncol* 2021;22:e327–40.
6 Moja L, Tagliabue L, Balduzzi S, *et al*. Trastuzumab containing regimens for early breast cancer. The Cochrane database of systematic reviews. 2012:CD006243.
7 Bradley R, Braybrooke J, Gray R, *et al*. Early breast cancer trialists' collaborative G. Trastuzumab for early-stage, Her2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. *The Lancet Oncology* 2021;22:1139–50.
8 Brollo J, Curigliano G, Disalvatore D, *et al*. Adjuvant Trastuzumab in elderly with HER-2 positive breast cancer: a systematic review of randomized controlled trials. *Cancer Treat Rev* 2013;39:44–50.
9 Partridge AH, Gelber S, Piccart-Gebhart MJ, *et al*. Effect of age on breast cancer outcomes in women with human Epidermal growth factor receptor 2-positive breast cancer: results from a Herceptin adjuvant trial. *JCO* 2013;31:2692–8.
10 Jawa Z, Perez RM, Garlie L, *et al*. Risk factors of Trastuzumab-induced cardiotoxicity in breast cancer: A meta-analysis. *Medicine (Baltimore)* 2016;95:e5195.
11 Mazzotta M, Krasniqi E, Barchiesi G, *et al*. Long-term safety and real-world effectiveness of Trastuzumab in breast cancer. *J Clin Med* 2019;8:254.
12 Lidbrink E, Chmielowska E, Otremba B, *et al*. A real-world study of cardiac events in > 3700 patients with Her2-positive early breast cancer treated with Trastuzumab: final analysis of the OHERA study. *Breast Cancer Res Treat* 2019;174:187–96.
13 Chavez-MacGregor M, Zhang N, Buchholz TA, *et al*. Trastuzumab-related cardiotoxicity among older patients with breast cancer. *JCO* 2013;31:4222–8.
14 Bonifazi M, Franchi M, Rossi M, *et al*. Trastuzumab-related cardiotoxicity in early breast cancer: a cohort study. *Oncologist* 2013;18:795–801.
15 Dall P, Lenzen G, Göhler T, *et al*. Trastuzumab in the treatment of elderly patients with early breast cancer: results from an observational study in Germany. *J Geriatr Oncol* 2015;6:462–9.
16 Goldhar HA, Yan AT, Ko DT, *et al*. The temporal risk of heart failure associated with adjuvant Trastuzumab in breast cancer patients: A population study. *J Natl Cancer Inst* 2016;108:djv301.
17 Palmieri C, Shah D, Krell J, *et al*. Management and outcome of Her2-positive early breast cancer treated with or without Trastuzumab in the adjuvant Trastuzumab era. *Clin Breast Cancer* 2011;11:93–102.
18 Piotrowski G, Gawor R, Stasiak A, *et al*. Cardiac complications associated with Trastuzumab in the setting of adjuvant chemotherapy for breast cancer overexpressing human Epidermal growth factor receptor type 2 - a prospective study. *Arch Med Sci* 2012;8:227–35.
19 Freedman RA, Vaz-Luis I, Barry WT, *et al*. Patterns of chemotherapy, toxicity, and short-term outcomes for older women receiving adjuvant Trastuzumab-based therapy. *Breast Cancer Res Treat* 2014;145:491–501.
20 Dall P, Koch T, Göhler T, *et al*. Trastuzumab in human Epidermal growth factor receptor 2-positive early breast cancer: results of a prospective, Noninterventional study on routine treatment between 2006 and 2012 in Germany. *Oncologist* 2017;22:131–8.
21 Dall P, Koch T, Göhler T, *et al*. Trastuzumab without chemotherapy in the adjuvant treatment of breast cancer: subgroup results from a large observational study. *BMC Cancer* 2018;18:51.
22 Ring A, Battisti NML, Reed MWR, *et al*. Bridging the age gap: observational cohort study of effects of chemotherapy and Trastuzumab on recurrence, survival and quality of life in older women with early breast cancer. *Br J Cancer* 2021;125:209–19.
23 Cadoo KA, Morris PG, Cowell EP, *et al*. Adjuvant chemotherapy and Trastuzumab is safe and effective in older women with small, node-negative, Her2-positive early-stage breast cancer. *Clin Breast Cancer* 2016;16:487–93.
24 Benchimol EI, Smeeth L, Guttmann A, *et al*. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
25 Sardesai S, Sukumar J, Kassem M, *et al*. Clinical impact of interruption in adjuvant Trastuzumab therapy in patients with operable HER-2 positive breast cancer. *Cardiooncology* 2020;6:26.
26 Henson KE, Elliss-Brookes L, Coupland VH, *et al*. Data resource profile: National cancer registration Dataset in England. *Int J Epidemiol* 2020;49:16–16h.
27 Herbert A, Wijlaars L, Zylbersztejn A, *et al*. Data resource profile: hospital episode Statistics admitted patient care (HES APC). *Int J Epidemiol* 2017;46:1093–1093i.
28 Bright CJ, Lawton S, Benson S, *et al*. Data resource profile: the systemic anti-cancer therapy (SACT) Dataset. *Int J Epidemiol* 2020;49:15–15l.
29 Henson KE, Brock R, Shand B, *et al*. Cohort profile: prescriptions dispensed in the community linked to the National cancer Registry in England. *BMJ Open* 2018;8:e020980.
30 Boyle JM, Cowling TE, Kuryba A, *et al*. Development and validation of a coding framework to identify severe acute toxicity from systemic anti-cancer therapy using hospital administrative data. *Cancer Epidemiol* 2022;77:S1877-7821(22)00001-7.
31 Gannon MR, Park MH, Miller K, *et al*. Concordance of cancer drug therapy information derived from routinely collected hospital admissions data and the systemic anti-cancer therapy (SACT) Dataset, for older women diagnosed with early invasive breast cancer in England. *Cancer Epidemiol* 2023;83:S1877-7821(23)00017-6.
32 Boyle JM, Kuryba A, Braun MS, *et al*. Validity of chemotherapy information derived from routinely collected Healthcare data: A national cohort study of colon cancer patients. *Cancer Epidemiol* 2021;73:S1877-7821(21)00088-6.
33 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.
34 Baselga J, Perez EA, Pienkowski T, *et al*. Adjuvant Trastuzumab: a milestone in the treatment of HER-2-positive early breast cancer. *Oncologist* 2006;11 Suppl 1:4–12.
35 Onitilo AA, Engel JM, Stankowski RV. Cardiovascular toxicity associated with adjuvant Trastuzumab therapy: prevalence, patient characteristics, and risk factors. *Ther Adv Drug Saf* 2014;5:154–66.
36 Jackisch C, Stroyakovskiy D, Pivot X, *et al*. Subcutaneous vs intravenous Trastuzumab for patients with Erbb2-positive early breast cancer: final analysis of the Hannah phase 3 randomized clinical trial. *JAMA Oncol* 2019;5:e190339.
37 Pivot X, Gligorov J, Müller V, *et al*. Patients' preferences for subcutaneous Trastuzumab versus conventional intravenous infusion for the adjuvant treatment of Her2-positive early breast cancer: final

analysis of 488 patients in the International, randomized, two-cohort Prefher study. *Ann Oncol* 2014;25:1979–87.
38 Vaz-Luis I, Keating NL, Lin NU, *et al*. Duration and toxicity of adjuvant Trastuzumab in older patients with early-stage breast cancer: a population-based study. *J Clin Oncol* 2014;32:927–34.
39 Tarantini L, Cioffi G, Gori S, *et al*. Trastuzumab adjuvant chemotherapy and cardiotoxicity in real-world women with breast cancer. *J Card Fail* 2012;18:113–9.
40 Seferina SC, de Boer M, Derksen MW, *et al*. Cardiotoxicity and cardiac monitoring during adjuvant Trastuzumab in daily Dutch practice: A study of the southeast Netherlands breast cancer consortium. *Oncologist* 2016;21:555–62.
41 Naumann D, Rusius V, Margiotta C, *et al*. Factors predicting Trastuzumab-related cardiotoxicity in a real-world population of women with Her2+ breast cancer. *Anticancer Res* 2013;33:1717–20.
42 Thavendiranathan P, Abdel-Qadir H, Fischer HD, *et al*. Breast cancer therapy-related cardiac dysfunction in adult women treated in routine clinical practice: A population-based cohort study. *J Clin Oncol* 2016;34:2239–46.
43 Cohen JB, Brown NJ, Brown S-A, *et al*. Cancer therapy-related hypertension: A scientific statement from the American heart Association. *Hypertension* 2023;80:e46–57.
44 Szczepaniak P, Siedlinski M, Hodorowicz-Zaniewska D, *et al*. Breast cancer chemotherapy induces vascular dysfunction and hypertension through a Nox4-dependent mechanism. *J Clin Invest* 2022;132:e149117.
45 Earl HM, Hiller L, Vallier A-L, *et al*. 6 versus 12 months of adjuvant Trastuzumab for Her2-positive early breast cancer (PERSEPHONE): 4-year disease-free survival results of a randomised phase 3 non-inferiority trial. *Lancet* 2019;393:2599–612.
46 Brain E, Caillet P, de Glas N, *et al*. Her2-targeted treatment for older patients with breast cancer: an expert position paper from the International society of geriatric oncology. *Journal of Geriatric Oncology* 2019;10:1003–13.
47 Sawaki M, Taira N, Uemura Y, *et al*. Adjuvant Trastuzumab without chemotherapy for treating early Her2-positive breast cancer in older patients: A propensity score-adjusted analysis of a prospective cohort study. *Breast* 2022;66:245–54.
48 Gannon MR, Dodwell D, Jauhari Y, *et al*. Initiation of adjuvant chemotherapy and Trastuzumab for human Epidermal growth receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England. *J Geriatr Oncol* 2020;11:836–42.
49 Morris PG, Hudis CA. Trastuzumab-related cardiotoxicity following anthracycline-based adjuvant chemotherapy: how worried should we be *JCO* 2010;28:3407–10.

5.4. Research Paper 6

Title

Survival following adjuvant trastuzumab-based treatment among older patients with HER2-positive early invasive breast cancer: A national population-based cohort study using routine data.

Journal details

Under review at European Journal of Cancer. This is the version that was submitted to the journal at the time of thesis submission. Supplementary material can be found in the Appendices.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT

T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	P1805830	**Title**	Mrs
First Name(s)	Melissa Ruth		
Surname/Family Name	Gannon		
Thesis Title	Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.		
Primary Supervisor	Professor David Cromwell		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	European Journal of Cancer
Please list the paper's authors in the intended authorship order:	Melissa Ruth Gannon, David Dodwell, Katie Miller, Jibby Medina, Karen Clements, Kieran Horgan, Min Hae Park, David Alan Cromwell

Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed the work, processed, analysed and interpreted the data, drafted the article, approved the final version to be published

SECTION E

Student Signature	
Date	

Supervisor Signature	
Date	

Survival following adjuvant trastuzumab-based treatment among older patients with HER2-positive early invasive breast cancer: A national population-based cohort study using routine data

Melissa Ruth Gannon (M.R.G.)[a b], David Dodwell (D.D.)[c], Katie Miller (K.M.)[a b], Jibby Medina (J.M.)[a b], Karen Clements (K.C.)[d], Kieran Horgan (K.H.)[e], Min Hae Park (M.H.P.)[a b], David Alan Cromwell (D.A.C.)[a b]

[a] Department of Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK, [b] Clinical Effectiveness Unit, The Royal College of Surgeons of England, London, UK, [c] Nuffield Department of Population Health, University of Oxford, Oxford, UK, [d] National Cancer Registration and Analysis Service, NHS England, 5th Floor, 23 Stephenson Street, Birmingham, UK, [e] Department of Breast Surgery, St James's University Hospital, Leeds, UK.

ABSTRACT

Background: Randomised controlled trials (RCTs) reported adjuvant trastuzumab-based treatment improved overall survival (OS) among patients with HER2-positive early invasive breast cancer (EIBC). Few RCTs included older patients or those with comorbidity/frailty. This study aimed to determine whether the effect of adjuvant trastuzumab-based treatment on survival outcomes varies by patient age and fitness, using national data from routine care.

Methods: Women (50+ years) newly-diagnosed with HER2-positive EIBC between 2014 and 2019 were identified from England Cancer Registry data. Registration records were linked to Systemic Anti-Cancer Therapy data for treatment details and ONS death register for mortality details. A propensity score analysis employing the inverse probability of treatment weighting method was used to balance the patient variables across treatment groups. Cox models were used to evaluate whether the effect of treatment on OS was associated with patient age and fitness; competing risks regression models were used for breast cancer-specific survival (BCSS).

Results: 5,238 women initiated adjuvant trastuzumab-based treatment. Median follow-up was 56.7 months. Comparison with 3,421 women who did not receive adjuvant trastuzumab highlighted differences at diagnosis in relation to age, fitness, grade, nodal involvement, surgery type and use of radiotherapy. Weighted survival analysis found trastuzumab was associated with improved OS (hazard ratio HR 0.56, 95%CI: 0.45-0.70) and improved BCSS (subHR 0.62, 95%CI: 0.47-0.82). We found no evidence of a difference in effect by age or patient fitness for either outcome.

Conclusion: In this national dataset, adjuvant trastuzumab was associated with improvements in survival, with an OS effect size similar to RCTs evidence. The effect size was not found to vary by patient age or fitness. Chronological age and fitness alone should not be barriers to receipt of effective adjuvant targeted treatment.

Introduction

Trastuzumab was approved for use within English NHS services for patients with HER2-positive early invasive breast cancer (EIBC) in 2006, following evidence of efficacy from several randomised controlled trials (RCTs), most notably the HERA trial.(1) A subsequent meta-analysis, which included 11,991 patients across eight RCTs with a median follow-up of 36 months, reported a hazard ratio (HR) of 0.66 (95% confidence interval [CI]: 0.57-0.77) for overall survival (OS).(2) National and international guidelines, including those published by the European Society of Medical Oncology (ESMO), NICE and the American Society of Clinical Oncology (ASCO), recommend treatment with chemotherapy and trastuzumab.(3-5) Additionally the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA) make comparable recommendations for the management of older patients with breast cancer.(6, 7)

Women aged 70 years and older account for more than one-third of breast cancers diagnosed annually in England. However, the evidence base for trastuzumab in this population is narrow due to the limited number of older women participating in trials; an estimated 2.5% of all patients in RCTs evaluating adjuvant trastuzumab were aged 70+ years. An individual patient data meta-analysis found no evidence that the benefit of trastuzumab, compared with chemotherapy alone, differed by patient age but this result was based on trials that included few older women.(8) The resulting lack of outcome evidence in older patients can result in considerable variation of management in routine care according to age.(9, 10) In the context of trastuzumab, we have previously reported increasing age was associated with a reduced used of adjuvant trastuzumab, even after accounting for tumour characteristics and comorbidities.(11)

Among patient populations where evidence is limited, observational studies can provide information to support treatment decision-making. Within a UK setting, a handful of small studies have described survival among women treated with adjuvant trastuzumab, but none looked at differences in treatment effect across age groups. For example, a study that included patients with HER2-positive EIBC treated with trastuzumab at a single London NHS trust from 2006 to 2008 reported an association between trastuzumab use and improved disease-free survival at 3-years, but had insufficient deaths to look at OS and did not examine if effect size was associated with patient factors.(12) A study conducted in South-East Wales reported improved 3-year OS among women treated with trastuzumab, but did not report a formal estimate of possible treatment effect.(13) In the absence of trials to provide the evidence, an observational study can be designed to mimic a randomised trial and provide estimate of treatment effect from the hypothetical “target trial”, an approach that is advocated to reduce the risk of bias from limitations of study design.(14)

The aim of this study was to evaluate whether the effect of adjuvant trastuzumab-based treatment on survival varies by patient age and fitness, among women aged 50+ years diagnosed with HER2-positive EIBC using a national, population-based dataset. We hypothesised that there would be no difference in treatment effect by age or fitness. The study also aimed to investigate factors associated with differences in survival following treatment with trastuzumab, to understand which patients had better/worse survival.

The study is reported according to the RECORD extension to STROBE guidelines for observational studies using routinely collected data.(15)

Patients and Methods

Study design

This non-randomised, retrospective population-based cohort study was designed using the methodological framework of a target trial emulation approach.(14) Within this approach, the observational study is designed to emulate the set-up of a hypothetical RCT, i.e. the 'Target Trial' by applying analogous criteria relating to patient eligibility, treatment assignment, definition of the follow-up period and analysis plan.(14, 16, 17) Supplementary Table S1 summarises the steps followed to guide selection of the study cohort and the conduct of the statistical analysis. We adopt this method primarily to reduce the risk of bias in the estimated hazard ratio of initiating trastuzumab-based treatment versus not. The study results are not interpreted as estimating the causal effects of treatment.

Data Source

This study was undertaken as part of the National Audit of Breast Cancer in Older Patients (NABCOP; see www.nabcop.org.uk for full details). The NABCOP received pseudonymised patient-level Cancer Registry data for all women aged 50+ years, diagnosed with breast cancer (BC) from 1 January 2014 to 31 December 2019 within NHS trusts in England.(18) Records were linked at tumour-level to data from the Cancer Outcomes and Services Dataset (COSD); Hospital Episode Statistics Admitted Patient Care (HES-APC) (19); Systemic Anti-Cancer Therapy (SACT) dataset (20); national Radiotherapy Dataset (RTDS); and at patient-level to the Primary Care Prescription Database (PCPD).(21)

Study cohort

The study cohort was defined to include women aged 50+ years diagnosed with HER2-positive EIBC (stage 1-3A; ICD-10 C50) who received surgery within six months of diagnosis. BC was classified as HER2-positive where HER2 status was reported as either positive or borderline but with a positive HER2-FISH (fluorescence in situ hybridization) or equivalent test result. Primary surgery was defined as either breast-conserving surgery (BCS) or mastectomy, identified from Office of Population Censuses and Surveys (OPCS) procedure codes entered within HES-APC.

In line with the emulated trial design, the following exclusion criteria were applied. First, the records of women were excluded if they had: (i) neoadjuvant chemotherapy or trastuzumab recorded, (ii) adjuvant chemotherapy or trastuzumab started more than 4 months after surgery, or (iii) adjuvant treatment included other HER2-targeting agents. Additionally records with missing information on patient fitness, tumour stage, nodal stage and invasive grade, or where the date of death or censoring was before the landmark point were excluded (see Statistical analysis section).

Treatment assignment

A clinically relevant defined grace period of 4 months to treatment assignment (initiated trastuzumab-based treatment or not) was applied, to reflect the time required for decision-making in routine clinical practice. Patients were defined as initiating adjuvant trastuzumab-based treatment if their SACT data records contained any of trastuzumab, Herceptin or trastuzumab biosimilar in the drug name field, within four months after surgery or after chemotherapy that started within four months of surgery. Records of trastuzumab-emtansine or other HER2-targeting therapy such as pertuzumab were categorised separately as "other HER2-targeting therapy".

Adjuvant chemotherapy, as part of trastuzumab-based treatment, was identified from SACT data where the first recorded cycle was within four months after surgery. It was categorised as: "sequential" where cycles were administered prior to the first trastuzumab date, with no cycles delivered during the trastuzumab cycles; and "concurrent" where any cycles were administered either on the same day as trastuzumab or between trastuzumab cycles (including when chemotherapy started prior to and continued during trastuzumab).

HES-APC was used to supplement data on trastuzumab and chemotherapy from SACT.(22) This provided additional cycle-level information for use of adjuvant trastuzumab and/or chemotherapy and associated treatment initiation date.

Outcome and follow-up period

Overall survival was defined as death from any cause. BC-specific survival (BCSS) was defined where the cause of death was recorded as BC. Death details from linked Civil Registration (death) records, including date and cause, were provided within Cancer Registry data.

Time zero (start of follow-up) was defined, based on the grace period, as 4 months after date of surgery. Each patient was followed up to date of death or administrative censoring (October 2021). Mortality data were available up to October 2021.

Study variables

Patient and tumour characteristics were taken from Cancer Registry/COSD. These were: age at diagnosis (years), deprivation, tumour stage (T0-3), nodal stage (N0-2), hormone receptor status (positive, negative/unknown), tumour grade (G1-3). BC was classified as hormone receptor-positive where either of estrogen or progesterone receptor status were recorded as positive. Deprivation was measured using the Index of Multiple Deprivation 2019 rank, based on the patient's postcode at diagnosis, and assigned to national quintiles of deprivation (most [1] to least [5] deprived).

Patient fitness was measured based on comorbidity and frailty. Comorbidity burden (0, 1, 2+; defined using the Royal College of Surgeons of England Charlson Comorbidity Index - CCI) and frailty (fit, mild frailty, moderate-

severe frailty; defined using the Secondary Care Administrative Records Frailty index – SCARF index) were determined using ICD-10 codes recorded in HES-APC within two years prior to diagnosis.(23)

Other treatment characteristics were: type of surgery (BCS, mastectomy), receipt of radiotherapy and receipt of endocrine therapy (ET). Use of radiotherapy was identified based on records within the RTDS dated during the initial treatment episode following diagnosis, defined by sequential use of treatments with no more than an eight month gap. ET use was identified from the PCPD.(21)

Statistical analysis

All analyses were performed in Stata Version 17.

Median follow-up was determined through reverse-censoring on death, in which survival is treated as the event and death as censoring.

Initial analysis investigated whether patient factors were associated with differences in OS following treatment with trastuzumab and was carried out only among patients initiating adjuvant trastuzumab-based treatment for HER2-positive EIBC. Standard survival analysis methods were used to analyse time-to-event data, with OS/BCSS calculated as time from starting treatment to death/death from BC. Kaplan-Meier survival curves were used to visually inspect OS across patient groups. Cox proportional hazards models were used to analyse the association between OS and the study variables, and Fine and Gray regression models for BCSS.

The main analysis estimated the association between use of adjuvant trastuzumab-based treatment and survival outcomes. Patients were included in treatment groups according to their assigned treatment strategy. We employed a landmark approach in which analyses were timed from 4 months after surgery, to allow for treatment to be started and reduce the risk of immortal time bias.(24) Patients were included if they had at least 4 months' follow-up from surgery and had not experienced the outcome of interest (death) within the first 4 months.

To balance the study variables across the treatment groups and thereby minimise bias due to measured confounders, a propensity score analysis employing the inverse probability of treatment weighting (IPTW) method was used.(25) The IPTW method used all patients in the cohort, and the propensity score corresponded to the probability of a patient receiving trastuzumab. The score was calculated for each patient using a logistic regression model that included all factors that could confound the relationship between treatment and the outcome, along with factors prognostic of the outcome.(26) The model included age, deprivation, patient fitness, tumour stage, nodal involvement, invasive grade, hormone receptor status, type of surgery, radiotherapy. Covariate balance was assessed using the standardised mean difference (SMD) with a value of greater than 0.1 taken to indicate significant imbalance.(27) Stabilised weights were calculated for each patient on the basis of the estimated propensity score (28)

Survival curves were created with IPTW-adjusted Kaplan-Meier plots. IPTW-weighted Cox proportional hazard models with a robust “sandwich” variance estimator were used to calculate an IPTW-adjusted HR as an estimate of the relative effect of trastuzumab-based treatment on OS. IPTW-weighted Fine and Gray regression models were used for competing risk analysis of BCSS. An HR below 1·00 favoured the use of trastuzumab. To determine whether the effect of trastuzumab varied by age and patient fitness, interaction terms with treatment were included in the weighted models.

A sensitivity analysis was carried out looking at the impact of including patients with a record of chemotherapy initiation but not trastuzumab; patients were first included in the “trastuzumab” group and then in the “no trastuzumab” group. A further sensitivity analysis was carried out looking at the impact of the landmark time point.

All tests were two-sided, with confidence intervals presented at the 95% level.

Results

A total of 156,375 women aged 50+ years were diagnosed with EIBC between 1 January 2014 and 31 December 2019. There were 14,936 women with HER2-positive EIBC, of whom 11,584 (77.6%) women proceeded to surgery within 6 months of diagnosis.

A total of 2,014 women who received trastuzumab in combination with another HER2-targeting therapy, started adjuvant treatment more than 4 months after surgery or had received neoadjuvant treatment were excluded, along with 40 women who died within four months of surgery, and 187 women with incomplete data. This left 5,238 women who received adjuvant trastuzumab-based treatment with no prior treatment and 3,421 who received no adjuvant treatment (supplementary Figure S1). 684 women who received adjuvant chemotherapy but no trastuzumab contributed to the sensitivity analysis of treatment assignment.

Among 5,238 women receiving trastuzumab-based treatment median trastuzumab duration was 11.7 months (IQR 11.0-12.1). 22.2% received chemotherapy prior to starting trastuzumab. 20.2% were aged 70+ years, 30.6% were recorded to have node-positive EIBC. 68.3% had BCS and 79.3% received radiotherapy.

Overall survival among patients receiving trastuzumab-based treatment

Median follow-up from initiation of adjuvant trastuzumab-based treatment was 59.0 months (interquartile range: 41.5-73.9), at which point 6.5% (n=338/5,238) of the cohort had died. OS estimates were 99.5%, 96.7% and 92.9% at 1, 3 and 5 years respectively from the start of treatment.

Supplementary Figure S2 shows Kaplan-Meier OS estimates overall and stratified by patient and tumour characteristics. For analyses stratified by age at diagnosis, 5 year OS estimates decreased with increasing age, being 96.2%, 94.2%, 84.8% and 64.3% for women aged 50-59, 60-69, 70-79 and 80+ respectively (supplementary Table S2). Estimates were lowest among those with any comorbidity (86.1%) or moderate-severe frailty (81.7%). OS estimates also decreased with increasing grade, increasing tumour stage, nodal involvement and negative/unknown hormone receptor status. Where chemotherapy was given, there was no difference in OS according to whether this was given sequentially or concurrently with trastuzumab.

Adjusted hazard ratios (aHRs) of OS according to patient subgroups, estimated from proportional hazard models, highlighted differences by age with worse OS as age increased ($p<0.0001$). Worse OS was also associated with having any comorbidity (aHR 1.55, 95%CI 1.14-2.11), nodal involvement (aHR 2.00, 95%CI 1.58-2.53), and larger tumours (T2/3 compared to T1), independently of other factors. Conversely hormone receptor-positive EIBC, use of radiotherapy and use of taxanes were independently associated with improved OS.

Association of trastuzumab-based treatment with overall survival

Comparison of patients who received adjuvant trastuzumab-based treatment with those who did not receive treatment, among a total of 8,659 patients, highlighted substantial differences in characteristics (Table 1). Differences were seen in relation to age, fitness (CCI and SCARF Index), grade, nodal involvement, surgery type,

use of radiotherapy. Specifically a higher percentage of women not receiving treatment were older, had at least one comorbidity, had some level of frailty, had grade 1 tumours, no nodal involvement, had had mastectomy and didn't have radiotherapy.

Median follow-up from the landmark time (4 months after surgery) was comparable among women who received adjuvant trastuzumab-based treatment and those who did not receive treatment (overall 56.7 months, interquartile range: 38.2-71.9). Unadjusted OS estimates differed by treatment group, at 92.8% at five years from the landmark time among women who received trastuzumab-based treatment compared with 75.8% among women who did not.

Of the ten variables used to produce the propensity score, seven exhibited substantial imbalance (SMD >0.1) pre-weighting. Following IPTW, the intra-group differences were substantially reduced, and a SMD of <0.1 was achieved for all variables (supplementary Figure S3). The distribution of propensity scores in the two groups is shown in supplementary Figure S4.

In the IPTW Cox regression landmark analysis, we found use of adjuvant trastuzumab-based treatment was associated with improved OS, compared with no treatment (HR 0.56, 95%CI: 0.45-0.70). Use was also associated with improved BCSS (subHR 0.62, 95%CI: 0.47-0.82).

Figure 1 presents overall survival estimates for treatment groups, by dichotomised age (50-69/70+). We found no statistical evidence of effect modification by age. Figure 2 shows the effect of treatment on OS across patient subgroups. We found no evidence of a difference by the presence of comorbidity (interaction p=0.822) or frailty (interaction p=0.923). Additionally, tumour stage (T1/T2/T3; interaction p=0.773), nodal involvement (N0/N+; interaction p=0.535285), grade (G1/G2/G3; interaction p=0.212), or hormone receptor-positive status (interaction p=0.853) were not associated with differences in OS. Similar associations were seen when looking at BCSS (Figure 3).

Sensitivity analysis which included those women who received adjuvant chemotherapy but no trastuzumab firstly in the trastuzumab group and secondly in the no trastuzumab group did not alter these findings. Additionally both shorter and longer grace periods/landmark points did not alter the findings.

Discussion

This population-based cohort study investigated the impact of factors including patient age and fitness on survival outcomes following initiation of adjuvant trastuzumab-based treatment, among women aged 50+ years diagnosed with HER2-positive EIBC in England.

Five-year OS of women treated with adjuvant trastuzumab-based treatment (93%) was consistent with estimates from both RCTs (89% 4-year OS in HERA, ~90% in N9831 and NSABP B-31, 91-92% 5-year OS in BCIRG-006) and several other population-based studies.(1, 29-34) Increasing age, comorbidity, nodal involvement and larger tumour size were linked to worse OS, whilst hormone receptor-positive EIBC, use of radiotherapy and use of taxanes were independently associated with improved OS. These findings are consistent with those reported by an Italian study looking at predictors of survival.(35)

Adjuvant trastuzumab-based treatment was associated with improved OS (HR 0.56) and BCSS (subHR 0.62), when patients who received adjuvant trastuzumab-based treatment were compared with those who did not. Overall findings were consistent with evidence from RCTs. In particular, the HERA trial reported an unadjusted HR of 0.53 for OS at 4-year follow-up.(36) This study however included more than twice as many patients than were in the HERA trial. The findings are also consistent with other real-world studies estimating the effect of treatment on OS. A study in the Netherlands among patients diagnosed from 2005 to 2007 reported comparable 5-year OS estimates and an associated adjusted HR of 0.48, whilst another study among women in the Netherlands from 2006-12 found adjuvant trastuzumab considerably improved OS for small tumours (adjusted HR 0.35).(30, 37)

This study included patients with an upper age range older than in the RCTs, with 20.2% of women aged 70+ years. We found no evidence that the impact of treatment varied by patient age at diagnosis or fitness, as measured using comorbidity and frailty scores.

The study has a number of strengths. Use of a large, population-based sample of women diagnosed with HER2-positive EIBC over a period of six years (2014-2019) with mortality data to October 2021 means the findings reflect the diversity of women with breast cancer and current survival outcomes. The evidence from this study is more representative of the general population than previously published small observational studies or randomised trials. Use of the propensity score weighting and landmark analysis are recognised methods for reducing bias introduced by patient selection for treatment. The study demonstrated good balance among the prognostic factors associated with treatment selection and clinical outcomes, and this provides confidence in the possible treatment effect estimates.

There are various limitations of this study. Firstly, the use of routine data raises the potential of bias from treatment misclassification, unmeasured confounding, and missing data. Misclassification might arise because some hospitals may not enter data into SACT on all treatments, however we used HES-APC data to identify patients who received treatment and thereby reduced misclassification. SACT provides data on prescribed therapies and some patients may not have received trastuzumab; which means that approach in this paper is

analogous to an intention-to-treat analysis as would be carried out within an RCT. There is potential for errors in patient and tumour characteristics within the England Cancer Registry and COSD datasets, however the cancer registration service has various validation steps when compiling the national registration data and the overall effect of coding errors should therefore be minimal. Secondly, propensity score analysis will not account for unmeasured confounding, and there may be residual bias. This should be small in comparison to the estimated treatment effect because of the large number of variables used to derive the propensity score. Finally, current NICE guidance (NG101) was changed in 2018 and recommended patients are offered neoadjuvant chemotherapy / HER2-targeting therapy for HER2-positive EIBC.(4) This might have had the effect of changing who was offered adjuvant therapy in later years, and increased the number of patients with better prognosis in the cohort for later years. Inclusion of year of diagnosis as a sensitivity analysis demonstrated no impact on the findings.

Conclusions

This study found the use of adjuvant trastuzumab-based treatment, initiated in routine care for women with HER2-positive EIBC, was associated with increased overall survival. This was seen regardless of patient age or fitness. Chronological age and fitness alone should not be barriers to the receipt of effective adjuvant targeted treatment.

Figures & Tables

Table 1: Distribution of patient, tumour and treatment characteristics in women with HER2-positive, early invasive breast cancer, by receipt of adjuvant trastuzumab-based treatment.

	Total	Trastuzumab-based treatment	No treatment	*Chemotherapy only*
		N=5,238	N=3,421	*N=684*
Age				
	50-59 years	2226 (42.5%)	587 (17.2%)	*293 (42.8%)*
	60-69 years	1952 (37.3%)	913 (26.7%)	*235 (34.4%)*
	70-79 years	970 (18.5%)	1010 (29.5%)	*147 (21.5%)*
	80+ years	90 (1.7%)	911 (26.6%)	*9 (1.3%)*
IMD				
	1 - Most deprived	794 (15.2%)	486 (14.2%)	*141 (20.6%)*
	2	886 (16.9%)	579 (16.9%)	*131 (19.2%)*
	3	1119 (21.4%)	678 (19.8%)	*130 (19%)*
	4	1200 (22.9%)	833 (24.3%)	*146 (21.3%)*
	5 - Least deprived	1239 (23.7%)	845 (24.7%)	*136 (19.9%)*
CCI				
	0	4744 (90.6%)	2760 (80.7%)	*609 (89%)*
	1	395 (7.5%)	416 (12.2%)	*59 (8.6%)*
	2+	99 (1.9%)	245 (7.2%)	*16 (2.3%)*
SCARF Index				
	Fit	4478 (85.5%)	2453 (71.7%)	*579 (84.6%)*
	Mild frailty	492 (9.4%)	450 (13.2%)	*71 (10.4%)*
	Moderate - severe frailty	268 (5.1%)	518 (15.1%)	*34 (5%)*
Grade				
	G1	100 (1.9%)	232 (6.8%)	*11 (1.6%)*
	G2	1925 (36.8%)	1631 (47.7%)	*279 (40.8%)*
	G3	3213 (61.3%)	1558 (45.5%)	*394 (57.6%)*
Tumour stage				
	T1	2954 (56.4%)	1923 (56.2%)	*341 (49.9%)*
	T2	2113 (40.3%)	1385 (40.5%)	*314 (45.9%)*
	T3	171 (3.3%)	113 (3.3%)	*29 (4.2%)*
Nodal stage				
	N0	3633 (69.4%)	2607 (76.2%)	*404 (59.1%)*
	N+	1605 (30.6%)	814 (23.8%)	*280 (40.9%)*
Hormone receptor-positive				
	Yes	3601 (68.7%)	2519 (73.6%)	*487 (71.2%)*
	No/Unknown	1637 (31.3%)	902 (26.4%)	*197 (28.8%)*
Surgery type				
	Breast conserving surgery	3577 (68.3%)	2072 (60.6%)	*432 (63.2%)*
	Mastectomy	1661 (31.7%)	1349 (39.4%)	*252 (36.8%)*
Radiotherapy				
	No	1082 (20.7%)	1333 (39.0%)	*160 (23.4%)*
	Yes	4156 (79.3%)	2088 (61.0%)	*524 (76.6%)*
Endocrine therapy				
	No	1627 (31.1%)	984 (28.8%)	*199 (29.1%)*
	Yes	3611 (68.9%)	2437 (71.2%)	*485 (70.9%)*
Chemotherapy				
	No	72 (1.4%)	3421 (100%)	*0 (0%)*
	Other chemotherapy	51 (1%)	-	*327 (47.8%)*
	Taxanes	2120 (40.5%)	-	*52 (7.6%)*
	Anthracyclines	969 (18.5%)		*179 (26.2%)*
	Taxane & anthracycline	2026 (38.7%)		*126 (18.4%)*
	Death reported (Y)	338 (6.5%)	741 (21.7%)	*76 (11.1%)*

Key: IMD = Index of Multiple Deprivation; CCI = Charlson Comorbidity Index; SCARF = Secondary Care Administrative Records Frailty.

Note: Anthracyclines = doxorubicin, epirubicin, mitoxantrone recorded in SACT.

Taxanes = docetaxel, cabazitaxel, paclitaxel, nab-paclitaxel recorded in SACT.

Figure 1: Weighted Kaplan-Meier curves (including 95% confidence intervals) for overall survival in patients with HER2-positive EIBC receiving adjuvant trastuzumab-based treatment compared with no treatment, by age at diagnosis

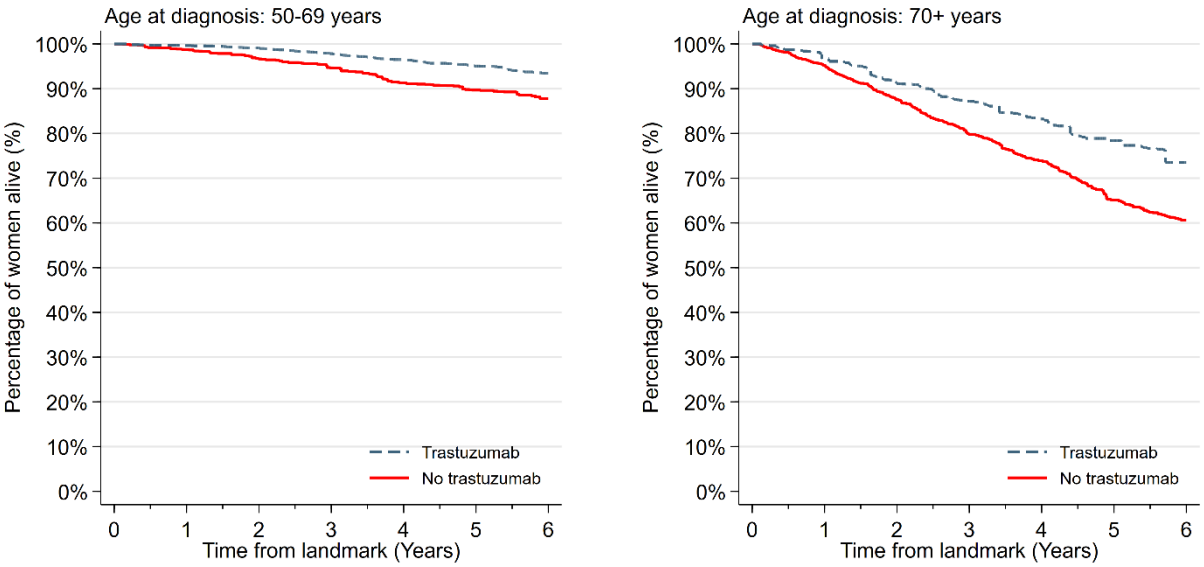


Figure 2: Forest plot of estimated hazard ratios for overall survival (trastuzumab versus no trastuzumab) in patients with HER2-positive EIBC

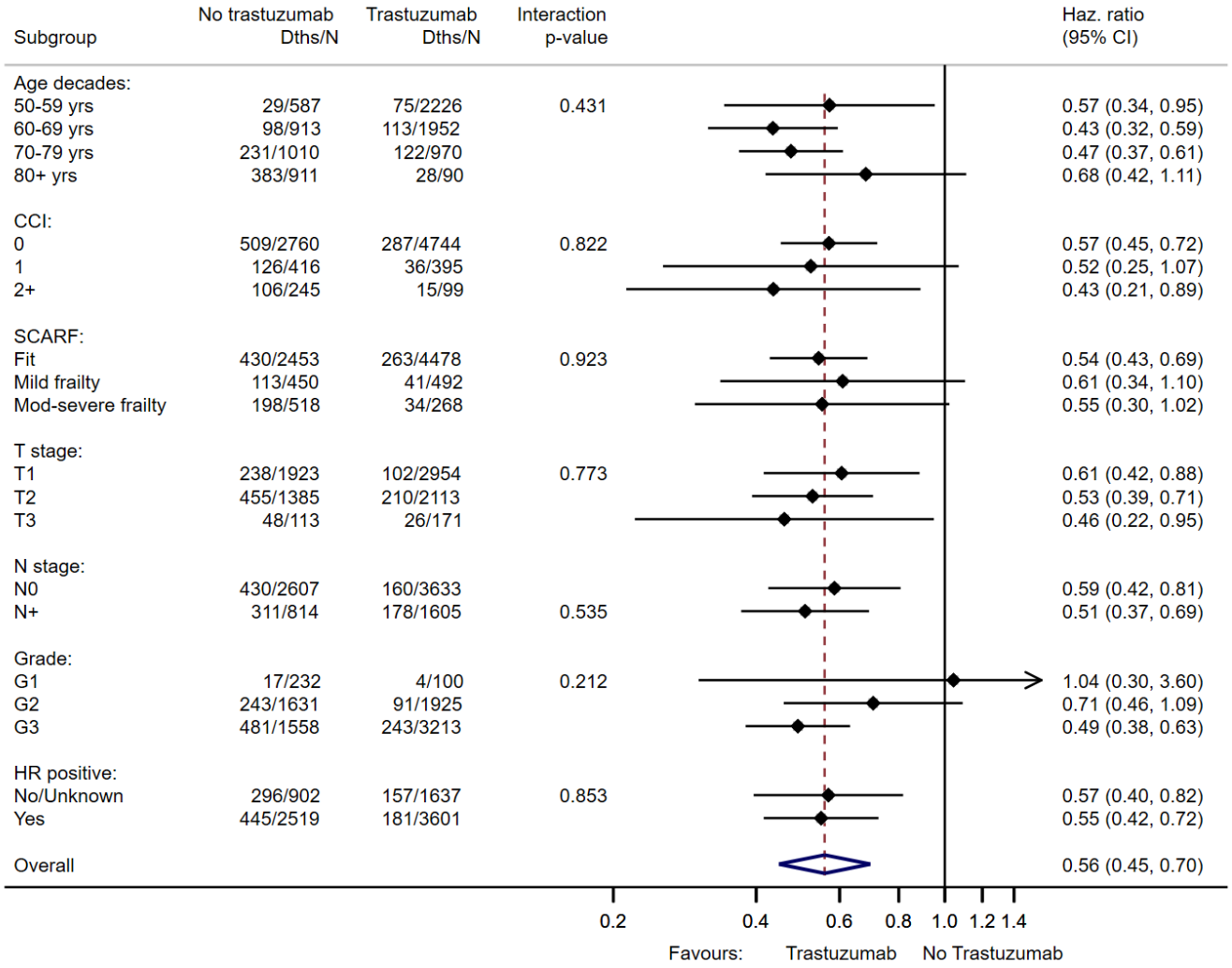
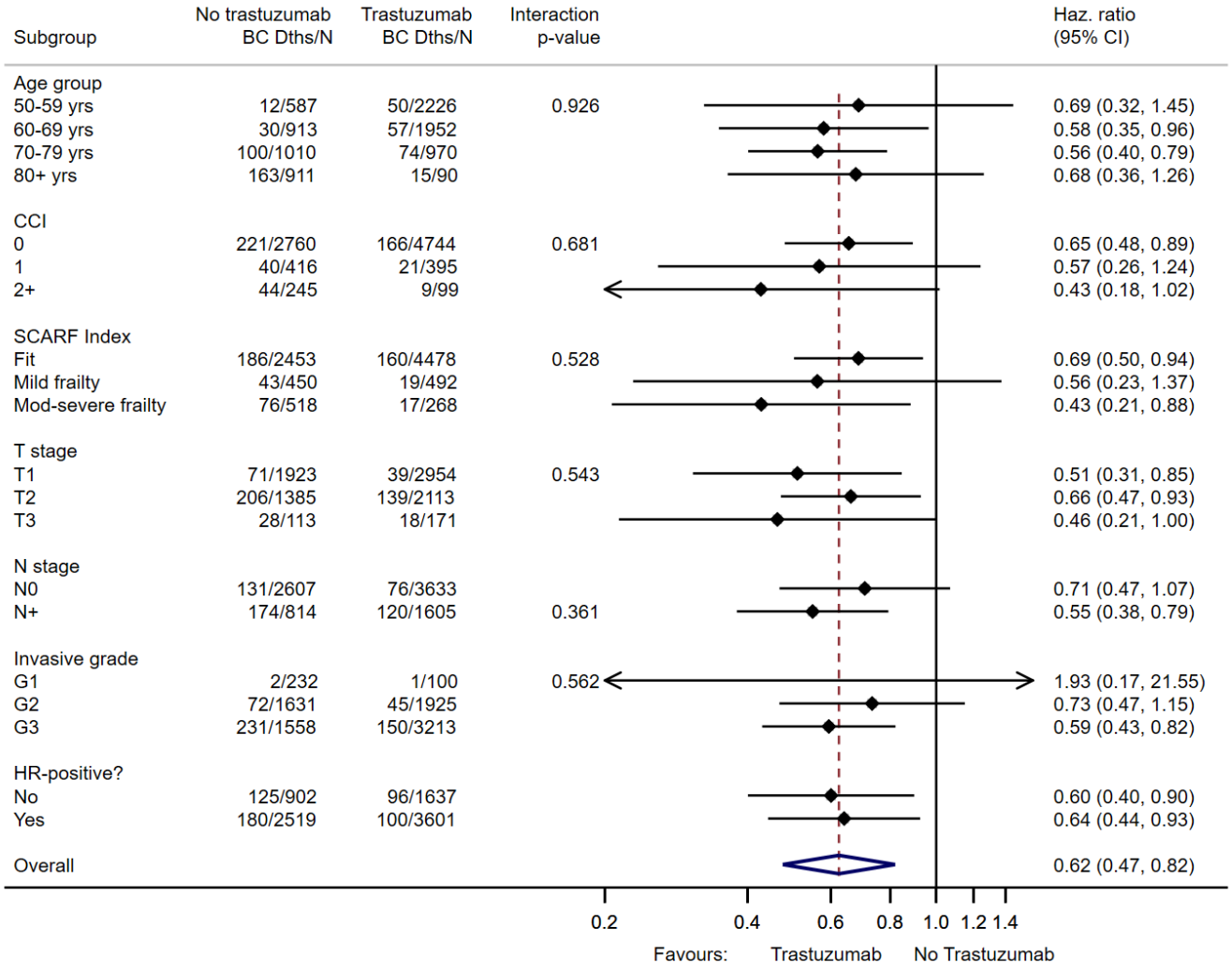


Figure 3: Forest plot of estimated sub hazard ratios breast cancer-specific survival (trastuzumab versus no trastuzumab) in patients with HER2-positive EIBC



References

1. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, Goldhirsch A, Untch M, Smith I, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. The New England journal of medicine. 2005;353(16):1659-72.

2. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, et al. Trastuzumab containing regimens for early breast cancer. The Cochrane database of systematic reviews. 2012(4):CD006243.

3. Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26 Suppl 5:v8-30.

4. NICE. Early and locally advanced breast cancer: diagnosis and management.2018. Available from: www.nice.org.uk/guidance/ng101.

5. Denduluri N, Somerfield MR, Eisen A, Holloway JN, Hurria A, King TA, et al. Selection of Optimal Adjuvant Chemotherapy Regimens for Human Epidermal Growth Factor Receptor 2 (HER2) –Negative and Adjuvant Targeted Therapy for HER2-Positive Breast Cancers: An American Society of Clinical Oncology Guideline Adaptation of the Cancer Care Ontario Clinical Practice Guideline. Journal of Clinical Oncology. 2016;34(20):2416-27.

6. Biganzoli L, Wildiers H, Oakman C, Marotti L, Loibl S, Kunkler I, et al. Management of elderly patients with breast cancer: updated recommendations of the International Society of Geriatric Oncology (SIOG) and European Society of Breast Cancer Specialists (EUSOMA). The Lancet Oncology. 2012;13(4):e148-60.

7. Giordano SH, Temin S, Chandarlapaty S, Crews JR, Esteva FJ, Kirshner JJ, et al. Systemic Therapy for Patients With Advanced Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: ASCO Clinical Practice Guideline Update. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2018;36(26):2736-40.

8. Early Breast Cancer Trialists' Collaborative g. Trastuzumab for early-stage, HER2-positive breast cancer: a meta-analysis of 13 864 women in seven randomised trials. The Lancet Oncology. 2021;22(8):1139-50.

9. Ring A. The influences of age and co-morbidities on treatment decisions for patients with HER2-positive early breast cancer. Critical reviews in oncology/hematology. 2010;76(2):127-32.

10. Lavelle K, Todd C, Moran A, Howell A, Bundred N, Campbell M. Non-standard management of breast cancer increases with age in the UK: a population based cohort of women > or =65 years. British journal of cancer. 2007;96(8):1197-203.

11. Gannon MR, Dodwell D, Jauhari Y, Horgan K, Clements K, Medina J, et al. Initiation of adjuvant chemotherapy and trastuzumab for human epidermal growth receptor 2-positive early invasive breast cancer in a population-based cohort study of older women in England. Journal of geriatric oncology. 2020;11(5):836-42.

12. Palmieri C, Shah D, Krell J, Gojis O, Hogben K, Riddle P, et al. Management and outcome of HER2-positive early breast cancer treated with or without trastuzumab in the adjuvant trastuzumab era. Clin Breast Cancer. 2011;11(2):93-102.

13. Webster RM, Abraham J, Palaniappan N, Caley A, Jasani B, Barrett-Lee P. Exploring the use and impact of adjuvant trastuzumab for HER2-positive breast cancer patients in a large UK cancer network. Do the results of international clinical trials translate into a similar benefit for patients in South East Wales? British journal of cancer. 2012;106(1):32-8.

14. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.

15. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885.

16. Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of clinical epidemiology. 2016;79:70-5.

17. Antoine A, Perol D, Robain M, Delaloge S, Lasset C, Drouet Y. Target trial emulation to assess real-world efficacy in the ESME metastatic breast cancer cohort. J Natl Cancer Inst. 2023.

18. Henson KE, Elliss-Brookes L, Coupland VH, Payne E, Vernon S, Rous B, et al. Data Resource Profile: National Cancer Registration Dataset in England. Int J Epidemiol. 2020;49(1):16-h.

19. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Int J Epidemiol. 2017;46(4):1093-i.

20. Bright CJ, Lawton S, Benson S, Bomb M, Dodwell D, Henson KE, et al. Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) dataset. Int J Epidemiol. 2020;49(1):15-l.

21. Henson KE, Brock R, Shand B, Coupland VH, Elliss-Brookes L, Lyratzopoulos G, et al. Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England. BMJ open. 2018;8(7):e020980.

22. Gannon MR, Park MH, Miller K, Dodwell D, Horgan K, Clements K, et al. Concordance of cancer drug therapy information derived from routinely collected hospital admissions data and the Systemic Anti-Cancer Therapy (SACT) dataset, for older women diagnosed with early invasive breast cancer in England. Cancer Epidemiol. 2023;83:102337.

23. Jauhari Y, Gannon MR, Dodwell D, Horgan K, Clements K, Medina J, 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(5):e035395.

24. Dafni U. Landmark analysis at the 25-year landmark point. Circ Cardiovasc Qual Outcomes. 2011;4(3):363-71.

25. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Statistics in medicine. 2014;33(7):1242-58.

26. Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, et al. An introduction to inverse probability of treatment weighting in observational research. Clin Kidney J. 2022;15(1):14-20.

27. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Statistics in medicine. 2009;28(25):3083-107.

28. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine. 2015;34(28):3661-79.

29. Tang M, Schaffer A, Kiely BE, Daniels B, Simes RJ, Lee CK, et al. Treatment patterns and survival in HER2-positive early breast cancer: a whole-of-population Australian cohort study (2007-2016). British journal of cancer. 2019;121(11):904-11.

30. Seferina SC, Lobbezoo DJ, de Boer M, Dercksen MW, van den Berkmortel F, van Kampen RJ, et al. Real-Life Use and Effectiveness of Adjuvant Trastuzumab in Early Breast Cancer Patients: A Study of the Southeast Netherlands Breast Cancer Consortium. Oncologist. 2015;20(8):856-63.

31. Dall P, Koch T, Gohler T, Selbach J, Ammon A, Eggert J, et al. Trastuzumab in Human Epidermal Growth Factor Receptor 2-Positive Early Breast Cancer: Results of a Prospective, Noninterventional Study on Routine Treatment Between 2006 and 2012 in Germany. Oncologist. 2017;22(2):131-8.

32. Perez EA, Romond EH, Suman VJ, Jeong JH, Sledge G, Geyer CE, Jr., et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis of overall survival from NSABP B-31 and NCCTG N9831. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(33):3744-52.

33. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant trastuzumab in HER2-positive breast cancer. The New England journal of medicine. 2011;365(14):1273-83.

34. Dall P, Lenzen G, Gohler T, Lerchenmuller C, Feisel-Schwickardi G, Koch T, et al. Trastuzumab in the treatment of elderly patients with early breast cancer: Results from an observational study in Germany. Journal of geriatric oncology. 2015;6(6):462-9.

35. Bonifazi M, Franchi M, Rossi M, Zambelli A, Moja L, Zambon A, et al. Long term survival of HER2-positive early breast cancer treated with trastuzumab-based adjuvant regimen: a large cohort study from clinical practice. Breast. 2014;23(5):573-8.

36. Gianni L, Dafni U, Gelber RD, Azambuja E, Muehlbauer S, Goldhirsch A, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. The Lancet Oncology. 2011;12(3):236-44.

37. van Ramshorst MS, van der Heiden-van der Loo M, Dackus GM, Linn SC, Sonke GS. The effect of trastuzumab-based chemotherapy in small node-negative HER2-positive breast cancer. Breast cancer research and treatment. 2016;158(2):361-71.

6. Discussion

The research presented within this thesis explored the value of routinely collected healthcare data in the evaluation of systemic oncological treatments for newly-diagnosed invasive breast cancer. This involved concentrating on two areas of research, firstly methodological work investigated the recording of treatment information within routine healthcare data and secondly analytical work evaluated the utilisation of systemic oncological treatments in clinical practice, with a focus on aspects of adjuvant trastuzumab-based treatment for HER2-positive EIBC.

This chapter summarises the main findings of the work presented in this thesis, discusses the strengths and limitations of the work as a whole and then describes implications for clinical practice, policy and future research.

6.1. Summary of findings

6.1.1. Methodological Work

Understanding recording of treatment information in routine data

The SACT dataset is the primary source of information on systemic oncological treatment prescribed within routine care in the English NHS.

The initial piece of research was designed to evaluate the consistency of cancer drug therapy (CDT; chemotherapy and trastuzumab) information recorded within the SACT and HES-APC datasets for women with EIBC in England (Section 4.1). The research used HES-APC as the independent data source against which the completeness of treatment recording in the SACT dataset could be assessed. This was feasible during the study period because the majority of systemic oncological treatments were administered to patients as a day case hospital admission. A secondary objective of the research was to assess whether HES-APC data might be valuable in supplementing information within the SACT dataset for certain patient subgroups whose treatment information was found to be incomplete in the SACT dataset.

The analysis included a cohort of nearly 130,000 women aged 50 years and over diagnosed with EIBC in England from January 2014 to December 2019, with treatment data up to March 2021. Using linked patient-level data from the SACT dataset and HES-APC dataset the concordance of cancer drug therapy information across data sources was assessed for both the overall cohort and among patient subgroups.

Overall agreement on CDT use was high across both data sources, at 94%. Where both data sources captured CDT use there was also reasonable agreement on the number of cycles.

Among 31,693 women with CDT recorded in either source, 9% had CDT recorded only in HES-APC data. This was an important finding that highlights a potential limitation of using only the SACT dataset to describe CDT use in routine care. Moreover, it was older patients who were most likely to have CDT only recorded in HES-APC data.

Systematic differences in CDT capture in the SACT dataset were also highlighted by year of diagnosis, NHS trust, and type of drug administration (oral or intravenous/subcutaneous).

This work found it was possible to identify trastuzumab and FEC (fluorouracil, epirubicin, cyclophosphamide) chemotherapy regimen from OPCS code combinations in HES-APC data, with good agreement in the SACT dataset; however, it was more difficult to distinguish use of taxane chemotherapy. The results highlighted the value of incorporating information from HES-APC data when looking at use of systemic oncological treatments, particularly in older patients, to ensure the cohort of patients receiving treatment are accurately identified. This is of particular importance when estimating both use of treatment and the impact of treatment on subsequent outcomes. In the case of the latter scenario, misclassification of patients receiving treatment to a "no treatment" group would underestimate a treatment effect size, where there was one.

The second piece of methodological work looked at the recording of another systemic oncological treatment commonly given alongside treatments such as chemotherapy (Section 4.2). Endocrine therapy is a widely used treatment for women with hormone receptor-positive invasive breast cancer, with national guidelines recommending its use for five years either as adjuvant to surgery or as primary treatment. As such, it is a key variable to describe a person's care pathway and might be expected to be an important explanatory variable (confounder) in any analysis of comparative effectiveness.

The methodological work was novel in that it dealt with information across data sources from primary and secondary care settings, reflecting that endocrine therapy is typically initiated in secondary care and subsequently prescribed within primary care. Thus, combining data from primary and secondary care had the potential to provide a more complete and in-depth understanding of the use of this cancer treatment.

This research used patient-level data from the Primary Care Prescription Database (PCPD) and several secondary care data sources. The completeness and consistency of endocrine therapy information recorded within secondary care data and the PCPD were compared and factors associated with missed recording of ET use in secondary care data were identified.

The study included a cohort of over 110,000 women aged 50 years and over diagnosed in England from January 2014 to December 2019. The recording of endocrine therapy in secondary care data was found to provide an incomplete representation of use, with under-reporting by age (lower completeness among younger patients), patient fitness (lowest completeness for patients with no comorbidity/frailty) and characteristics relating to the cancer (lowest completeness for patients with lower stage or HER2-positive breast cancer). Combining information from secondary care and PCPD data provided estimates of endocrine therapy use in line with clinical expectations, across all patient groups, whilst clearly highlighting gaps in completeness within the secondary care data. Linking information from secondary care to the PCPD enabled a more comprehensive account of ET use, including treatment duration and adherence, than using secondary care data alone, establishing PCPD data as crucial in providing complete and longitudinal information on ET use.

This work highlighted the value of combining routine data sources from across care settings in examining data completeness and understanding patterns of care using routinely collected health data. Understanding the various aspects of treatment delivery is of importance in the estimation of treatment effects on subsequent outcomes, particularly when accounting for other treatments is required in order to single out the effect of the treatment of interest. Additionally, where treatments are given over long periods of time, it is beneficial to have longitudinal information to account for this ongoing exposure and any temporal changes. The PCPD not only has the potential to provide more complete reporting of treatment for patients with breast cancer, but given the extent of information collated within the database, the implications of this study extend beyond the breast cancer example presented.

Both studies highlighted the value of combining treatment information from across routine data sources, particularly where sources captured differing but complementary aspects of treatment. The studies highlighted approaches to be taken to increase the likelihood that estimates of treatment use within an NHS setting, based on routine national data, were as complete and reflective of real-world use as possible. This informed the work done to understand systemic oncological treatment in clinical practice.

6.1.2. Understanding systemic oncological treatment in clinical practice

Utilisation of NICE-recommended systemic oncological drugs

A key requirement for the use of real-world data in CER studies is for there to be sufficient uptake of a treatment for the data to represent a mature pattern of use. The work presented in this paper assessed the use of oncological drugs in routine care after they had been recommended by NICE, for women diagnosed with invasive breast cancer in routine practice (Section 5.1). The study also investigated differences in uptake by age, comorbidity burden and geographical region. Where there were sufficient time periods, the impact of NICE guidance on use was evaluated using interrupted time series analysis.

The study used information captured in the SACT dataset and defined a cohort of more than 160,000 women aged 50 years and over diagnosed with invasive breast cancer in England from January 2014 to December 2019, with patient-level data on use of oncological drugs up to February 2022.

The uptake of oncological treatments recommended by NICE within the last 20 years for the treatment of invasive breast cancer was variable. Use was consistently lower than expected among older women, and increasing levels of comorbidity burden were typically associated with lower likelihood of use (where numbers were sufficient to investigate). For drugs recommended as part of first-line treatment since 2014, the publication of the NICE recommendation was generally associated with an increase in use. This was the first time the uptake of recommended drugs for breast cancer had been documented at a national level and all in one place, enabling comparison across drugs.

The results highlighted trastuzumab as the most well-established and frequently used drug. In doing so it set the scene for more in-depth evaluation of the use of, and outcomes following, trastuzumab-based treatment among older, more frail patients.

Use of and outcomes following adjuvant trastuzumab-based treatment for HER2-positive EIBC

The final three papers in this thesis reported on adjuvant trastuzumab-based treatment (trastuzumab given in combination with chemotherapy) for women with a registered diagnosis of HER2-positive EIBC, using prospective, routinely collected national data. Using clinical trials as a backdrop, the first of these three papers evaluated what happened in clinical practice (Section 5.2). With an ageing population, it is increasingly important to understand how the drug is used in this population. The analysis used multilevel mixed-effects logistic regression to evaluate factors associated with variation in the initiation of adjuvant chemotherapy and trastuzumab.

The study found age was associated with low use, regardless of other factors including tumour characteristics or comorbidity burden. Additionally, geographical variation in use was found to be an issue among women aged 50 to 70 years, whilst for older women there was consistently low use across the country.

While trastuzumab-based treatment for HER2-positve EIBC improves survival, its use is associated with increased cardiotoxicity risk. With low numbers of older patients within the trials, little was known about the rates of treatment-related severe acute toxicity among patients receiving adjuvant trastuzumab-based treatment for HER2-positive EIBC in routine care. Specifically, there were no studies providing a comprehensive and extensive breakdown of the safety of trastuzumab-based treatment for older patients. Published data on treatment-related toxicity in older patients is valuable to inform discussions around treatment decisions and help in counselling older patients on the side-effects of treatment.

The second of these three papers characterised the cohort of women (aged 50+ years) with HER2-positive EIBC who received adjuvant trastuzumab-based treatment in routine care in England (Section 5.3). Treatment-related overnight hospital admissions with severe acute toxicity events (SATE) captured in hospital administrative data were identified using ICD-10 diagnosis codes previously validated in colon cancer.[105] Differences by age, comorbidity and trial eligibility were evaluated. The study cohort included over 5,000 women aged 50 years and over diagnosed from January 2014 to December 2019, with linked hospital administrative data up to March 2021. Patient-level English cancer registry data, linked to other secondary care datasets including detailed information on systemic oncological treatment prescriptions, were used to compare the characteristics of patients with HER2-positive EIBC (initiating trastuzumab) with those of patients included in the RCTs evaluating adjuvant trastuzumab, and to compare toxicity rates. This enabled a clearer understanding of which subgroups of patients were under-represented in the RCTs. Comparison was made to a cohort of women who received chemotherapy for HER2-negative EIBC to understand the added burden of trastuzumab.

Among women treated with adjuvant trastuzumab-based treatment in routine care, one in three had at least one SATE. Rates of cardiac events (including congestive cardiac failure) were much lower, but there was

evidence that increasing age was associated with increased odds of cardiac events. The study also presented information on treatment-related toxicity specifically within the population of older women (aged 70+ years) who received adjuvant trastuzumab in routine care, providing information about the safety of treatment among patients underrepresented in RCTs, and quantifying current levels of treatment-related toxicity in this patient group.

The final piece of work in this thesis looked at survival among the cohort of women receiving adjuvant trastuzumab for HER2-positive EIBC in routine care (Section 5.4). The work evaluated whether existing statistical methods for comparative effectiveness research using observational datasets could replicate the estimates of trastuzumab effectiveness reported in RCTs of adjuvant trastuzumab for HER2-positive EIBC, and whether estimates differed across patient subgroups defined based on patient age and fitness. In this, routine healthcare data were analysed using an approach designed to mimic a hypothetical trial including patients with an upper age range older than in the trials. The study cohort included over 8500 women aged 50 years and over diagnosed with HER2-positive EIBC from January 2014 to December 2019, with mortality data up to October 2021.

This piece of work highlighted that analysis of routine data using a trial emulation approach can produce comparable survival estimates to those published for RCTs.[115] The work was focused on mortality outcomes rather than intermediate outcomes such as recurrence due to the known issues with the recording of recurrence within routine data.[79] Among patients diagnosed and treated in routine care adjuvant trastuzumab-based treatment was associated with improved survival. Overall study findings were consistent with evidence from published trials. There was no evidence of difference in the impact of treatment by patient age or fitness, providing real-world evidence of the impact of trastuzumab treatment on survival in patient subgroups underrepresented in RCTs.

6.2. Strengths and limitations

For each study conducted as part of this thesis, the related strengths and limitations are discussed in detail within the associated chapter. This section therefore summarises those strengths and limitations relating to the project as a whole.

6.2.1. Routinely Collected Data

The main strength of the work presented in this thesis was the use of Cancer Registry data linked at patient/tumour-level to other national cancer and administrative datasets. This meant the work had national coverage and used information which covered all areas of clinical practice from the point of diagnosis, including tumour pathology, primary treatment and systemic therapy. Data were available on all women aged 50 years and over with a confirmed diagnosis of breast cancer from 2014 onwards. For England, this equated to around 31,000 women diagnosed with invasive breast cancer per calendar year from all NHS organisations providing breast cancer services. This means that findings from the research are generalisable to the patient population of

women aged 50 years and over, with many of the studies' cohort numbers considerably larger than previous studies published.

A concern in using these data was the potential for errors in patient and tumour characteristics recorded within the England Cancer Registry and COSD datasets, the data sources primarily used to define the patient cohort within each piece of work, which would introduce bias due to potential misclassification of patients.[101,116] Examples of errors which might occur include the wrong information being recorded (for example invalid dates or an incorrect diagnosis code) or changes in information not being reflected within the data. An example of the latter includes the recorded stage information: clinical (pre-treatment) stage may change from an initial recorded stage following further investigation, down staging, or progression before treatment begins. The patient would be treated based on this updated stage, but this may not be reflected in the data. The cancer registration service has various validation steps when compiling the national registration data and so the overall effect of coding errors should be small. For tumour characteristics such as staging which may change over time, data are provided at the different points in the pathway, although it is noted that this information is typically more complete for staging done post-surgery.

For analysis where a comparison with trial populations was carried out, there were limitations around the use of routine data. One aspect was that the full set of information used to determine trial eligibility was not routinely collected, specifically where functional tests (such as cardiac function, liver function) are required and a threshold met to confirm patient eligibility. As such, presence of comorbidity was used in this instance which may have wrongly defined some patients as "trial-eligible" who may not have a formal comorbidity diagnosis recorded in the HES-APC data but in practice would not meet the trial inclusion criteria, meaning it was not possible to comprehensively select patients from routine practice in the same way that a trial cohort is selected. Because of the nature of the way data used to define comorbidity were recorded within the routine data, it was not possible to know if information was missing. A second aspect relates to the completeness of data such as staging, HER2 status and WHO performance status, which were often included in trial eligibility criteria. In the case of WHO performance status, for patients with breast cancer this was poorly completed (<50%), and so when creating a flag for the two RCTs whose inclusion criteria required patients to be WHO performance status 0 or 1 this will have resulted in creating a smaller group of patients defined as being "trial eligible".

SACT dataset

The work undertaken in this thesis focused on systemic oncological treatments, of which there are multiple regimens for breast cancer. Drug details (name and cycle dates) were only available for England (in the SACT dataset), and unavailable for patients treated in Wales.

Although data returns for the SACT dataset were mandatory from April 2014, with full compliance reported from July 2014, lower than expected data returns for some NHS trusts may still be an issue. The work in this thesis comparing capture with that in hospital admissions data has highlighted this to be the case, and mirrors findings from work carried out in both lung and colon cancer.[102,103]

This has implications for studies looking at treatment use where receipt of treatment is of central importance. In addition to treatment information not being completed in some instances, another limitation is that it is prescribed treatment that is recorded in the SACT dataset, of which not all is ultimately administered.[28] This may be because of a time difference in when the treatment decision was made, and entered in the SACT dataset, and when the patient was due to receive treatment, during which time the patient's condition may have deteriorated and therefore their fitness for treatment changed. Within the context of breast cancer, with such large numbers of patients receiving treatment, the impact of this is likely to be considerably smaller than in other cancer types where patient numbers are smaller. Additionally, for studies investigating treatment use from the perspective of equitable access to treatment, information on prescribed rather than administered treatment is useful for showing intent to treat.

Work done within the data cleaning process for analysis looking at systemic oncological treatment cycles, within this thesis, found some cycles in the SACT data with a recorded dose of zero. For these instances, it was unclear how to interpret the treatment record in terms of whether this meant treatment was not administered for these cycles. Using HES-APC data was valuable in these cases, as in cases where the cycle date could be linked to an admission for treatment with the same date, it could be assumed that treatment was given.

Within SACT data, several variables provide information on drugs given, from the combined drug regimen details (within analysis group or benchmark group) and single drug details provided in the drug group variable. The analysis and benchmark groups are reliant on the accurate mapping of drug combinations to regimens which reference each drug given and so in some instances treatment details may be misclassified either by including the name of a drug which wasn't prescribed or where mapping does not completely identify all drugs prescribed. To avoid the issue of potential misclassification in analysis and benchmark groups, the work in this thesis used the individual drug name variable to determine drug combinations and translate them into regimens.

HES-APC data

HES-APC data were a valuable source for checking the recording of systemic oncological treatment in the SACT dataset and were useful in providing extra information for some aspects of systemic oncological treatment. Previous validation work has found the data in HES-APC to be accurate when compared to clinical notes, and as codes are collected for the primary means of financial reimbursement for services provided they are less likely to be prone to reporting bias which might be incurred where data are collected for the purposes of monitoring treatment use, for example. However, when considering systemic oncological treatments beyond chemotherapy and trastuzumab, or indeed when considering details of the type of chemotherapy given for breast cancer (for which there are many regimens and combinations), this data source currently has limited utility due to the use of codes to identify this aspect of care.[117] Additionally coding may include errors and coding practices may vary between units which will have some impact on the accuracy of treatment information gained from this source.

HES-APC data were also used to estimate measures of patient fitness (CCI and SCARF Index) used to define patient subgroups and included within statistical models in relevant analyses.[106,107] These were calculated based on ICD-10 codes recorded in the diagnosis fields, which are the fields used to capture either the diagnostic

reason for an admission or known diagnoses the patient had at the admission. Any comorbidity not recorded within this data source would not have been included in the measure of fitness. Data from primary care general practices was something which would likely have given a more comprehensive picture of patient comorbidity. For example, studies have shown how linkage of the Clinical Practice Research Datalink (CPRD) with other datasets can indicate the level of coverage of secondary care datasets for patients with a specified disease.[118,119] However, the CPRD captures information on only 7% of the population and, due to sample size considerations, was considered to be of limited value to this thesis.[120] Moreover, the cancer registration service aims to capture 100% of all cancer patients. Reassuringly, in the context of fitness for treatment, information required to define eligibility for treatment is primarily clinical information, which would be investigated within the secondary care setting and therefore available within secondary care sources. A paper in 2015, by Crooks et al, found that although CPRD records captured more comorbidity, the Charlson index calculated from secondary care records performed as well as that derived from primary care in predicting survival.[121] The quality of coding in hospital administrative datasets has been shown to be increasingly accurate.[122,123] The RCS Charlson Score was developed for use with administrative data; specifically, using HES-APC data the score was validated among patients receiving surgery, showing good discrimination and calibration, and associated with known risk factors for comorbidity and subsequent outcomes.[106]

6.2.2. Statistical design and analysis

The observational studies carried out within this thesis have many strengths. The use of national routine data based on cancer registrations, linked to other datasets, has enabled large study sizes. For analysis looking at patient subgroups and exploring interactions, the research had large patient numbers to ensure good statistical power to understand patterns across patient subgroups. Additionally, use of data with national coverage allows for the findings from work within this thesis to have national impact, being relevant for the whole population.

The work used validated methods for identifying treatment-related severe acute toxicity events recorded within hospital admissions data, which allowed for estimation of an important aspect of the safety profile of adjuvant trastuzumab-based treatment among women treated routine care.[105]

There are of course limitations associated with different statistical approaches. Firstly, there is the general issue of either unmeasured or unmeasurable variables. For analysis looking at treatment uptake, it was not possible to account for patient preference and clinical discussions when understanding variation in use. As such some level of selection bias and impact from unmeasured confounders will have been introduced. The use of risk adjustment and multilevel mixed-effects models will go some way to accounting for this aspect of bias in the data and cohort selection.

A further problem linked to confounders is the mechanisms for imputing missing information, and any differential missingness among patient subgroups. For methodological work, variables were categorical and so missing values were included as their own category because this was informative in understanding differences in recording. Within the papers looking at use of treatments in routine care, identification of patient groups was by

stage and receptor status. For receptor status, where information was missing, because of the way in which routine pathology data are typically returned from hospitals to NCRAS via automated feeds, missing information most likely meant that the pathology test to assess receptor status was not performed and so this information would not have been available to the hospital to inform treatment decisions; in this scenario the unknown receptor status will not have affected treatment. If this mechanism for data flows did not apply (for example where a hospital did not have the capabilities for automated feeds and instead pathology data were either returned in free text pathology reports or not routinely returned), it is unlikely that recording of information systematically differed according to the patient subgroup. Rather, where differences occur by patient subgroup, it is most likely to reflect systematic differences in the approach to staging and receptor status assessment. For example, anecdotally it is known that some hospitals have guidance that patients over a certain age don't have the HER2 status of their tumour assessed, or older, less fit patients may not be fully staged due to poor fitness for treatment. Missing data in other variables of interest was minimal.

The final paper looked at understanding the impact of treatment on survival outcomes. Endpoints such as disease recurrence were not routinely available within the data, with completeness of variables flagging local or distant recurrence being an issue. In a similar vein to this, factors of interest which would align with variables measured within RCTs were limited to those investigations and patient factors known within the secondary care setting and routinely collected in the cancer and hospital datasets. This might potentially limit the degree to which comparison with trial populations could highlight the full extent of differences with patients treated in routine care. This analysis used propensity scores (PS) for estimating the effect of treatment across patient subgroups, specifically using the inverse probability of treatment weighting method. Other methods were considered for this analysis, such as entropy balancing and instrumental variables analysis however, the PS method was chosen for several reasons, including (i) it was deemed most understandable to a non-statistical audience, (ii) within the routine data available there was no clear instrumental variable which would be a strong predictor of the treatment of interest whilst having no impact on the outcome (survival in this instance) except through treatment, which is a key requirement for this approach.

6.3. Implications

The work presented in this thesis has implications across two differing but complementary areas within healthcare. Firstly the work has highlighted both the immense value and potential uses of routine data, and has also shone a light on aspects that should be considered before using routine data for treatment evaluation. Secondly there are policy implications for the ongoing monitoring of treatment utilisation, safety and effectiveness in the patient population. In the context of those patients not well represented in RCTs, routine data provides an opportunity for them to be counted, to be understood and to have their access to life-prolonging treatments evaluated and improved.

The research presented in this thesis is novel within breast cancer, and has a major impact on our ongoing understanding of the benefits of routine national data and the use of oncological treatments among patients treated in routine care. The implications from this work are discussed in more detail within the next sections.

6.3.1. Routine data

The methodological work within this thesis, demonstrated the value of using multiple sources where capture may be limited by organisational and time factors, which impacted on data completeness. In the case of chemotherapy and trastuzumab these represented two treatment types where recording of both OPCS codes for treatment procurement and delivery enabled distinction of treatment administration in hospital admissions data. Recent guidelines on the recording of OPCS codes have moved to only OPCS codes for delivery being required, which will have an impact on the utility of this data source in identifying treatment administration where it is helpful to distinguish between treatment types.[124] Coding practices would benefit in this circumstance from codes being clearly delineated between different regimens. This is of increasing importance given the work in this thesis which highlighted the number of NICE-approved drugs becoming available since 2002. Implications for the routine data collection, coming out of this work comparing recording in the SACT and HES-APC datasets that would benefit most users (and future-proof the routine data available) would be for efforts to focus on ensuring there is clear reporting of drug details within the SACT dataset as the primary and most comprehensive source of data on systemic oncological treatment. The capabilities of the SACT dataset, in terms of what can be collected, rival any trials database on treatment provision, although data completeness of peripheral items such as regimen outcomes information (treatment delays, dose reductions, completion), currently limits the value of the SACT dataset as more than a basic drug recording tool. Adding to this that not all drugs recorded are eventually administered raises the question of whether a variable flagging this dynamic aspect of treatment would be of value, as something that would be useful for users of the data without having a large effect on data entry burden.

Another implication from this piece of work relates to the value of linked primary and secondary care data, in particular, the merits of electronic prescribing data. The second methodological paper demonstrated the value of having linked data from different healthcare settings to fully understand the patterns of treatment delivery and care received by patients. This was exemplified in the case of endocrine therapy, a major component of treatment for women with hormone receptor-positive invasive breast cancer, but for which comparison of recording across secondary care and primary care data sources highlighted systematic differences in recording by age and patient fitness. With much of treatment started in secondary care and continued within a primary care setting, primary care data were found to be essential in accurately capturing information on the number of patients having this as part of treatment. The implications of this work extend beyond both breast cancer and endocrine therapy.

Finally, scenarios where data recording is found to systematically differ according to patient age suggest a difference in the process of care for patients, determined by their age, rather than differential recording of information by age. For users of routine data, these differences in recording are crucial to be aware of in order to prevent drawing incorrect conclusions about equity of care. For example using secondary care data alone, one may infer that use of endocrine therapy varies according to age, although linking with information from primary care highlighted that use of endocrine therapy was high regardless of age, and so the age difference was likely to be reflective of differences in where and at what point treatment was initiated. This work also draws out the

importance of a collaborative approach to using routine data, with input from both analysts and clinicians being imperative to ensure that data can be understood and interpreted correctly.

There could be considerable benefits from data source cross-checks being carried out by the data controllers. Such checks could in many ways be automated, with rules assigned based on clinical knowledge of practice. Validations of data completeness would then not need to prevent data release but could be communicated with the release of data as known issues to be aware of.

6.3.2. Clinical practice & policy

Ongoing monitoring of treatment uptake

The second part of the work presented in this thesis, looking at use of NICE-approved drugs for invasive breast cancer in routine care and then focusing in on use of adjuvant trastuzumab-based treatment for EIBC, highlights the importance of monitoring the translation of RCT findings and NICE approvals into equitable use among patients as well as displaying the value of routine data in this process.

Firstly, the variation in utilisation of approved drugs highlights a need for both clear guidance on treatment use, to ensure consistency in use that is not dependent on factors which are not linked to treatment indication, and clear information on the use of treatments in those patient populations poorly represented in the evidence base for treatment.

Secondly, where there are multiple drug options for patients, national guidance should acknowledge this and provide information to enable clear routes for treatment decision-making.

Evidence of treatment outcomes in routine care

The work presenting a thorough breakdown of severe acute toxicity events provides valuable information on the prevalence of treatment-related toxicity among patients receiving adjuvant trastuzumab-based treatment in routine care. This is crucial information needed by clinicians and patients in conversations around treatment options, as well as being informative for healthcare policy makers and regulators.

Further to this, analysis of survival outcomes following initiation of adjuvant trastuzumab-based treatment found no evidence of differential effect by age, which has direct implications for guidance wording and the treatment decision-making process. National quality standards for breast cancer referred to the principle that treatment decisions should be made “irrespective of age” in the initial briefing document, however this did not make it to the final published version.[125,126] My work would suggest that treatment decisions should not be determined by chronological age, but rather the patient’s additional health conditions and fitness should inform whether treatment is contraindicated. Chronological age is not analogous to fitness or the predictor either of life expectancy or ability to tolerate treatment it perhaps once was, however being aged 70 or older appears to be contributing to a reduced likelihood of treatment. A contributing factor to this may well be a patients’ perception of the impact of treatment on their quality of life compared with the impact of the cancer. In this situation being able to provide clear and understandable information to patients on this among patients like

them treated in routine care, and not just those in trials, is vital to encourage use of beneficial treatments and provide context which is familiar to them. Guidance and local policy should reflect this variation in practice and make attempts to discourage treatment decision-making based on chronological age.

Beyond implications for policy which directly refers to treatment provision for patients, the work carried out for this thesis has implications for healthcare policy where improvement goals and areas for research are defined. There is a need for evidence of treatment safety profiles and effectiveness in routine care, for which this research has provided evidence in the case of adjuvant trastuzumab-based treatment for HER2-positive EIBC. However, there are many scenarios beyond this, whether it be for more advanced cancer or other treatments, which would benefit from evidence production and clear guidance around implications for patients not represented in trials. The work in this thesis has demonstrated that routine data is of value to meet such needs.

6.3.3. Future research

Areas which would benefit from future research are touched on in the previous sections. More detail is given here for those elements of work directly resulting from the work presented in this thesis.

Recording of treatment in routine data sources

The methodological work presented in this thesis focused on chemotherapy and trastuzumab, as the most well-established and commonly used treatments for invasive breast cancer. However there have been considerable developments in the treatment options available for breast cancer, particularly in targeted therapies. In light of the areas of treatment recording in routine data highlighted within the methodological work carried out, future research would benefit from extending this work to those more recently available treatments, looking at validating recording within hospital administrative datasets and other relevant data sources, to understand the completion rate of these within the SACT dataset, and whether these are of value for such drugs and can highlight where data completeness requires focused attention in the SACT dataset.

This aspect of methodological work focused on treatments given in the context of primary surgery for early breast cancer, and so it would be of interest to look both at whether the SACT dataset has better capture of treatments given where they are the primary treatment and also how well the SACT dataset captures treatments given as part of subsequent treatment for cancer such as in the case of recurrence or as part of further lines of treatment.

Given the cohort of women available for this PhD were all aged 50 years or older, future research should look to understand the recording of treatments in routine data sources among women with breast cancer younger than 50 years of age.

Treatment utilisation and outcomes in routine care

With the most recent NICE guidance for advanced breast cancer published in 2009 and many treatments approved since, work to understand uptake, use and associated outcomes focusing on this patient group would be beneficial.[52] Additionally more regular monitoring of treatment use, safety and survival outcomes in routine

care, would be advantageous in providing more up-to-date, generalisable information for clinical decision-making and patient experience. Understanding variation in patient care, whether geographically or based on patient factors, is particularly crucial as the number of drug options increases and as developments in other disease areas contribute to new complex treatment patterns and interactions.[127]

The work looking at use of NICE-approved drugs in routine care, highlighted that in the case of CDK 4/6 inhibitors for hormone receptor-positive, HER2-negative invasive breast cancer there are several similar treatment options. Additionally for HER2-positive breast cancer there are several options. Future research on the sequencing of drug treatments where more than one are options as well as to provide clear information on the benefit and risk profiles of each treatment outside of a trial setting would be incredibly beneficial for informing discussions in clinic. For patients who are older or less fit such research could provide valuable information on the risk-benefit considerations specific to them.

Work presented in this thesis looked at both safety and survival outcomes following use of adjuvant trastuzumab-based treatment for HER2-positive EIBC in routine care. With the experience of adverse events contributing to treatment decision discussions, considering the limitations of routine administrative data in understanding those events experienced by patients but not resulting in an overnight hospital admission, future research would benefit from looking at ways of understanding these types of treatment-related toxicity events. Use of systems enabling patient self-report of outcomes including symptoms around treatment would be valuable in providing information on this important aspect of care.[128-130] Additionally, work to understand the ability of routine data to provide information on intermediate survival outcomes, such as recurrence would be incredibly beneficial.[131-134]

Finally, future research should look at addressing gaps in the evidence base for use of systemic oncological treatments, particularly where there may be grey areas evidenced by variation in utilisation. This PhD focused attention on patient age, and fitness, but there are other factors, such as ethnicity and socioeconomic deprivation status, where evidence may be lacking and which independently determine the likelihood of a patient receiving treatment.[135-137]

7. Conclusion

Overall, the research presented in this thesis has demonstrated that routine data are of value in understanding real-world treatment use and associated outcomes. Additionally they are valuable in overcoming some of the limitations of RCTs where generalisability is an issue. The work has highlighted systematic differences in the recording of oncological treatments within the SACT dataset, which particularly affected its utility within older patients. Work to ensure routine data are complete and accurate is vital in making sure these data can be of the utmost value for use in future health services research and in studies evaluating treatment effectiveness among patients treated in routine care. Specifically, efforts should be made to ensure full information is recorded in the SACT dataset for all patients receiving treatment as this is a primary data collection tool for details on treatment and is incredibly valuable for future studies looking at utilisation of oncological treatments, along with aspects of treatment such as adherence and clinical management of patients experiencing treatment-related toxicity.

Routine data can be used to estimate outcomes following the initiation of adjuvant systemic oncological treatment with respect to safety and survival. This information is beneficial when discussing treatment options with patients. In the case of adjuvant trastuzumab-based treatment, work within this thesis highlighted variation in use by age, independent of other factors such as patient fitness. Where trastuzumab was part of a patient's treatment however severe toxicity was found to be well managed among older patients. Additionally, after accounting for systematic differences in patient and tumour characteristics between those patients receiving trastuzumab and those who did not effectiveness was not observed to differ according to patient age or fitness.

With a clinical need for information on treatment for older patients and a keenness to understand real-world treatment use and outcomes, routine data provide a means to satisfy both needs, providing evidence which not only complement trial findings but also enable monitoring of the risks and benefits associated with treatment in clinical practice.

8. References

1. Elting LS, Cooksley C, Bekele BN, Frumovitz M, Avritscher EB, Sun C, Bodurka DC. Generalizability of cancer clinical trial results: prognostic differences between participants and nonparticipants. Cancer. 2006;106(11):2452-8.

2. Hutchins LF, Unger JM, Crowley JJ, Coltman CA, Jr., Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. The New England journal of medicine. 1999;341(27):2061-7.

3. Kemeny MM, Peterson BL, Kornblith AB, Muss HB, Wheeler J, Levine E, et al. Barriers to clinical trial participation by older women with breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2003;21(12):2268-75.

4. Townsley CA, Selby R, Siu LL. Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2005;23(13):3112-24.

5. Ridda I, MacIntyre CR, Lindley RI, Tan TC. Difficulties in recruiting older people in clinical trials: an examination of barriers and solutions. Vaccine. 2010;28(4):901-6.

6. Kazmierska J. Do we protect or discriminate? Representation of senior adults in clinical trials. Rep Pract Oncol Radiother. 2012;18(1):6-10.

7. Ludmir EB, Mainwaring W, Lin TA, Miller AB, Jethanandani A, Espinoza AF, et al. Factors Associated With Age Disparities Among Cancer Clinical Trial Participants. JAMA oncology. 2019;5(12):1769-73.

8. Forsat ND, Palmowski A, Palmowski Y, Boers M, Buttgereit F. Recruitment and Retention of Older People in Clinical Research: A Systematic Literature Review. J Am Geriatr Soc. 2020;68(12):2955-63.

9. van Marum RJ. Underrepresentation of the elderly in clinical trials, time for action. Br J Clin Pharmacol. 2020;86(10):2014-6.

10. Payne JK, Hendrix CC. Clinical trial recruitment challenges with older adults with cancer. Appl Nurs Res. 2010;23(4):233-7.

11. Macmillan Cancer Support. The Age Old Excuse: The Under Treatment of Older Cancer Patients. 2012. Available from: https://www.macmillan.org.uk/documents/getinvolved/campaigns/ageoldexcuse/ageoldexcusereport-macmillancancersupport.pdf. (accessed 17/08/2019)

12. VanderWalde A, Hurria A. Early breast cancer in the older woman. Clin Geriatr Med. 2012;28(1):73-91.

13. Shachar SS, Hurria A, Muss HB. Breast Cancer in Women Older Than 80 Years. J Oncol Pract. 2016;12(2):123-32.

14. Ring A, Reed M, Leonard R, Kunkler I, Muss H, Wildiers H, et al. The treatment of early breast cancer in women over the age of 70. British journal of cancer. 2011;105(2):189-93.

15. Enger SM, Thwin SS, Buist DS, Field T, Frost F, Geiger AM, et al. Breast cancer treatment of older women in integrated health care settings. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2006;24(27):4377-83.
16. Lavelle K, Todd C, Moran A, Howell A, Bundred N, Campbell M. Non-standard management of breast cancer increases with age in the UK: a population based cohort of women > or =65 years. British journal of cancer. 2007;96(8):1197-203.
17. Allemani C, Storm H, Voogd AC, Holli K, Izarzugaza I, Torrella-Ramos A, et al. Variation in 'standard care' for breast cancer across Europe: a EUROCARE-3 high resolution study. Eur J Cancer. 2010;46(9):1528-36.
18. Derks MGM, Bastiaannet E, Kiderlen M, Hilling DE, Boelens PG, Walsh PM, et al. Variation in treatment and survival of older patients with non-metastatic breast cancer in five European countries: a population-based cohort study from the EURECCA Breast Cancer Group. British journal of cancer. 2018;119(1):121-9.
19. LeMasters T, Madhavan SS, Sambamoorthi U, Hazard-Jenkins HW, Kelly KM, Long D. Receipt of Guideline-Concordant Care Among Older Women With Stage I-III Breast Cancer: A Population-Based Study. J Natl Compr Canc Netw. 2018;16(6):703-10.
20. Exman P, Burstein HJ. How Old is Too Old? Breast Cancer Treatment in Octogenarians. Ann Surg Oncol. 2018;25(6):1458-60.
21. Reeder-Hayes K, Peacock Hinton S, Meng K, Carey LA, Dusetzina SB. Disparities in Use of Human Epidermal Growth Hormone Receptor 2-Targeted Therapy for Early-Stage Breast Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2016;34(17):2003-9.
22. Ring A. The influences of age and co-morbidities on treatment decisions for patients with HER2-positive early breast cancer. Critical reviews in oncology/hematology. 2010;76(2):127-32.
23. Schonberg MA, Marcantonio ER, Li D, Silliman RA, Ngo L, McCarthy EP. Breast cancer among the oldest old: tumor characteristics, treatment choices, and survival. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(12):2038-45.
24. Department of Health. The impact of patient age on clinical decision-making in oncology. 2012. Available from: https://assets.publishing.service.gov.uk/media/5a75043840f0b6399b2affdb/dh_133095.pdf (accessed 17/08/2019)
25. Foot C, Harrison T. How to improve cancer survival: Explaining England's relatively poor rates. 2011. Available from: www.kingsfund.org.uk/sites/default/files/How-to-improve-cancer-survival-Explaining-England-poor-rates-Kings-Fund-June-2011.pdf. (acessed 17/08/2019)
26. CRUK. Advancing Years: Treating and Caring for an Ageing Population. 2018. Available from: www.cancerresearchuk.org/about-us/we-develop-policy/our-policy-on-access-to-cancer-treatments/treating-and-caring-for-an-ageing-population (acessed 17/08/2019)

27. Bate A, Juniper J, Lawton AM, Thwaites RM. Designing and incorporating a real world data approach to international drug development and use: what the UK offers. Drug discovery today. 2016;21(3):400-5.
28. Bright CJ, Lawton S, Benson S, Bomb M, Dodwell D, Henson KE, et al. Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) dataset. Int J Epidemiol. 2020;49(1):15-l.
29. National Audit of Breast Cancer in Older Patients (NABCOP). NABCOP 2019 Annual Report. 2019. Available from: https://www.nabcop.org.uk/reports/nabcop-2019-annual-report/. (acessed 09/05/2019)
30. Lim E, Metzger-Filho O, Winer EP. The natural history of hormone receptor-positive breast cancer. Oncology (Williston Park). 2012;26(8):688-94, 96.
31. Loibl S, Gianni L. HER2-positive breast cancer. Lancet. 2017;389(10087):2415-29.
32. O'Connor T, Shinde A, Doan C, Katheria V, Hurria A. Managing breast cancer in the older patient. Clin Adv Hematol Oncol. 2013;11(6):341-7.
33. Breast Cancer Now. Breast cancer facts and statistics. 2022. Available from: https://breastcancernow.org/about-us/media/facts-statistics. (acessed 12/05/2022)
34. Cancer Research UK. Cancer incidence for common cancers. 2020. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared#heading-Two. (acessed 12/05/2022)
35. Cancer Research UK. Breast cancer statistics. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer. (acessed 18/10/2023)
36. ONS. Cancer registration statistics, England. 2017.
37. Maddams J, Utley M, Moller H. Projections of cancer prevalence in the United Kingdom, 2010-2040. British journal of cancer. 2012;107(7):1195-202.
38. Smittenaar CR, Petersen KA, Stewart K, Moitt N. Cancer incidence and mortality projections in the UK until 2035. British journal of cancer. 2016;115(9):1147-55.
39. Harnett A, Smallwood J, Titshall V, Champion A, Guideline Development G. Diagnosis and treatment of early breast cancer, including locally advanced disease--summary of NICE guidance. BMJ. 2009;338:b438.
40. Moo TA, Sanford R, Dang C, Morrow M. Overview of Breast Cancer Therapy. PET Clin. 2018;13(3):339-54.
41. World Health Organization. Breast cancer. 2023. Available from: https://www.who.int/news-room/fact-sheets/detail/breast-cancer. (acessed 12/10/2023)
42. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134-50.
43. NICE. Early and locally advanced breast cancer: diagnosis and management. 2018. Available from: www.nice.org.uk/guidance/ng101.
44. Ferguson T, Wilcken N, Vagg R, Ghersi D, Nowak AK. Taxanes for adjuvant treatment of early breast cancer. The Cochrane database of systematic reviews. 2007(4):CD004421.

45. Early Breast Cancer Trialists' Collaborative G. Anthracycline-containing and taxane-containing chemotherapy for early-stage operable breast cancer: a patient-level meta-analysis of 100 000 women from 86 randomised trials. Lancet. 2023;401(10384):1277-92.
46. Early Breast Cancer Trialists' Collaborative G. Multi-agent chemotherapy for early breast cancer. The Cochrane database of systematic reviews. 2002(1):CD000487.
47. Early Breast Cancer Trialists' Collaborative G, Peto R, Davies C, Godwin J, Gray R, Pan HC, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. Lancet. 2012;379(9814):432-44.
48. Kerr AJ, Dodwell D, McGale P, Holt F, Duane F, Mannu G, et al. Adjuvant and neoadjuvant breast cancer treatments: A systematic review of their effects on mortality. Cancer Treat Rev. 2022;105:102375.
49. Verrill M. Chemotherapy for early-stage breast cancer: a brief history. British journal of cancer. 2009;101 Suppl 1(Suppl 1):S2-5.
50. Willson ML, Burke L, Ferguson T, Ghersi D, Nowak AK, Wilcken N. Taxanes for adjuvant treatment of early breast cancer. The Cochrane database of systematic reviews. 2019;9:CD004421.
51. Aebi S, Karlsson P, Wapnir IL. Locally advanced breast cancer. Breast. 2022;62 Suppl 1(Suppl 1):S58-S62.
52. NICE. Advanced breast cancer: diagnosis and treatment. 2009. Available from: https://www.nice.org.uk/guidance/cg81.
53. NICE. Guidance on the use of trastuzumab for the treatment of advanced breast cancer 2002 [Available from: https://www.nice.org.uk/guidance/ta34.
54. NICE. Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. 2006. Available from: https://www.nice.org.uk/guidance/ta107.
55. Moja L, Tagliabue L, Balduzzi S, Parmelli E, Pistotti V, Guarneri V, D'Amico R. Trastuzumab containing regimens for early breast cancer. The Cochrane database of systematic reviews. 2012(4):CD006243.
56. Balduzzi S, Mantarro S, Guarneri V, Tagliabue L, Pistotti V, Moja L, D'Amico R. Trastuzumab-containing regimens for metastatic breast cancer. The Cochrane database of systematic reviews. 2014;2014(6):CD006242.
57. NHS Digital. Cancer Outcomes and Services Data set (COSD). Available from: https://digital.nhs.uk/ndrs/data/data-sets/cosd. (accessed 03/07/2023)
58. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Int J Epidemiol. 2017;46(4):1093-i.
59. Henson KE, Brock R, Shand B, Coupland VH, Elliss-Brookes L, Lyratzopoulos G, et al. Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England. BMJ open. 2018;8(7):e020980.

60. Emanuel G, Henson KE, Broggio J, Charman J, Horgan K, Dodwell D, Darby SC. Endocrine therapy in the years following a diagnosis of breast cancer: A proof of concept study using the primary care prescription database linked to cancer registration data. Cancer Epidemiol. 2019;61:185-9.
61. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004;58(4):265-71.
62. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet. 2002;359(9300):57-61.
63. Steckler A, McLeroy KR. The importance of external validity. Am J Public Health. 2008;98(1):9-10.
64. Jones DS, Podolsky SH. The history and fate of the gold standard. Lancet. 2015;385(9977):1502-3.
65. Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the Gold Standard--Lessons from the History of RCTs. The New England journal of medicine. 2016;374(22):2175-81.
66. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand. 2018;97(4):380-7.
67. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601-5.
68. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet. 2005;365(9453):82-93.
69. Lewis JH, Kilgore ML, Goldman DP, Trimble EL, Kaplan R, Montello MJ, et al. Participation of patients 65 years of age or older in cancer clinical trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2003;21(7):1383-9.
70. Markopoulos C, van de Water W. Older patients with breast cancer: is there bias in the treatment they receive? Ther Adv Med Oncol. 2012;4(6):321-7.
71. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol. 2003;3:28.
72. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. AHRQ Technical Reviews. Rockville (MD)2006.
73. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-World Evidence - What Is It and What Can It Tell Us? The New England journal of medicine. 2016;375(23):2293-7.
74. Aapro MS, Köhne C-H, Cohen HJ, Extermann M. Never Too Old? Age Should Not Be a Barrier to Enrollment in Cancer Clinical Trials. The Oncologist. 2005;10(3):198-204.

75. Leonard R, Ballinger R, Cameron D, Ellis P, Fallowfield L, Gosney M, et al. Adjuvant chemotherapy in older women (ACTION) study - what did we learn from the pilot phase? British journal of cancer. 2011;105(9):1260-6.
76. Crivellari D, Gray KP, Dellapasqua S, Puglisi F, Ribi K, Price KN, et al. Adjuvant pegylated liposomal doxorubicin for older women with endocrine nonresponsive breast cancer who are NOT suitable for a "standard chemotherapy regimen": the CASA randomized trial. Breast. 2013;22(2):130-7.
77. Buttgereit T, Palmowski A, Forsat N, Boers M, Witham MD, Rodondi N, et al. Barriers and potential solutions in the recruitment and retention of older patients in clinical trials-lessons learned from six large multicentre randomized controlled trials. Age Ageing. 2021;50(6):1988-96.
78. Todd A, Martin C, Morgan J, Herbert E, Bradburn M, Burton M, et al. Age specific recruitment and retention to a large multicentre observational breast cancer trial in older women: The Age Gap Trial. Journal of geriatric oncology. 2021;12(5):714-23.
79. National Audit of Breast Cancer in Older Patients (NABCOP). NABCOP 2020 Annual Report. 2020. Available from: www.nabcop.org.uk/reports/nabcop-2020-annual-report/ (accessed 12/05/2022)
80. National Audit of Breast Cancer in Older Patients (NABCOP). NABCOP 2022 Annual Report. 2022. Available from: https://www.nabcop.org.uk/reports/nabcop-2022-annual-report/ (accessed 12/05/2022)
81. Klungel OH, Martens EP, Psaty BM, Grobbee DE, Sullivan SD, Stricker BH, et al. Methods to assess intended effects of drug treatment in observational studies are reviewed. Journal of clinical epidemiology. 2004;57(12):1223-31.
82. Lu CY. Observational studies: a review of study designs, challenges and strategies to reduce confounding. Int J Clin Pract. 2009;63(5):691-7.
83. Maringe C, Benitez Majano S, Exarchakou A, Smith M, Rachet B, Belot A, Leyrat C. Reflections on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. Int J Epidemiol. 2020.
84. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
85. Stuart BL, Grebel LE, Butler CC, Hood K, Verheij TJM, Little P. Comparison between treatment effects in a randomised controlled trial and an observational study using propensity scores in primary care. The British journal of general practice : the journal of the Royal College of General Practitioners. 2017;67(662):e643-e9.
86. Petito LC, Garcia-Albeniz X, Logan RW, Howlader N, Mariotto AB, Dahabreh IJ, Hernan MA. Estimates of Overall Survival in Patients With Cancer Receiving Different Treatment Regimens: Emulating Hypothetical Target Trials in the Surveillance, Epidemiology, and End Results (SEER)-Medicare Linked Database. JAMA Netw Open. 2020;3(3):e200452.

87. D'Agostino RB, Jr., D'Agostino RB, Sr. Estimating treatment effects using observational data. JAMA. 2007;297(3):314-6.
88. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55.
89. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.
90. Williamson E, Morley R, Lucas A, Carpenter J. Propensity scores: from naive enthusiasm to intuitive understanding. Stat Methods Med Res. 2012;21(3):273-93.
91. Austin PC. A comparison of 12 algorithms for matching on the propensity score. Statistics in medicine. 2014;33(6):1057-69.
92. Rosenbaum PR. Covariance Adjustment in Randomized Experiments and Observational Studies. Statistical Science. 2002;17(3):286-327, 42.
93. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, et al. Propensity score analysis with partially observed covariates: How should multiple imputation be used? Stat Methods Med Res. 2019;28(1):3-19.
94. Lu B, Thomson S, Blommaert S, Tadrous M, Earle CC, Chan KKW. Use of Instrumental Variable Analyses for Evaluating Comparative Effectiveness in Empirical Applications of Oncology: A Systematic Review. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2023;41(13):2362-71.
95. Hainmueller J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis. 2012;20(1):25–46.
96. Swanson SA. Instrumental Variable Analyses in Pharmacoepidemiology: What Target Trials Do We Emulate? Curr Epidemiol Rep. 2017;4(4):281-7.
97. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. Statistics in medicine. 2014;33(13):2297-340.
98. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006;17(4):360-72.
99. Martinussen T, Vansteelandt S, Tchetgen Tchetgen EJ, Zucker DM. Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. Biometrics. 2017;73(4):1140-9.
100. NICE. NICE real-world evidence framework. 2022. Available from: https://www.nice.org.uk/corporate/ecd9/chapter/overview. (accessed 03/07/2023)
101. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. CMAJ. 2016;188(8):E158-E64.
102. McDonald L, Sammon C, Carroll R, Harish A, C AS, Tyas D, Ramagopalan SV. Consistency of recording of chemotherapy cycles in the National Cancer Registration and Analysis Service Systemic Anti-Cancer Therapy database and the Hospital Episode Statistics Admitted Patient Care database. Future Oncol. 2020;16(3):4455-60.

103. Boyle JM, Kuryba A, Braun MS, Aggarwal A, van der Meulen J, Cowling TE, Walker K. Validity of chemotherapy information derived from routinely collected healthcare data: A national cohort study of colon cancer patients. Cancer Epidemiol. 2021;73:101971.
104. Krzyzanowska MK, Enright K, Moineddin R, Yun L, Powis M, Ghannam M, Grunfeld E. Can Chemotherapy-Related Acute Care Visits Be Accurately Identified in Administrative Data? J Oncol Pract. 2018;14(1):e51-e8.
105. Boyle JM, Cowling TE, Kuryba A, Fearnhead NS, van der Meulen J, Braun MS, et al. Development and validation of a coding framework to identify severe acute toxicity from systemic anti-cancer therapy using hospital administrative data. Cancer Epidemiol. 2022;77:102096.
106. Armitage JN, van der Meulen JH. Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score. The British journal of surgery. 2010;97(5):772-81.
107. Jauhari Y, Gannon MR, Dodwell D, Horgan K, Clements K, Medina J, 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(5):e035395.
108. Henson KE, Elliss-Brookes L, Coupland VH, Payne E, Vernon S, Rous B, Rashbass J. Data Resource Profile: National Cancer Registration Dataset in England. Int J Epidemiol. 2020;49(1):16-h.
109. Moller H, Richards S, Hanchett N, Riaz SP, Luchtenborg M, Holmberg L, Robinson D. Completeness of case ascertainment and survival time error in English cancer registries: impact on 1-year survival estimates. British journal of cancer. 2011;105(1):170-6.
110. National Cancer Registration and Analysis Service. Matching SACT to Cancer Waiting Times Data. 2014. Available from: http://www.ncin.org.uk/publications/data_briefings/sact_cwt. (accessed 12/05/2022)
111. StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC2021.
112. Greenland S. Principles of multilevel modelling. Int J Epidemiol. 2000;29(1):158-67.
113. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Statistics in medicine. 2017;36(27):4391-400.
114. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association. 1999;94(446):496-509.
115. Dickerman BA, Garcia-Albeniz X, Logan RW, Denaxas S, Hernan MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019;25(10):1601-6.
116. Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. CMAJ. 2017;189(33):E1054-E5.
117. Health and Social Care Information Centre. Chemotherapy Regimens Clinical Coding Standards and Guidance OPCS-4. 2014. Available from:

http://www.nwisinformationstandards.wales.nhs.uk/sitesplus/documents/299/ChemoRegClinCodStanGuidOPCS-4v1.0.pdf. (accessed 12/05/2022)
118. Morgan A, Sinnott SJ, Smeeth L, Minassian C, Quint J. Concordance in the recording of stroke across UK primary and secondary care datasets: a population-based cohort study. BJGP Open. 2021;5(2).
119. Williams R, van Staa TP, Gallagher AM, Hammad T, Leufkens HGM, de Vries F. Cancer recording in patients with and without type 2 diabetes in the Clinical Practice Research Datalink primary care data and linked hospital admission data: a cohort study. BMJ open. 2018;8(5):e020827.
120. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827-36.
121. Crooks CJ, West J, Card TR. A comparison of the recording of comorbidity in primary and secondary care by using the Charlson Index to predict short-term and long-term survival in a routine linked data cohort. BMJ open. 2015;5(6):e007974.
122. Rockenschaub P, Nguyen V, Aldridge RW, Acosta D, Garcia-Gomez JM, Saez C. Data-driven discovery of changes in clinical code usage over time: a case-study on changes in cardiovascular disease recording in two English electronic health records databases (2001-2015). BMJ open. 2020;10(2):e034396.
123. Januel JM, Luthi JC, Quan H, Borst F, Taffe P, Ghali WA, Burnand B. Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. BMC health services research. 2011;11:194.
124. NHS England. National Clinical Coding Standards OPCS-4; Accurate data for quality information. 2023 Available from: https://classbrowser.nhs.uk/ref_books/OPCS-4.10_NCCS-2023.pdf.
125. National Institute for Health and Care Excellence. Health and social care directorate. Quality standards and indicators. Briefing paper. 2015. Available from: https://www.nice.org.uk/guidance/qs12/documents/briefing-paper. (accessed 03/07/2023)
126. National Institute for Health and Care Excellence. Breast Cancer (QS12). 2016. Available from: https://www.nice.org.uk/guidance/qs12. (accessed 03/07/2023)
127. Gannon MR, Dodwell D, Miller K, Horgan K, Clements K, Medina J, et al. Change in the Use of Fractionation in Radiotherapy Used for Early Breast Cancer at the Start of the COVID-19 Pandemic: A Population-Based Cohort Study of Older Women in England and Wales. Clin Oncol (R Coll Radiol). 2022;34(9):e400-e9.
128. Holch P, Warrington L, Bamforth LCA, Keding A, Ziegler LE, Absolom K, et al. Development of an integrated electronic platform for patient self-report and management of adverse events during cancer treatment. Annals of oncology : official journal of the European Society for Medical Oncology. 2017;28(9):2305-11.
129. Stamp E, Clarke G, Wright P, Velikova G, Crossfield SSR, Zucker K, et al. Collection of cancer Patient Reported Outcome Measures (PROMS) to link with primary and secondary

electronic care records to understand and improve long term cancer outcomes: A protocol paper. PloS one. 2022;17(4):e0266804.
130. Velikova G, Absolom K, Hewison J, Holch P, Warrington L, Avery K, et al. Electronic self-reporting of adverse events for patients undergoing cancer treatment: the eRAPID research programme including two RCTs. Programme Grants for Applied Research. Southampton (UK)2022.
131. Chubak J, Yu O, Pocobelli G, Lamerato L, Webster J, Prout MN, et al. Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. J Natl Cancer Inst. 2012;104(12):931-40.
132. Pedersen RN, Esen BO, Mellemkjaer L, Christiansen P, Ejlertsen B, Lash TL, et al. The Incidence of Breast Cancer Recurrence 10-32 Years After Primary Diagnosis. J Natl Cancer Inst. 2022;114(3):391-9.
133. Izci H, Tambuyzer T, Tuand K, Depoorter V, Laenen A, Wildiers H, et al. A Systematic Review of Estimating Breast Cancer Recurrence at the Population Level With Administrative Data. J Natl Cancer Inst. 2020;112(10):979-88.
134. Earle CC, Nattinger AB, Potosky AL, Lang K, Mallick R, Berger M, Warren JL. Identifying cancer relapse using SEER-Medicare data. Med Care. 2002;40(8 Suppl):IV-75-81.
135. Morimoto L, Coalson J, Mowat F, O'Malley C. Factors affecting receipt of chemotherapy in women with breast cancer. Int J Womens Health. 2010;2:107-22.
136. Gathani T, Chaudhry A, Chagla L, Chopra S, Copson E, Purushotham A, et al. Ethnicity and breast cancer in the UK: Where are we now? European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2021;47(12):2978-81.
137. Woods LM, Rachet B, Morris M, Bhaskaran K, Coleman MP. Are socio-economic inequalities in breast cancer survival explained by peri-diagnostic factors? BMC cancer. 2021;21(1):485.

9. Appendices

9.1. Appendix 1 – Confirmation of ethics approval

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
United Kingdom
Switchboard: +44 (0)20 7636 8636

www.lshtm.ac.uk

Observational / Interventions Research Ethics Committee

Mrs Melissa Gannon
LSHTM

5 July 2021

Dear Mrs Melissa Gannon

Study Title: Use of real world data to generate evidence on the effectiveness of oncological treatment provision in older/underrepresented populations of women diagnosed with breast cancer.

LSHTM Ethics Ref: 22508

Thank you for responding to the Observational Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Conditions of the favourable opinion

Approval is dependent on local ethical approval having been received, where relevant.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document Type	File Name	Date	Version
Local Approval	16.CAG.0079 final approval re-issued 19.09.16	19/09/2016	1
Investigator CV	M Gannon CV 2020	05/01/2021	1
Other	Gannon, M training record 26Apr21	26/04/2021	1
Protocol / Proposal	Gannon_ResearchProposal_PostUpgrading Version10_05_2021	10/05/2021	1
Covering Letter	LSHTM Ethics Clarifications - M Gannon 250621	25/06/2021	1

After ethical review

The Chief Investigator (CI) or delegate is responsible for informing the ethics committee of any subsequent changes to the application. These must be submitted to the Committee for review using an Amendment form. Amendments must not be initiated before receipt of written favourable opinion from the committee.

The CI or delegate is also required to notify the ethics committee of any protocol violations and/or Suspected Unexpected Serious Adverse Reactions (SUSARs) which occur during the project by submitting a Serious Adverse Event form.

An annual report should be submitted to the committee using an Annual Report form on the anniversary of the approval of the study during the lifetime of the study.

At the end of the study, the CI or delegate must notify the committee using an End of Study form.

All aforementioned forms are available on the ethics online applications website and can only be submitted to the committee via the website at: http://leo.lshtm.ac.uk

Additional information is available at: www.lshtm.ac.uk/ethics

Yours sincerely,

Professor Jimmy Whitworth
Chair

9.2. Appendix 2 – Supplementary material for Research Paper 1

Table A1. OPCS-4 codes for records of chemotherapy within HES-APC

OPCS-4 code	Classification
Drug procurement	
X701	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 1
X702	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 2
X703	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 3
X704	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 4
X705	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 5
X708	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X709	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X711	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 6
X712	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 7
X713	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 8
X714	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 9
X715	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 10
X718	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
X719	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
Drug delivery	
X721	Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first attendance
X722	Delivery of complex parenteral chemotherapy for neoplasm at first attendance
X723	Delivery of simple parenteral chemotherapy for neoplasm at first attendance
X724	Delivery of subsequent element of cycle of chemotherapy for neoplasm
X728	Other specified delivery of chemotherapy for neoplasm
X729	Unspecified delivery of chemotherapy for neoplasm
X731	Delivery of exclusively oral chemotherapy for neoplasm
X738	Other specified delivery of oral chemotherapy for neoplasm
X739	Unspecified delivery of oral chemotherapy for neoplasm
X748	Other specified other chemotherapy drugs
X749	Unspecified other chemotherapy drugs
Administration route	
X352	Intravenous chemotherapy
X373	Intramuscular chemotherapy
X384	Subcutaneous chemotherapy

Table A2. ICD-10 codes for records of chemotherapy within HES

ICD-10 code	Classification
Z082	Follow-up exam after chemotherapy for malignant neoplasm
Z292	Other prophylactic chemotherapy
Z511	Chemotherapy session for neoplasm
Z512	Other chemotherapy
Z542	Convalescence following chemotherapy

Figure A1: Details of patient selection from women aged 50 and over, diagnosed with invasive breast cancer in a NHS trust in England, between January 2014 and December 2019.

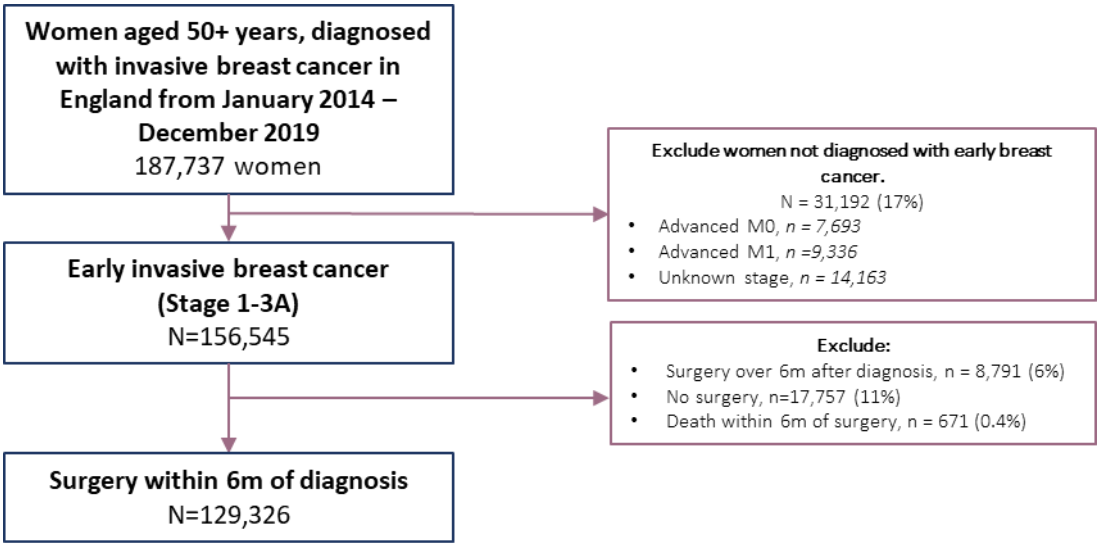
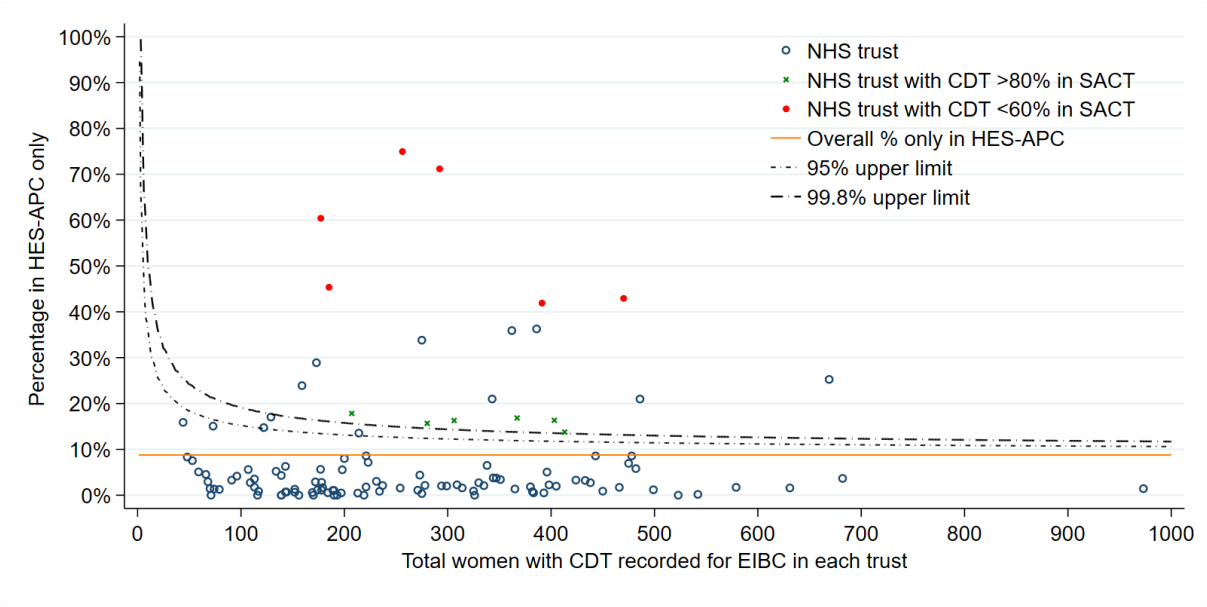
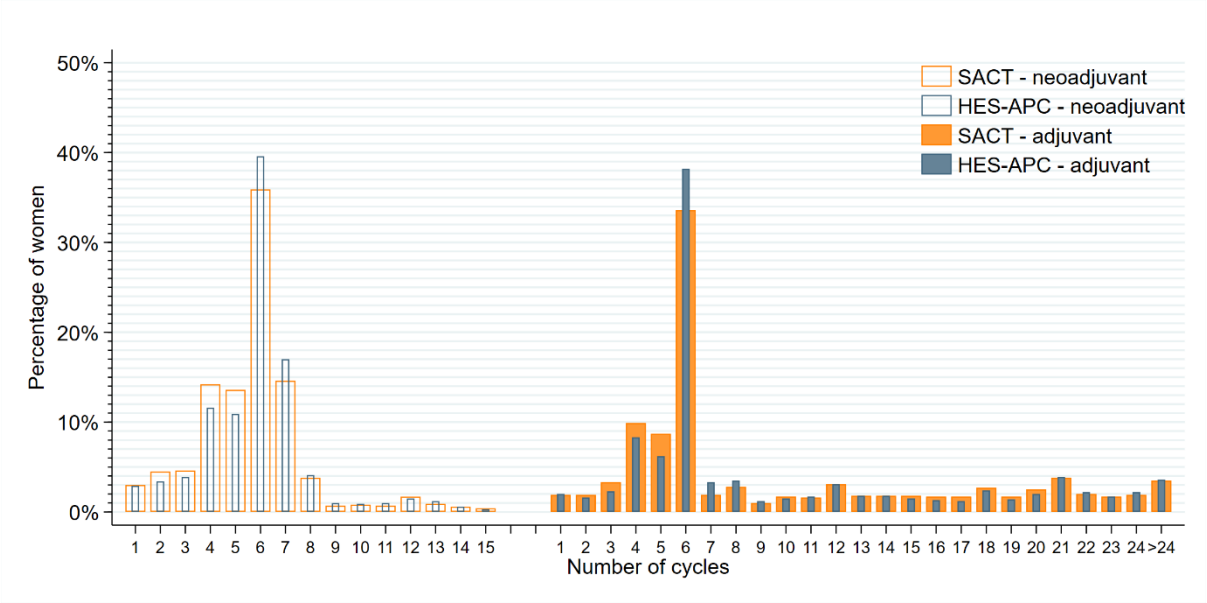


Figure A2. Funnel plot showing the percentage of CDT captured in HES-APC only, among women receiving CDT for early invasive breast cancer, by diagnosing NHS trust.



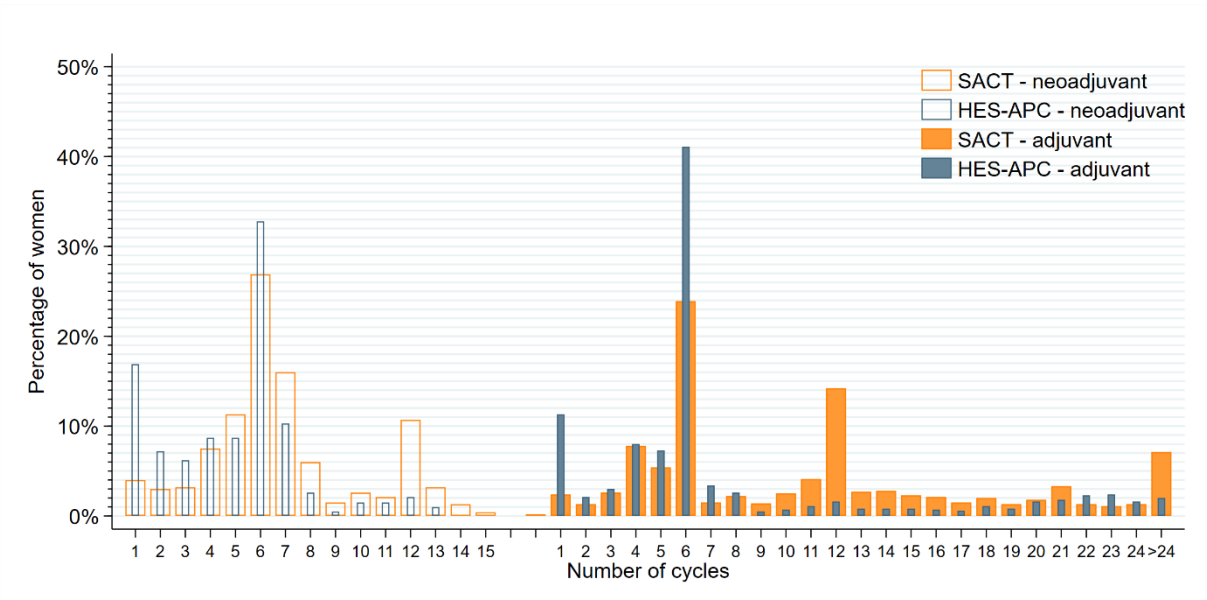
Key: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data.

Figure A3. Distribution of total numbers of cycles recorded in each source, among women with CDT for early invasive breast cancer recorded in both datasets, by CDT setting.



Key: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data.

Figure A4. Distribution of total numbers of cycles recorded in each source, among women with CDT for early invasive breast cancer recorded in only one dataset, by CDT setting.



Key: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data.

Table A3. Identification of SACT defined drug-regimens in HES-APC, among women with early invasive breast cancer with details in both SACT and HES-APC.

Drug name/ combination recorded in SACT	Route of administration	Number of SACT admin[n]	Expected OPCS code(s) for drug name/combination		Identification of drug in HES-APC			
			Proc	Delivery	% of matched HES admissions (n)	Number of HES admissions w/OPCS codes	% of HES admissions with a matching drug record in SACT	
							Drug alone	Drug given with other drugs
Trastuzumab *(HER2+ pts only)*	Intravenous or subcutaneous	52 743	X713	X723	88% (n=46 293)	48,480	95% (n=46 293)	4% (n=1700)
FEC	Intravenous	43 276	X702	X723	94% (n=40 867)	64 948	90% (n=58 122)	2% (n=1512)
EC	Intravenous	18 284			94% (n=17 255)			
Paclitaxel	Intravenous	26 923	X704 X705 X711 X712 X713 X714 X715 X703 X705 X715	X721 X721 X721 X721 X721 X721 X721 X722 X722 X722	58% (n=15 512)	26 522	58% (n=15 512)	13% (n=3357)
Docetaxel	Intravenous	20 819	X705 X711 X712 X714 X715 X712 X713 X714 X715 X703 X705 X712 X713 X715	X721 X721 X721 X721 X721 X722 X722 X722 X722 X723 X723 X723 X723 X723	79% (n=16 384)	109 465	15% (n=16 384)	11% (n=12 497)

Key: SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes Statistics Admitted Patient Care data; proc = procurement; admin[n] = administrations; OPCS = Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures; HER2+ = human epidermal growth receptor 2 positive; pts = patients; FEC = fluorouracil, epirubicin, cyclophosphamide; EC = epirubicin & cyclophosphamide.

9.3. Appendix 3 – Supplementary material for Research Paper 2

Table A1. Information recorded on endocrine therapy (ET) in each data source

Data item/information	SACT	COSD	Registry	CWT	PCPD
Fact of ET (Y)	X	X	X	X	X
Treatment start date	X	X	X	X	X
Drug name	X		X		X
Subsequent ET dates	X	Unclear	Unclear		X
Type of ET	X	X			
Treatment purpose (e.g. for recurrence)		X			

Key: SACT = Systemic Anti-Cancer Therapy dataset; COSD = Cancer Outcomes and Services data; Registry = Cancer Registry data; CWT = Cancer Waiting Times data; PCPD = Primary Care Prescription Database.

Note: The information on subsequent ET dates in COSD and Registry are marked as "unclear" because these data sets collect treatment start date but the data provided for some patients includes multiple treatment dates.

Table A2. Identification of endocrine therapy in the Systemic Anti-Cancer Therapy Dataset and the Primary Care Prescription Database

Recording of endocrine therapy based on the drug name field in the Systemic Anti-Cancer Therapy Dataset	The subset of prescribing data provided, based on the Virtual Medicinal Product (VMP) name field in the Primary Care Prescription Database
Aminoglutethimide	Aminoglutethimide
Anastrozole	Anastrozole
Buserelin	Buserelin
Exemestane	Exemestane
Formestane	Formestane
Fulvestrant	Fulvestrant
Goserelin Acetate	Goserelin Acetate
Gtx-024	Letrozole
Letrozole	Leuprorelin Acetate
Leuprorelin Acetate	Tamoxifen Citrate
Megestrol	Toremifene Citrate
Medroxyprogesterone	Triptorelin (Acetate)
Tamoxifen Citrate	
Toremifene Citrate	
Triptorelin (Acetate)	

Note: Details on drug name captured within Registry data are not provided as this information was poorly completed and is information collected in addition to treatment type (endocrine therapy).

Table A3: Characteristics of women diagnosed with hormone receptor-positive invasive breast cancer, and breakdown of the recording of endocrine therapy (ET) in COSD or PCPD by characteristic.

Characteristic		Number of patients (column %)		Patients with ET use recorded within 12m of diagnosis (row %)				
				Captured in either data source (COSD or PCPD)		Of which N/% in PCPD alone		Adjusted p-value*
	All women	110529		104389	94.4%	46545	44.6%	
Age at diagnosis	50-59 years	32303	29.2%	30206	93.5%	16763	55.5%	<0.0001
	60-69 years	35004	31.7%	33022	94.3%	16365	49.6%	
	70-79 years	25461	23.0%	24153	94.9%	9959	41.2%	
	80+ years	17761	16.1%	17008	95.8%	3458	20.3%	
Year of diagnosis	2015	17402	15.7%	16528	95.0%	6917	41.9%	<0.0001
	2016	23026	20.8%	21783	94.6%	9340	42.9%	
	2017	23201	21.0%	21953	94.6%	9764	44.5%	
	2018	24192	21.9%	22791	94.2%	10433	45.8%	
	2019	22708	20.5%	21334	93.9%	10091	47.3%	
IMD 2019	1 – Most deprived	16397	14.8%	15546	94.8%	6960	44.8%	<0.0001
	2	19387	17.5%	18306	94.4%	7963	43.5%	
	3	23366	21.1%	21985	94.1%	9559	43.5%	
	4	25175	22.8%	23802	94.5%	10590	44.5%	
	5 - Least deprived	26204	23.7%	24750	94.5%	11473	46.4%	
Charlson Comorbidity Index	0	90336	81.7%	85393	94.5%	40153	47.0%	<0.0001
	1	10914	9.9%	10292	94.3%	3767	36.6%	
	2+	6478	5.9%	6112	94.4%	1428	23.4%	
	Unknown	2801	2.5%	2592	92.5%	1197	46.2%	
SCARF index	Fit	82091	74.3%	77573	94.5%	37078	47.8%	<0.0001
	Mild-Moderate	20840	18.9%	19698	94.5%	7469	37.9%	
	Severe	4797	4.3%	4526	94.4%	801	17.7%	
	Unknown	2801	2.5%	2592	92.5%	1197	46.2%	
Disease stage	Early invasive	101855	92.2%	96635	94.9%	44144	45.7%	0.0031
	Locally advanced	4173	3.8%	3896	93.4%	1403	36.0%	
	Metastatic	4501	4.1%	3858	85.7%	998	25.9%	
Tumour stage	T1	61265	55.4%	57899	94.5%	26879	46.4%	<0.0001
	T2	38744	35.1%	36889	95.2%	16210	43.9%	
	T3	5629	5.1%	5270	93.6%	2533	48.1%	
	T4	3359	3.0%	3059	91.1%	581	19.0%	
	Unknown	1532	1.4%	1272	83.0%	342	26.9%	
Nodal stage	N0	76624	69.3%	72481	94.6%	31555	43.5%	<0.0001
	N+	31944	28.9%	30224	94.6%	14584	48.3%	
	Unknown	1961	1.8%	1684	85.9%	406	24.1%	
HER2 status	Positive	9276	8.4%	8257	89.0%	4500	54.5%	<0.0001
	Borderline	11733	10.6%	11170	95.2%	4928	44.1%	
	Negative	79588	72.0%	75670	95.1%	33477	44.2%	
	Unknown	9932	9.0%	9292	93.6%	3640	39.2%	
Invasive grade	G1	20136	18.2%	18676	92.7%	8054	43.1%	<0.0001
	G2	66499	60.2%	64180	96.5%	27915	43.5%	
	G3	22716	20.6%	20562	90.5%	10323	50.2%	
	Unknown	1178	1.1%	971	82.4%	253	26.1%	
Primary surgery	No surgery	17200	15.6%	16102	93.6%	2167	13.5%	<0.0001
	BCS	69694	63.1%	66004	94.7%	33500	50.8%	
	Mastectomy	23635	21.4%	22283	94.3%	10878	48.8%	
Radiotherapy	No	35670	32.3%	33124	92.9%	10642	32.1%	<0.0001
	Yes	74859	67.7%	71265	95.2%	35903	50.4%	
Chemotherapy	No	85008	76.9%	81083	95.4%	32526	40.1%	<0.0001
	Yes	25521	23.1%	23306	91.3%	14019	60.2%	

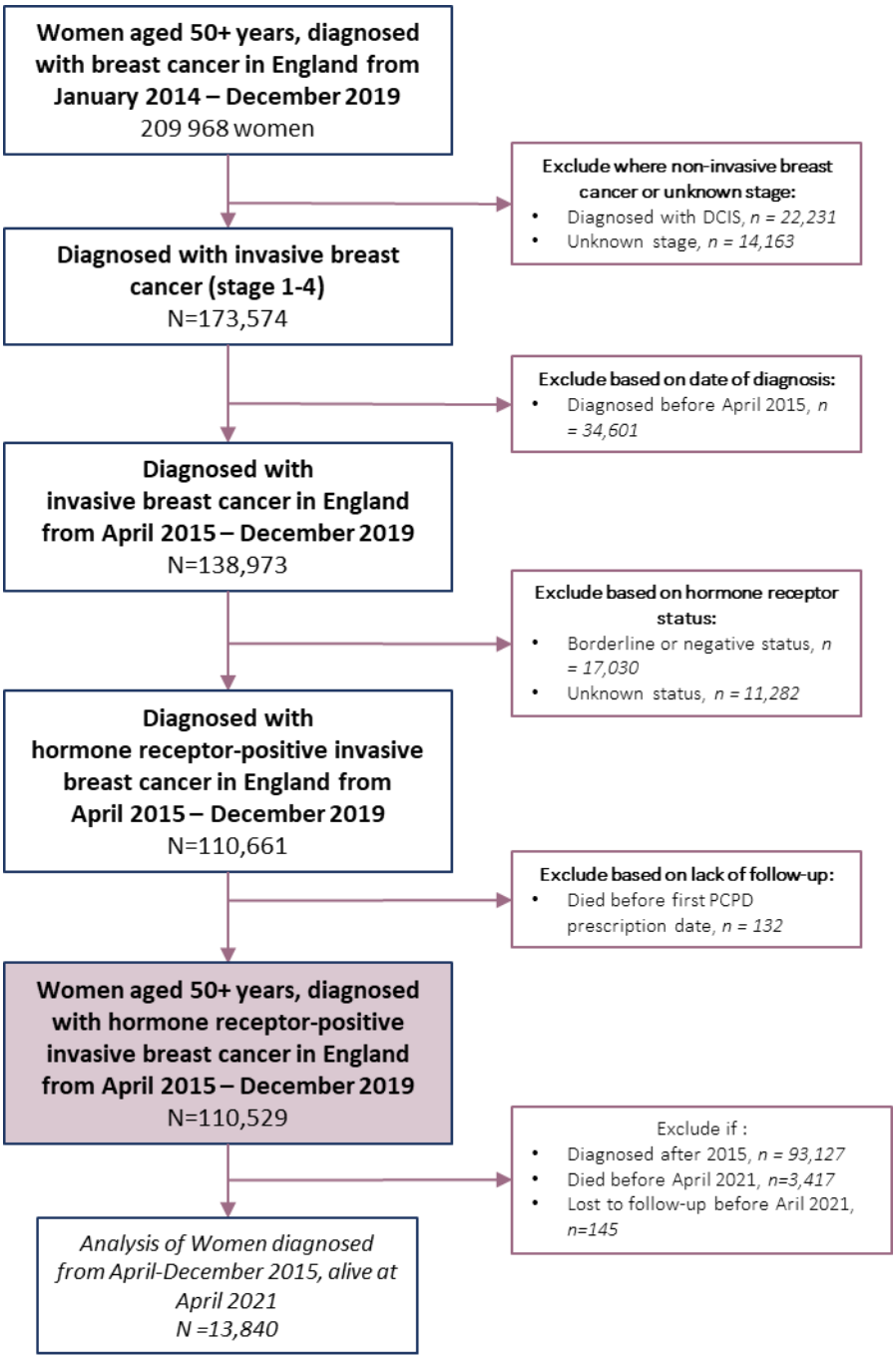
Key: ET = Endocrine Therapy; COSD = Cancer Outcomes and Services Dataset; PCPD = Primary Care Prescription Database;
IMD = Index of Multiple deprivation; HER2 = human epidermal growth receptor 2; SCARF = Secondary Care Administrative Records Frailty;
BCS = breast-conserving surgery.
*grouped p-value from multilevel mixed-effects logistic regression models including all factors in the table; outcome is ET use in COSD (with or without PCPD) vs PCPD only.

Table A4: Comparison of type of endocrine therapy (ET) by data source, among all women with hormone receptor-positive early invasive or locally advanced breast cancer.

PCPD	COSD					
	Neoadjuvant	Same month as surgery	Adjuvant	PET	No ET	Total
Neoadjuvant	4499	64	106	134	719	5522
Same month as surgery	857	1635	211	2	2177	4882
Adjuvant	610	184	34170	1	41129	76094
PET	1	0	0	11487	1522	13010
No ET	56	137	570	260	5497	6520
Total	6023	2020	35057	11884	51044	106,028

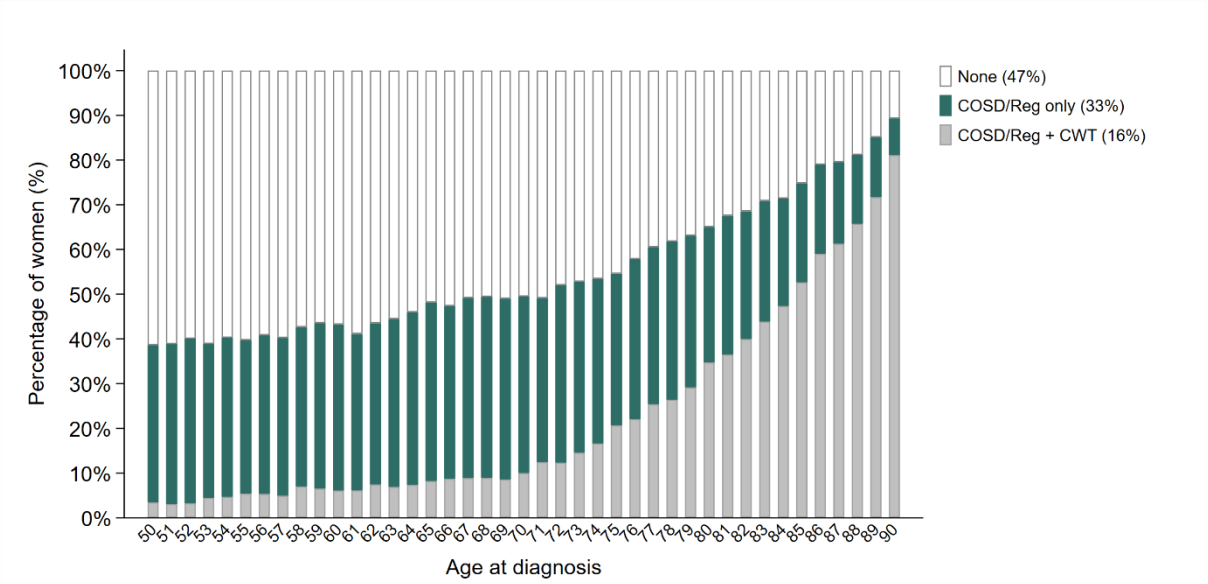
Key: COSD = Cancer Outcomes and Services Dataset; PCPD = Primary Care Prescription Database; PET = Primary Endocrine Therapy.

Figure A1: Details of patient selection from women aged 50 and over, diagnosed with breast cancer in a NHS trust in England, between 2014 and 2019.



Key: DCIS = ductal carcinoma in situ; PCPD = Primary Care Prescription Database.

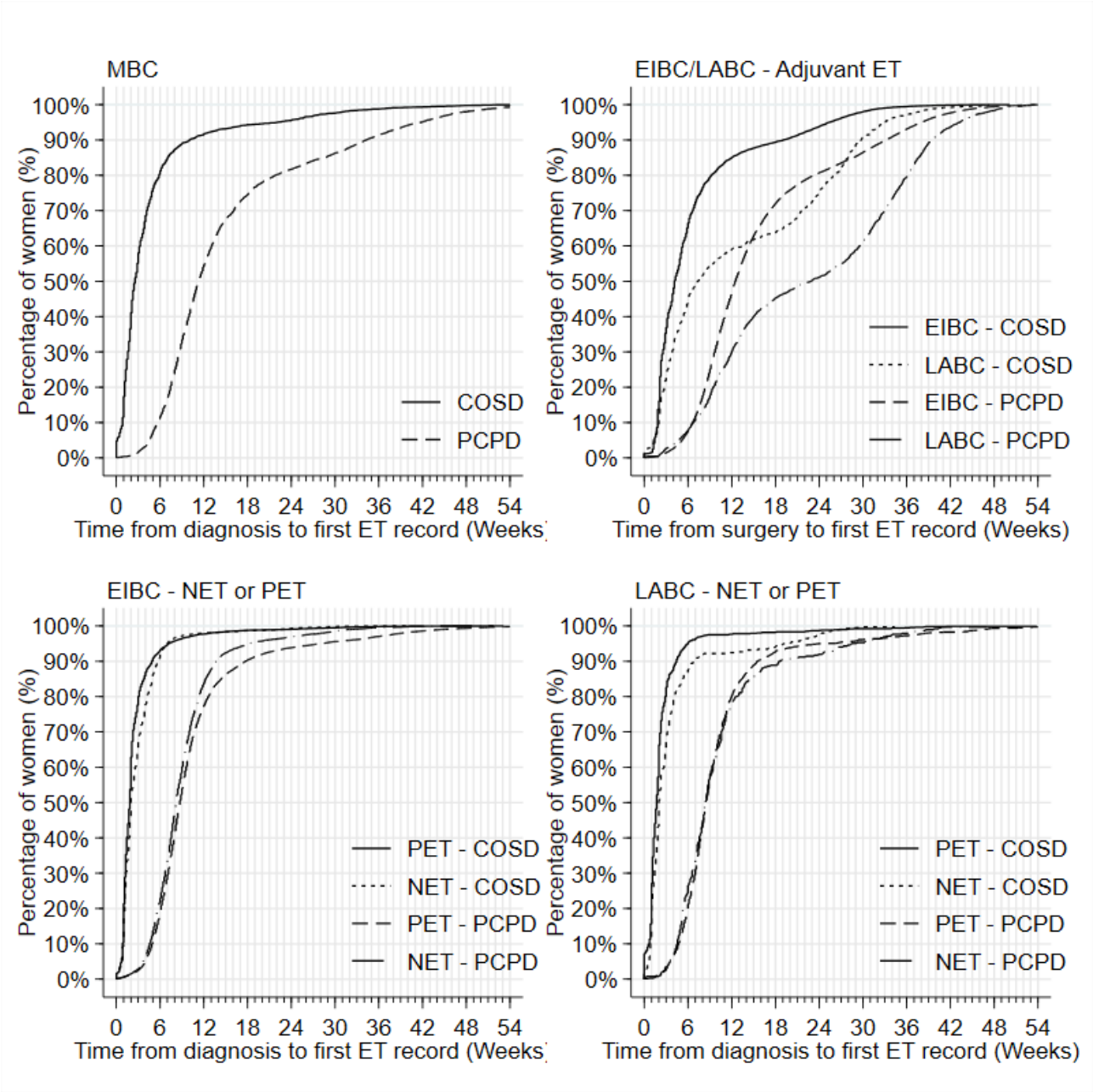
Figure A2. Recording of endocrine therapy within 12 months of diagnosis in secondary care data, by age at diagnosis, among all women with hormone receptor-positive invasive breast cancer.



Key: CWT = Cancer Waiting Times data; COSD = Cancer Outcomes and Services data; Registry = Cancer Registry data; SACT = Systemic Anti-Cancer Therapy dataset; All = CWT, COSD, Registry and SACT.

Note: Secondary care datasets with <5% are not presented in the figure above. These were: SACT+CWT = 0%; CWT only = 0.08%; SACT only = 0.3%; All datasets = 1%; SACT+COSD/Reg = 2%

Figure A3. Time to earliest endocrine therapy (ET) record* within 12 months of diagnosis, by disease stage and type of ET, among women with ET recorded in either data source.



*For ET prescriptions recorded in PCPD the date was taken to be the middle of the calendar month

Key: MBC = metastatic breast cancer; LABC = locally advanced breast cancer; EIBC = early invasive breast cancer; Adjuvant ET = adjuvant endocrine therapy (receiving surgery within 12m of diagnosis with no prior ET recorded); NET = neoadjuvant endocrine therapy (earliest ET prescription before date of surgery within 12m of diagnosis); PET = primary endocrine therapy (women with EIBC/LABC not receiving surgery within 12m of diagnosis).

9.4. Appendix 4 – Supplementary material for Research Paper 3

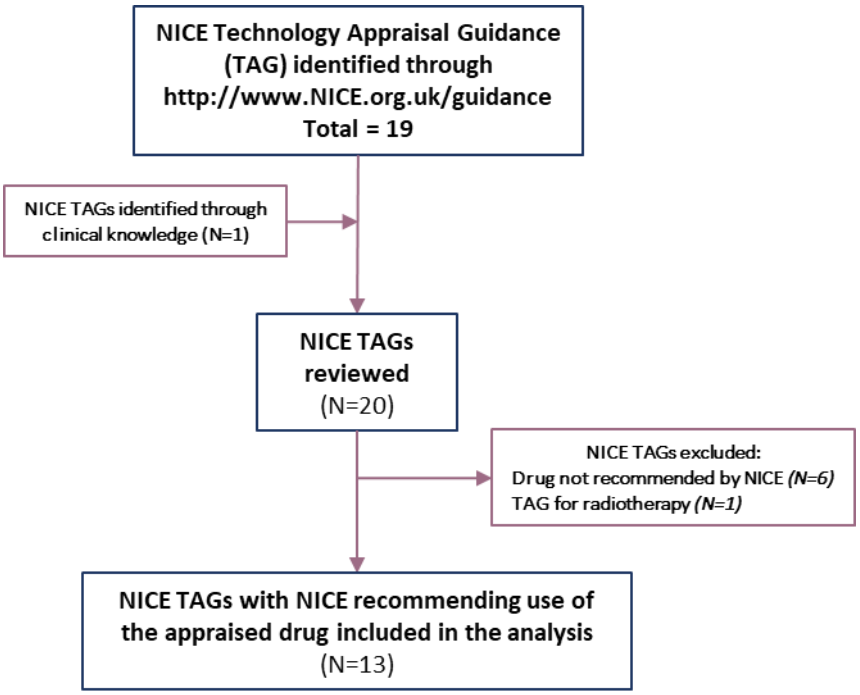
Appendix 1:

Steps taken to identify published NICE Technology Appraisal Guidance (TAGs) within https://www.nice.org.uk/guidance . Performed on 17/03/2022.

Step 1: Select "Technology Appraisal Guidance".

Step 2: Ensure you are on the "Published" tab.

Step 3: Apply filters. Last updated date: From date = "01/01/2000"; To date = "31/12/2019". Type: select "Guidance". Guidance programme: select "Technology appraisal guidance". Filter by title or keyword: type "breast cancer".



NICE Technology Appraisal Guidance excluded from the study.

Reference	Reason for exclusion
NICE. Bevacizumab in combination with a taxane for the first-line treatment of metastatic breast cancer. 2011 [Available from: https://www.nice.org.uk/guidance/ta214]	Drug not recommended by NICE
NICE. Fulvestrant for the treatment of locally advanced or metastatic breast cancer. 2011 [Available from: https://www.nice.org.uk/guidance/ta239]	Drug not recommended by NICE
NICE. Lapatinib or trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone-receptor-positive breast cancer that overexpresses HER2. 2012 [Available from: https://www.nice.org.uk/guidance/ta257]	Drug not recommended by NICE
NICE. Bevacizumab in combination with capecitabine for the first-line treatment of metastatic breast cancer. 2012 [Available from: https://www.nice.org.uk/guidance/ta263]	Drug not recommended by NICE
NICE. Fulvestrant for untreated locally advanced or metastatic oestrogen-receptor positive breast cancer. 2018 [Available from: https://www.nice.org.uk/guidance/ta503]	Drug not recommended by NICE
NICE. Intrabeam radiotherapy system for adjuvant treatment of early breast cancer. 2018 [Available from: https://www.nice.org.uk/guidance/ta501]	TAG for radiotherapy not an oncological drug
NICE. Eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen. 2018 [Available from: https://www.nice.org.uk/guidance/ta515]	Drug not recommended by NICE

Appendix 2:

Figure A1: Details of patient selection from women aged 50 and over, diagnosed with breast cancer in a NHS trust in England, between 2014 and 2019.

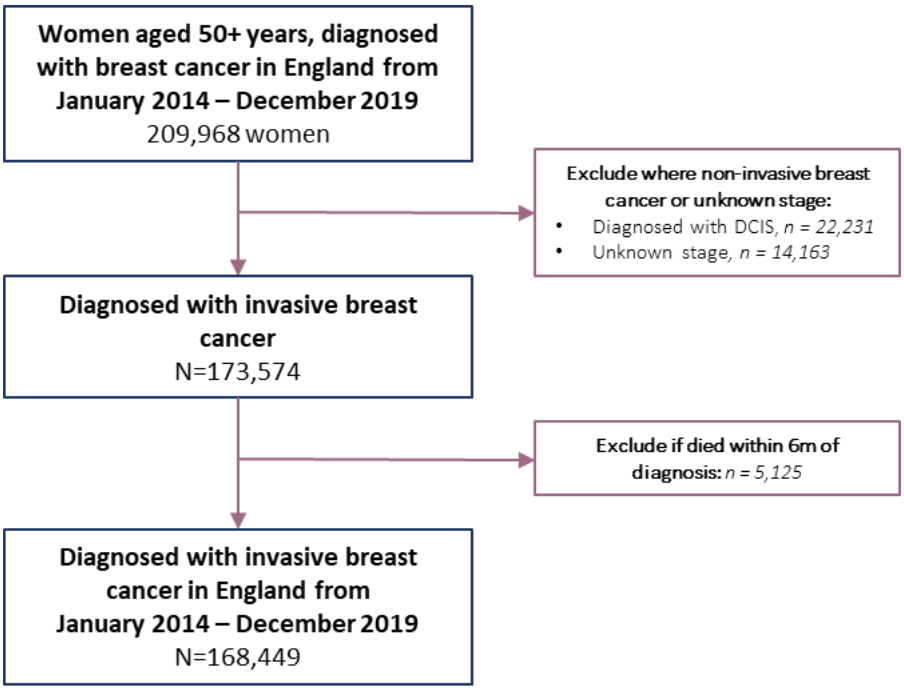
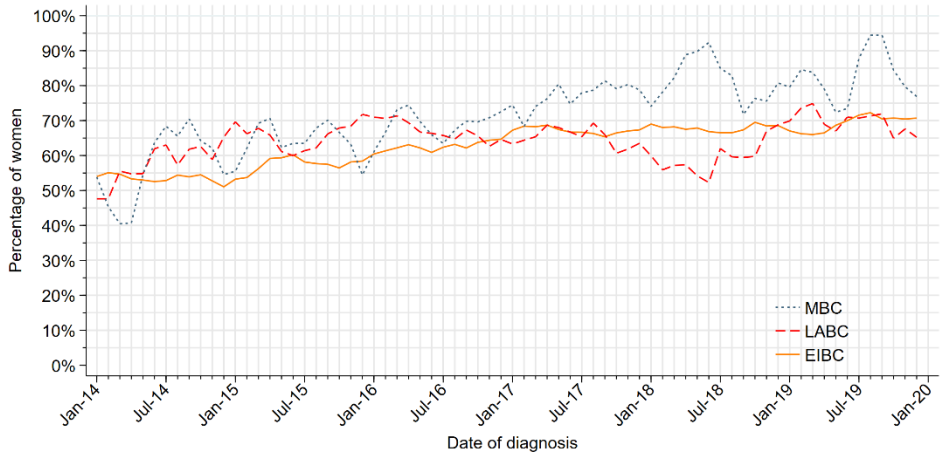


Figure A2. Trastuzumab utilisation among women initially diagnosed with HER2-positive early invasive breast cancer (EIBC), locally advanced breast cancer (LABC) or metastatic breast cancer (MBC) from 2014–2019

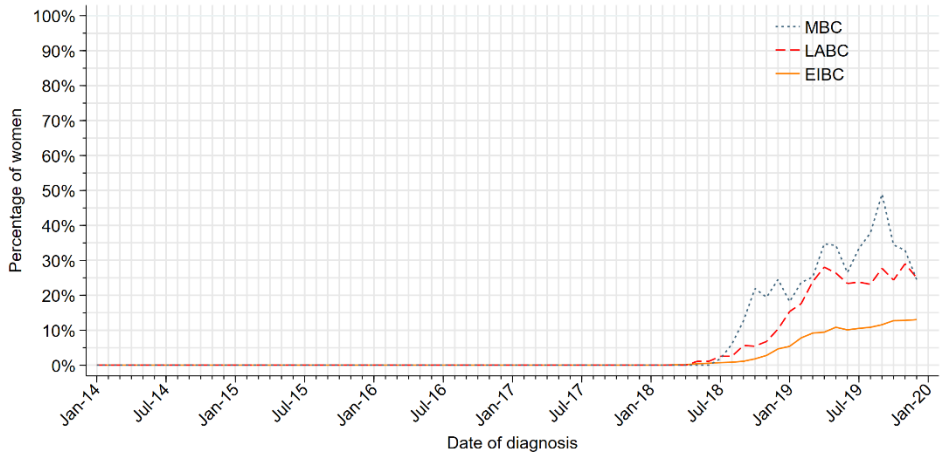
Figure A2a. Line plot of trastuzumab initiations, per 1000 women, by date of diagnosis



***Note**: line plot created using tsline, with smoothed moving average using uniform weights.*

Trastuzumab initiation includes trastuzumab (Herceptin) and trastuzumab biosimilars.

Figure A2b. Line plot of trastuzumab biosimilar initiations, per 1000 women, by date of diagnosis



***Note**: line plot created using tsline, with smoothed moving average using uniform weights.*

Figure A2c. Observed percentage of women initiating trastuzumab, by Government office Region (GOR) and age at diagnosis

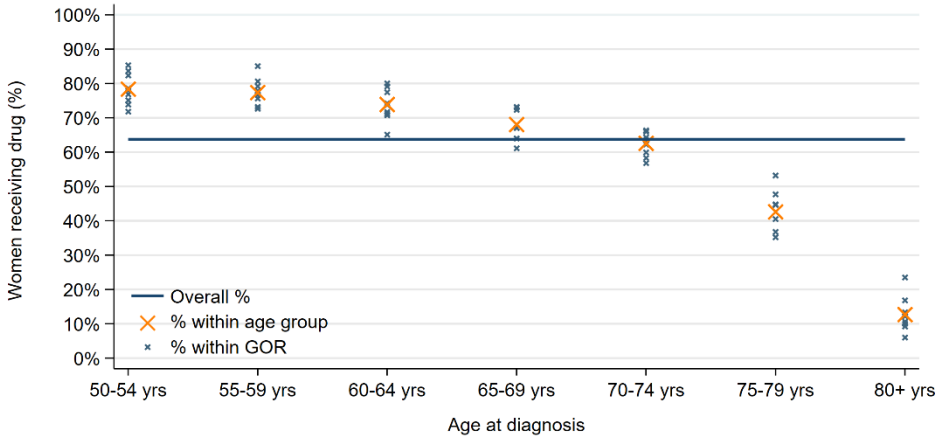


Figure A2d. Observed percentage of women initiating trastuzumab, by Government office Region (GOR) and age at diagnosis

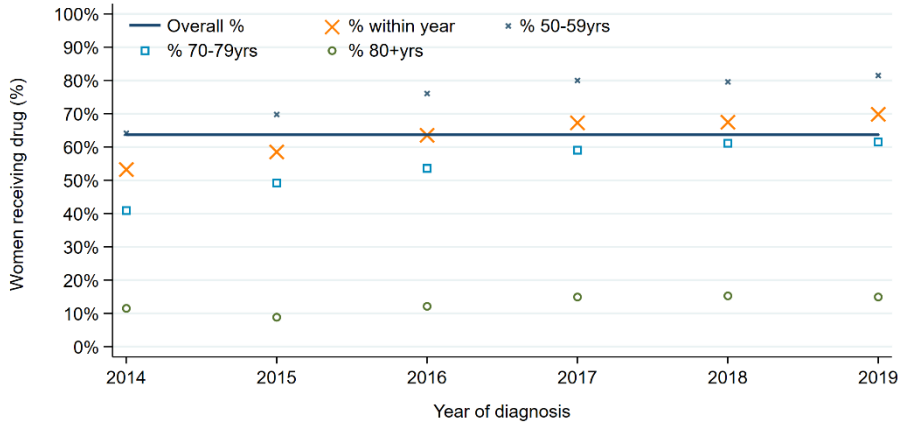


Figure A3. Post-publication pertuzumab utilisation among women initially diagnosed with HER2-positive early invasive breast cancer (EIBC) or locally advanced breast cancer (LABC)

Figure A3a. Observed percentage of women initiating neoadjuvant pertuzumab for EIBC or LABC, by Government office Region (GOR) and age at diagnosis

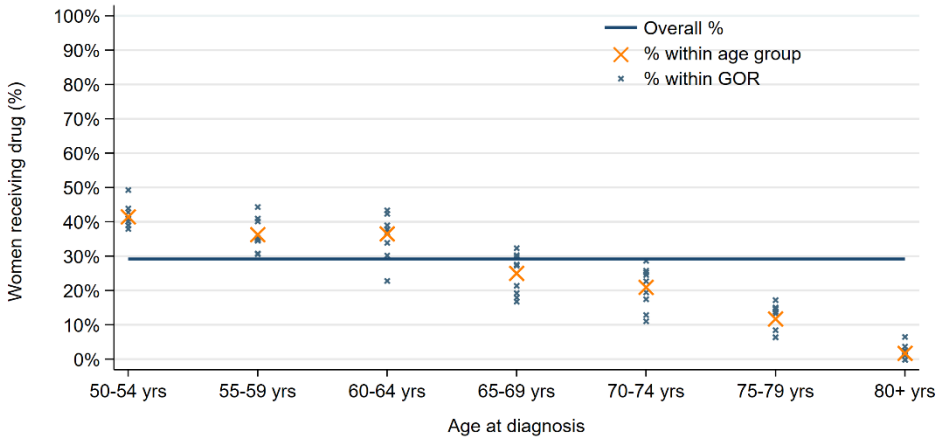


Figure A3b. Observed percentage of women initiating adjuvant pertuzumab for EIBC or LABC, by Government office Region (GOR) and age at diagnosis

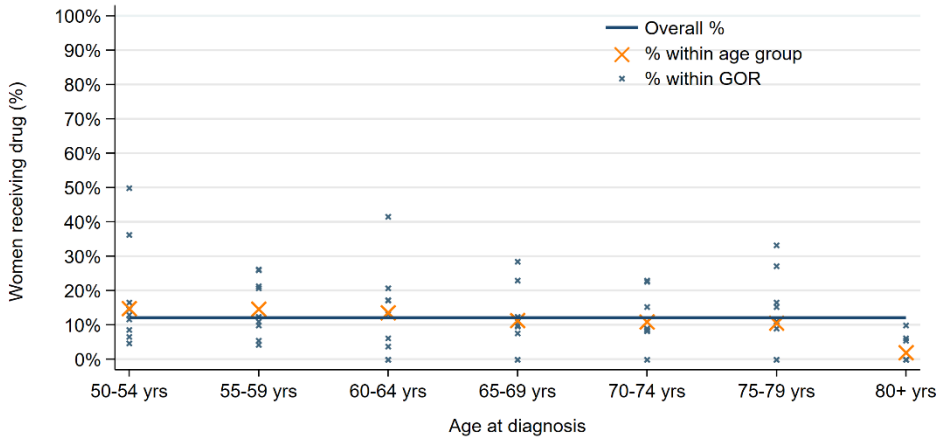


Figure A4. Post-publication palbociclib utilisation among women initially diagnosed with HER2-positive locally advanced breast cancer (LABC) or metastatic breast cancer (MBC)

Figure A4a. Observed percentage of women initiating palbociclib for LABC or MBC, by year of diagnosis and age at diagnosis

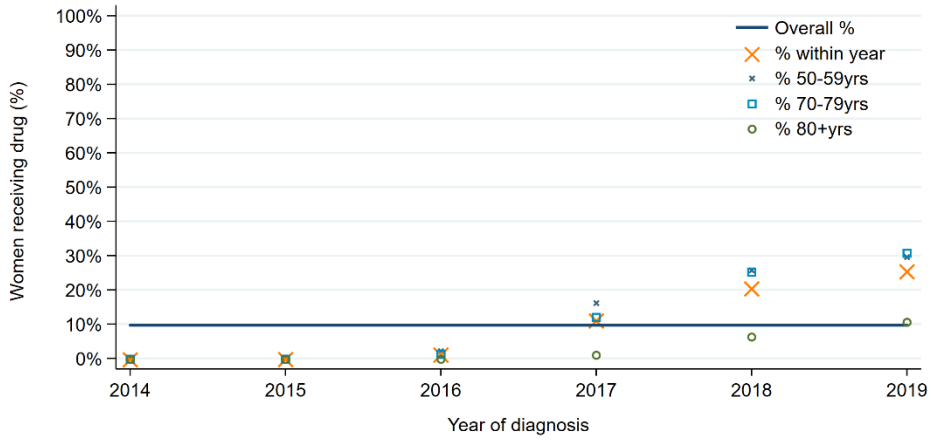


Figure A4b. Observed post-publication percentage of women initiating palbociclib for LABC or MBC, by Government office Region (GOR) and age at diagnosis

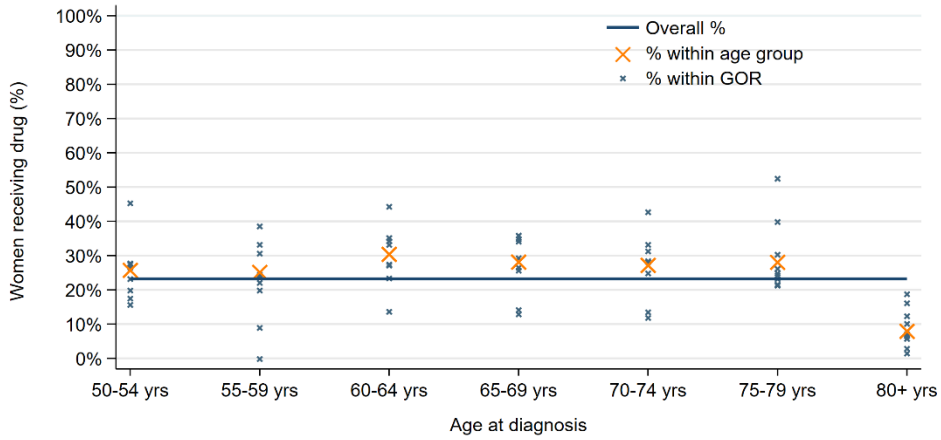


Table A1: Adjusted odds ratios of drug use recorded within SACT, for oncological treatments for breast cancer recommended by NICE for use as first-line treatment following initial diagnosis, among eligible women

Drug (NICE TAG date)	Indication/ Eligibility	Cohort of women diagnosed	Use % (n/N)	Adjusted* odds ratio (95% confidence interval)									
				50-59 years	60-69 years	70-79 years	80+ years	Overall p-value	CCI = 0	CCI = 1	CCI = 2+	Overall p-value	GOR p-value
Trastuzumab	HER2+ EIBC/LABC/MBC	Jan2014-Dec2019	63.7% (10776/16987)	1.00	0.73 (0.67-0.80)	0.28 (0.25-0.03)	0.03 (0.03-0.04)	<0.0001	1.00	0.64 (0.57-0.73)	0.30 (0.25-0.37)	<0.0001	<0.0001
Neoadjuvant Pertuzumab	HER2+ EIBC/LABC	Jan2017-Dec2019	29.2% (2068/7083)	1.00	0.78 (0.68-0.89)	0.25 (0.21-0.30)	0.02 (0.01-0.03)	<0.0001	1.00	0.55 (0.43-0.71)	0.33 (0.21-0.51)	<0.0001	<0.0001
Pertuzumab	HER2+ MBC	Jan2014-Dec2019	52.0% (485/932)	1.00	0.84 (0.56-1.26)	0.23 (0.15-0.35)	0.04 (0.02-0.08)	<0.0001	1.00	0.87 (0.52-1.44)	0.46 (0.22-0.95)	0.0025	0.1420
Adjuvant Pertuzumab	HER2+ EIBC/LABC	Apr2019-Dec2019	12.1% (129/1068)	1.00	1.23 (0.57-2.66)	0.23 (0.11-0.47)	0.03 (0.01-0.10)	<0.0001	1.00	0.53 (0.21-1.36)	0.28 (0.08-1.01)	0.0805	0.0102
Palbociclib	HER2-, HR+ LABC/MBC	Jan2018-Dec2019	23.2% (455/1962)	1.00	1.11 (0.79-1.57)	0.97 (0.70-1.36)	0.21 (0.13-0.32)	<0.0001	1.00	0.74 (0.52-1.07)	0.64 (0.38-1.07)	0.1245	0.1999

Key*: HER2 = human epidermal growth receptor 2; MBC = metastatic breast cancer; EIBC = early invasive breast cancer; LABC = locally advanced breast cancer; HR+ = hormone receptor-positive.*
**Odds ratio from multilevel mixed-effects logistic regression model, adjusting for age, comorbidity burden, diagnosis year, deprivation, ethnicity, stage group, tumour characteristics, geographical region.*

Table A2. Observed percentage of drug use recorded within SACT, for oncological treatments for breast cancer recommended by NICE for use as first-line treatment following initial diagnosis, among eligible women, by Government Office Region

Government Office Region	Trastuzumab		Neoadjuvant Pertuzumab				Pertuzumab for MBC				Adjuvant Pertuzumab				Palbociclib			
	All use		All use		Post-publication use		All use		Post-publication use		All use		Post-publication use		All use		Post-publication use	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
North East	1,232	58.4%	1,076	11.3%	439	24.6%	62	38.7%	14	64.3%	888	1.5%	68	11.8%	464	11.2%	157	26.8%
North West	2,186	69.6%	1,940	18.6%	962	35.1%	100	57.0%	34	52.9%	1,409	1.6%	134	3.0%	877	10.9%	294	24.5%
Yorkshire & the Humber	1,830	69.5%	1,592	13.0%	765	25.2%	110	46.4%	28	71.4%	1,253	3.7%	132	22.0%	633	9.5%	203	24.1%
East Midlands	1,364	62.8%	1,198	14.1%	577	27.4%	68	58.8%	25	72.0%	951	2.4%	90	12.2%	386	6.0%	121	14.9%
West Midlands	1,817	60.9%	1,604	15.7%	726	32.2%	91	44.0%	24	66.7%	1,210	2.3%	98	13.3%	621	9.2%	196	24.5%
East England	2,168	62.3%	1,846	16.4%	910	30.9%	139	47.5%	35	65.7%	1,335	1.3%	116	8.6%	800	8.8%	247	21.9%
London	1,783	60.4%	1,478	16.1%	751	27.4%	109	60.6%	23	78.3%	1,054	2.8%	116	12.9%	661	10.0%	186	22.0%
South East	2,607	64.5%	2,247	14.9%	1,111	26.5%	143	53.1%	41	68.3%	1,700	2.6%	192	11.5%	944	10.8%	323	23.8%
South West	1,910	61.9%	1,652	16.8%	842	30.4%	110	59.1%	32	78.1%	1,216	2.8%	122	13.9%	718	9.1%	235	23.0%

Table A3. Full interrupted time series analysis output of drug initiations (per 1000)

Time period	Drug *(associated figure)*	Coefficient	95% confidence interval		P-value
	Neoadjuvant pertuzumab *(Figure 1a)*				
Ref	Drug initiations, per 1000, in Jan-2014	2.8	-17.1	22.7	0.779
Ref	Monthly change in drug initiations, per 1000, from Jan-2014 to Jun-2016	0.4	-0.1	0.9	0.087
T	Immediate change in drug initiations, per 1000, in Jul-2016 (6m prior to NICE TAG publication)	5.9	-20.4	32.1	0.657
T	Monthly change in drug initiations, per 1000, from Jul-2016 to Dec-2016	37.2	19.6	54.9	<0.0001
PP	Immediate change in drug initiations, per 1000, in Jan-2017 (month following NICE TAG publication in Dec-2016)	15.9	-80.9	112.7	0.743
PP	Monthly change in drug initiations, per 1000, from Jan-2017 onwards	1.9	0.9	2.9	<0.0001
	Pertuzumab for MBC *(Figure 2a)*				
Ref	Drug initiations, per 1000, in Jan-2014	292.2	224.0	360.3	<0.0001
Ref	Monthly change in drug initiations, per 1000, from Jan-2014 to Mar-2017	6.9	4.0	9.7	<0.0001
T	Immediate change in drug initiations, per 1000, in Apr-2017 (12m prior to NICE TAG publication)	92.5	-80.4	265.3	0.288
T	Monthly change in drug initiations, per 1000, from Apr-2017 to Mar-2018	-1.1	-22.9	19.9	0.915
PP	Immediate change in drug initiations, per 1000, in Apr-2018 (month following NICE TAG publication in Mar-2018)	23.9	-181.4	229.1	0.817
PP	Monthly change in drug initiations, per 1000, from Apr-2018 onwards	5.8	-6.1	17.7	0.332
	Adjuvant pertuzumab *(Figure 3a)*				
Ref	Drug initiations, per 1000, in Jan-2014	10.4	1.0	19.8	0.030
Ref	Monthly change in drug initiations, per 1000, from Jan-2014 to Mar-2018	-0.2	-0.4	0.0	0.060
T	Immediate change in drug initiations, per 1000, in Apr-2018 (12m prior to NICE TAG publication)	-10.5	-30.1	9.1	0.289
T	Monthly change in drug initiations, per 1000, from Apr-2018 to Mar-2019	8.1	4.7	11.6	<0.0001
PP	Immediate change in drug initiations, per 1000, in Apr-2019 (month following NICE TAG publication in Mar-2019)	45.1	8.5	81.8	0.017
PP	Monthly change in drug initiations, per 1000, from Apr-2019 onwards	-3.3	-8.7	2.1	0.230
	Palbociclib *(Figure 4a)*				
Ref	Drug initiations, per 1000, in Jan-2014	-0.2	-13.9	13.5	0.979
Ref	Monthly change in drug initiations, per 1000, from Jan-2014 to Dec-2016	0.7	0.1	1.3	0.024
T	Immediate change in drug initiations, per 1000, in Jan-2017 (12m prior to NICE TAG publication)	7.8	-23.1	38.7	0.615
T	Monthly change in drug initiations, per 1000, from Jan-2017 to Dec-2017	16.7	12.2	21.1	<0.0001
PP	Immediate change in drug initiations, per 1000, in Jan-2018 (month following NICE TAG publication in Dec-2016)	-32.9	-79.0	13.2	0.158
PP	Monthly change in drug initiations, per 1000, from Jan-2018 onwards	3.5	1.0	6.1	0.007

Key: Ref = pre-publication time period, T = Transition period, PP = post-publication period
The coefficients in the reference (pre-publication) period are the baseline rate and monthly trend. The coefficients for the transition and post-publication periods are the change in the rate in the first month of the period and the trend for that period.

9.5. Appendix 5 – Supplementary material for Research Paper 4

Table A1: Description of OPCS codes used to define surgical procedure.

OPCS Code	Description
Mx Excision codes	
B27.1	Total Mx and excision of both pectoral muscles and part of chest wall.
B27.2	Total Mx and excision of both pectoral muscles NEC.
B27.3	Total Mx and excision of pec minor.
B27.4	Total Mx NEC (incl simple mastectomy.)
B27.5	Subcutaneous Mastectomy.
B27.6	Skin sparing mastectomy
B27.8	Other specified total excision of breast
B27.9	Unspecified total excision of breast
BCS Excision codes	
B28.1	Quadrantectomy of breast
B28.2	Partial excision of breast NEC
B28.3	Excision of lesion of breast NEC
B28.5	Wire guided partial excision of breast
B28.7	Wire guided excision of lesion of breast
B28.8	Other specified other excision of breast
B28.9	Unspecified other excision of breast

Figure A1: Details of patient selection from women aged 50 and over, diagnosed with early invasive breast cancer in a NHS trust in England, between January 2014 and December 2017.

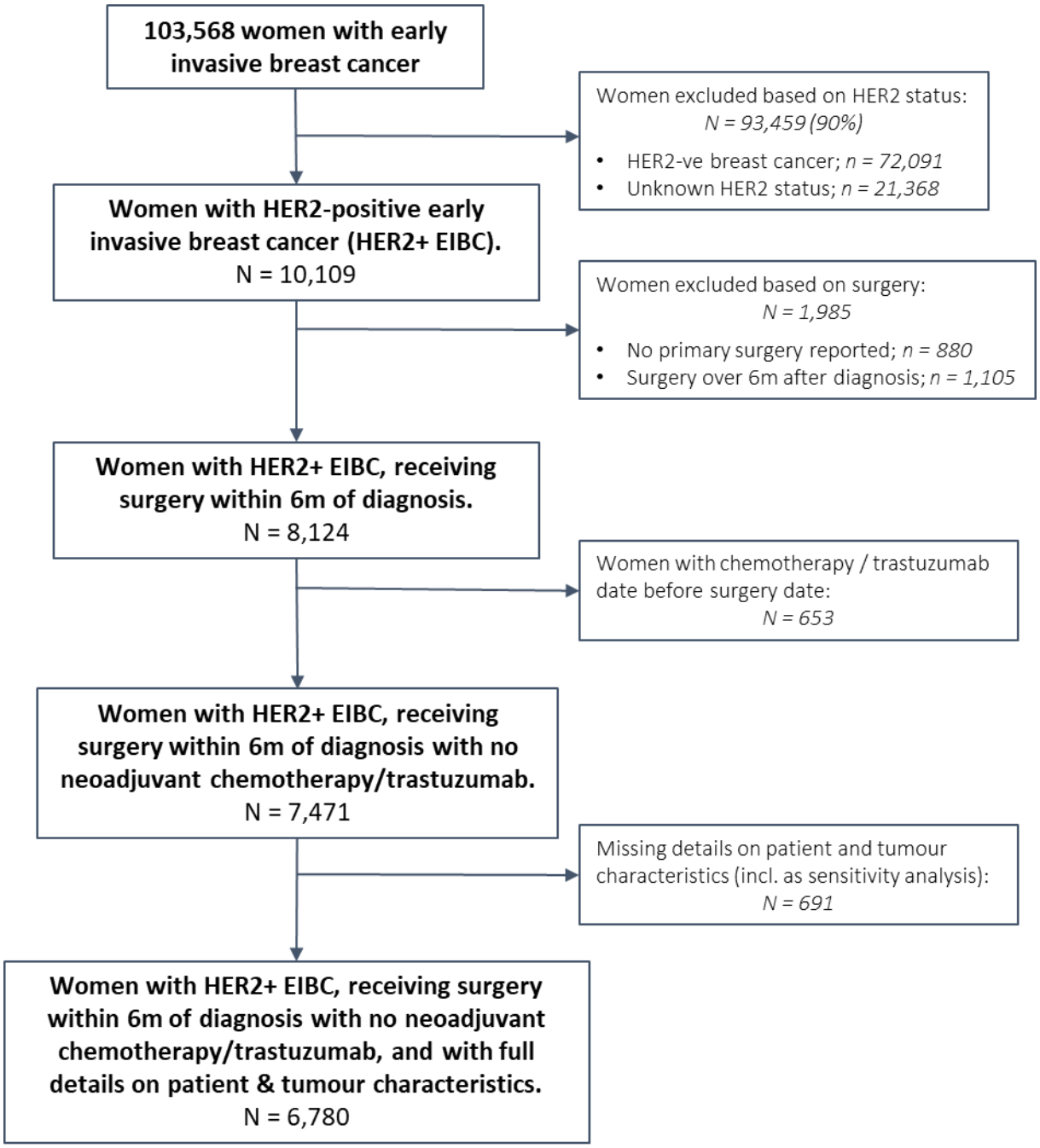
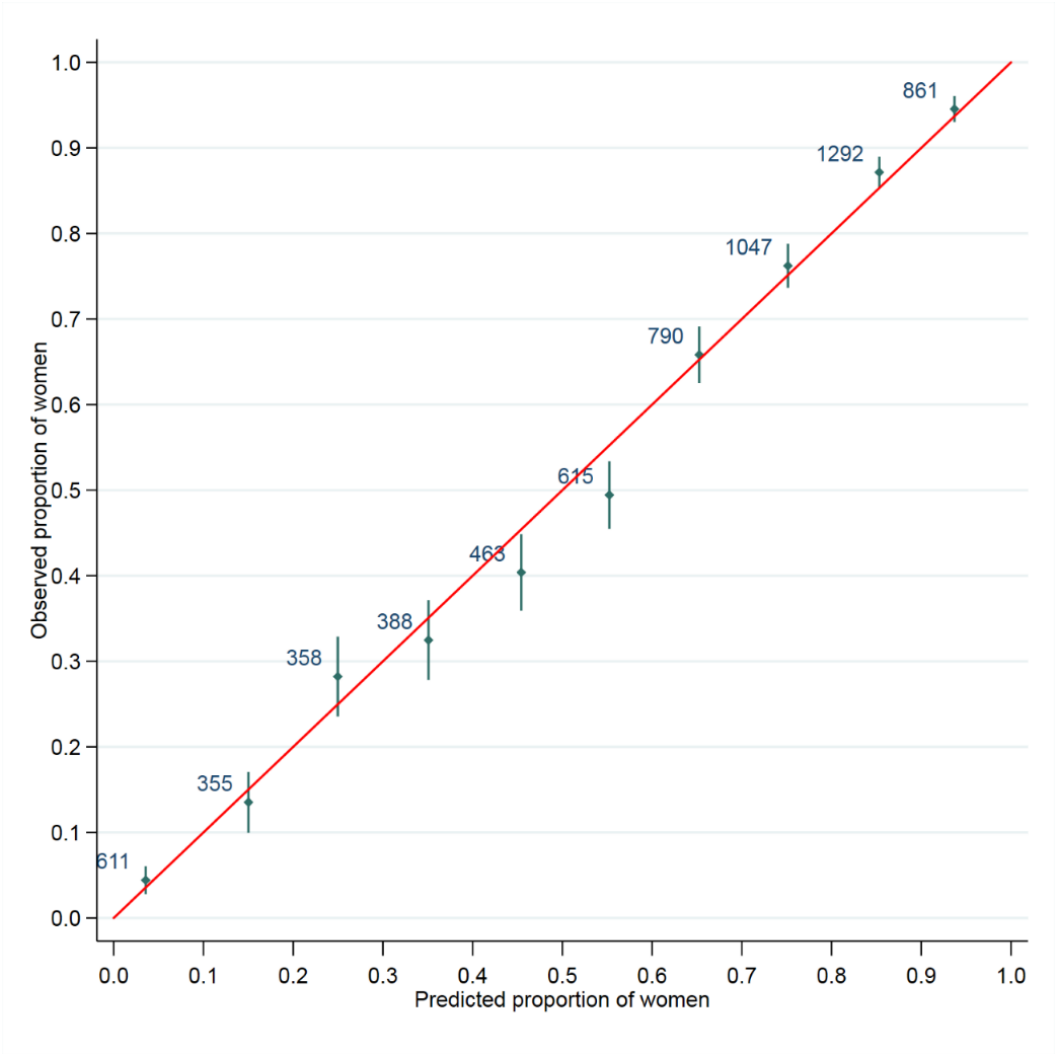


Figure A2: Calibration of prognostic multilevel mixed-effects logistic regression model by levels of predicted risk



9.6. Appendix 6 – Supplementary material for Research Paper 5

Figure A1: Time window for assessment of severe acute toxicity event.

Figure A2: Timing of chemotherapy in relation to commencing adjuvant trastuzumab.

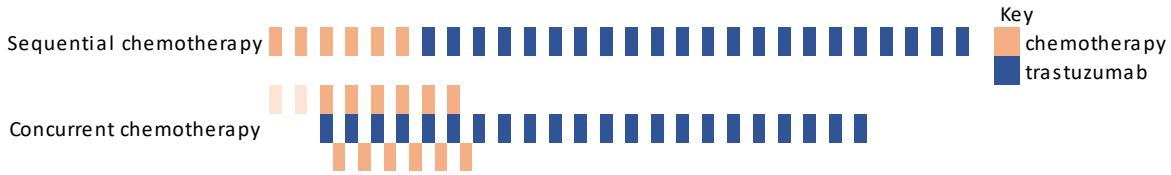


Figure A3: Details of patient selection from women aged 50 and over, diagnosed with breast cancer in England between 2014 and 2019.

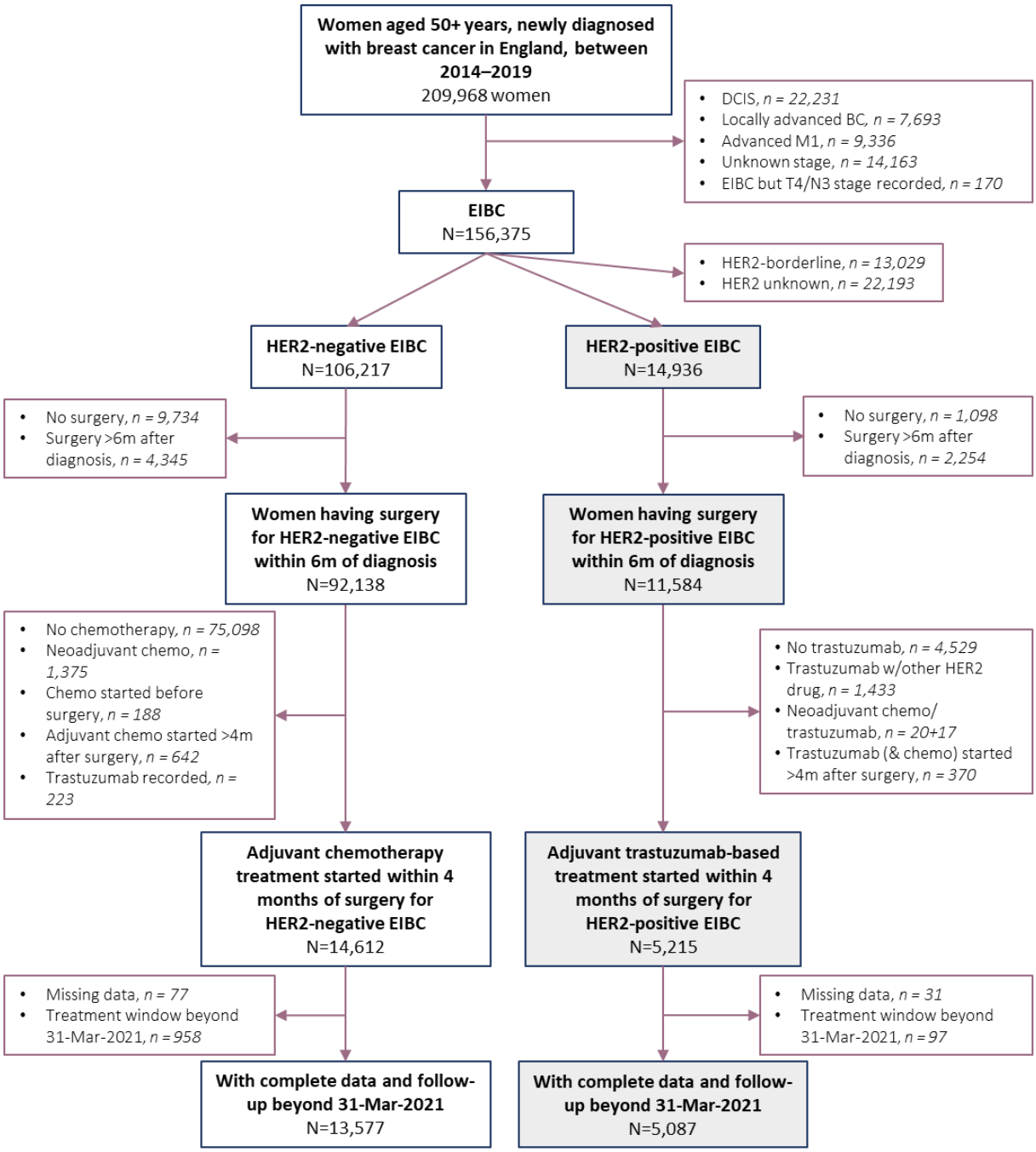
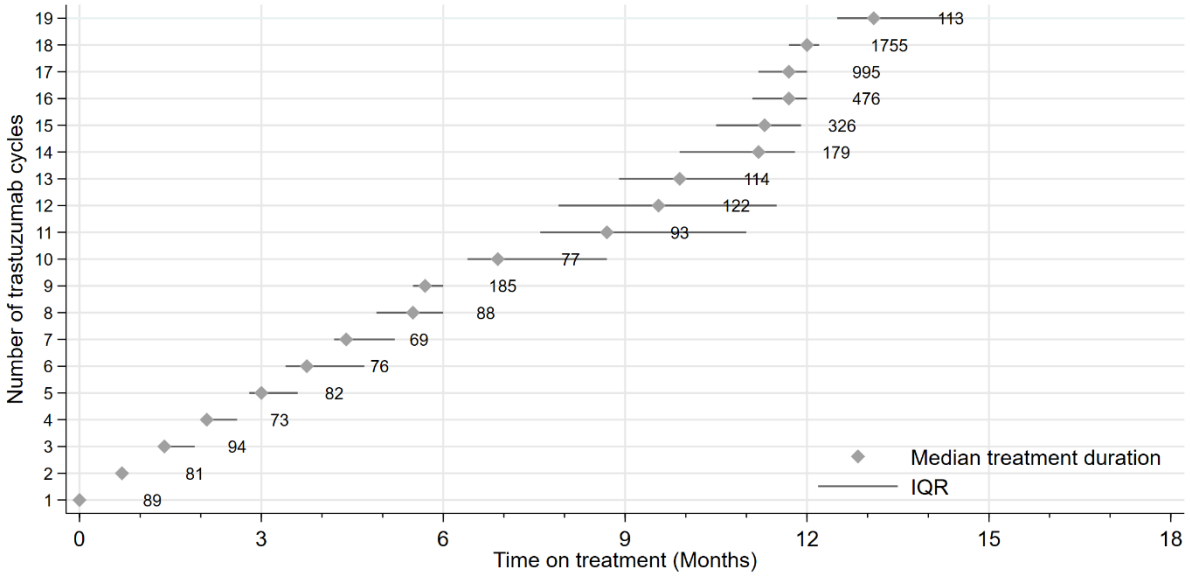


Figure A4: Trastuzumab treatment details among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer.

Duration of treatment by total number of cycles.



Frequency of trastuzumab cycles and percentage of women receiving each cycle, by cycle number, among women receiving more than one cycle of trastuzumab.

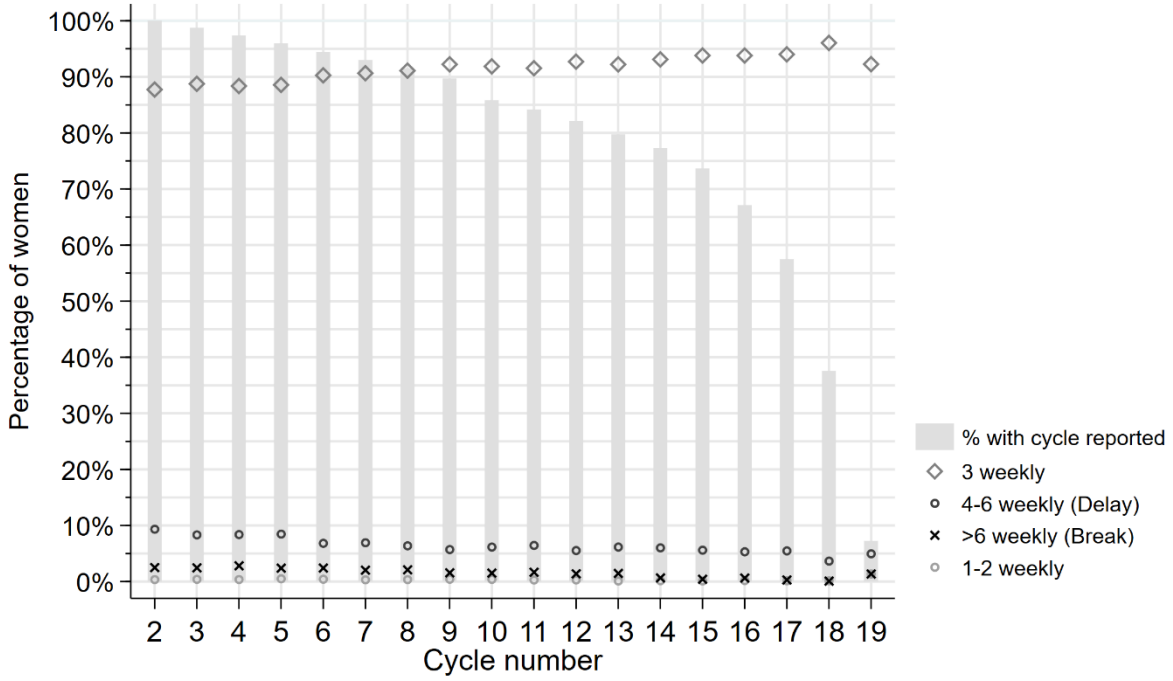


Figure A5. Time from first treatment cycle to first (a) severe acute toxicity event (b) cardiovascular disorder, among women with early invasive breast cancer, by HER2 status.

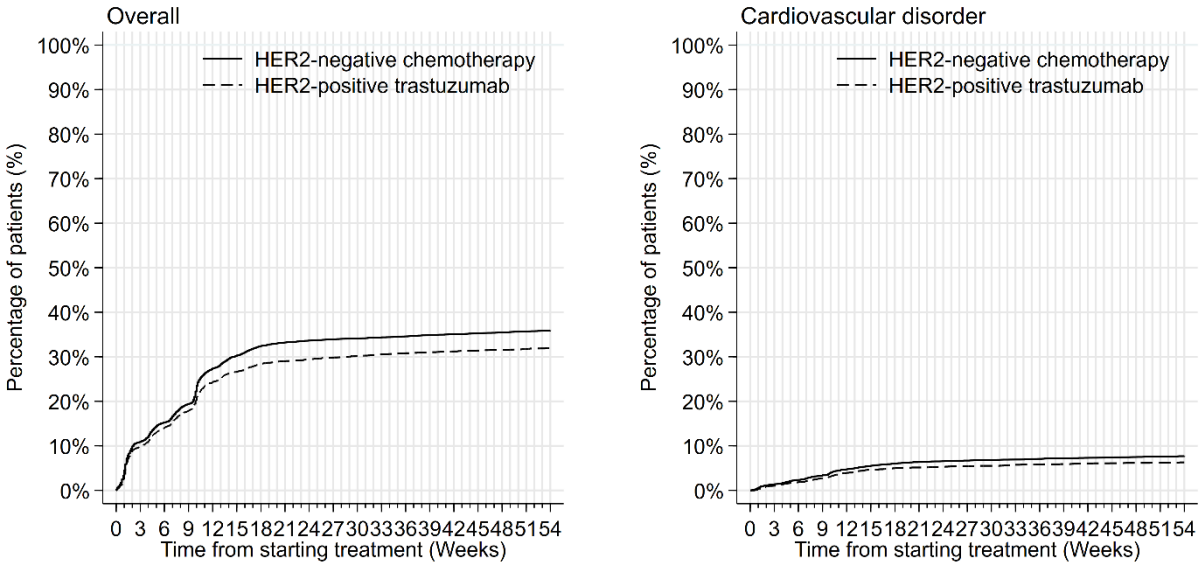


Figure A6: Percentage of women with a severe acute toxicity event, among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer, by age at diagnosis

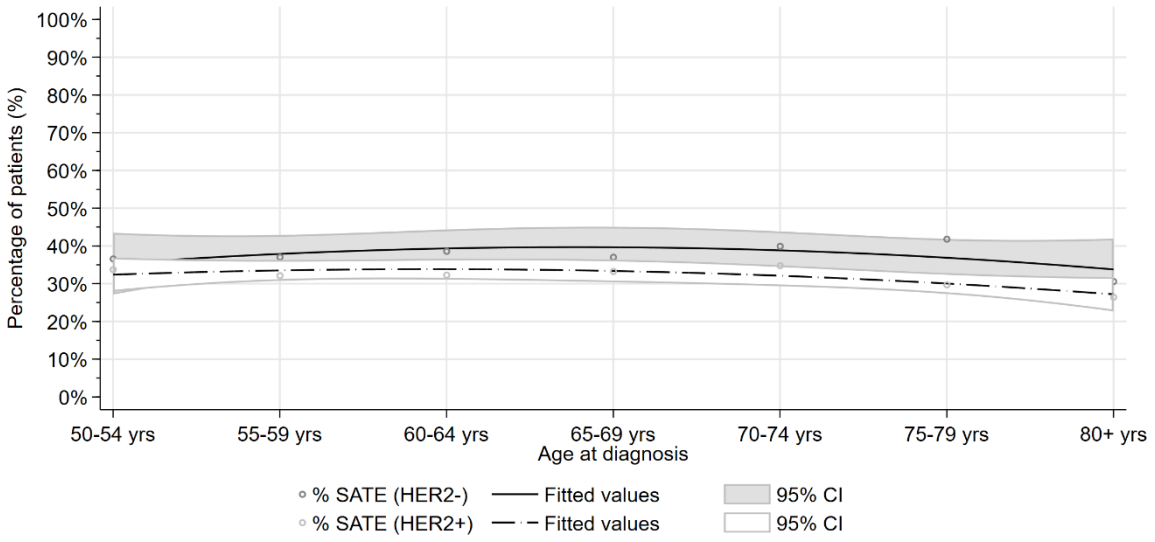


Table A1: Validated coding framework used to determine severe acute toxicity.

Event	ICD-10 code
Haematological	
Neutropenia	D701 D702 D703 D708 D709 D70X
Anaemia	D611 D618 D619 D648 D509* D630 D649*
Thrombocytopenia	D695 D696 D699 M311 R233
Infection	R502 R508 R509 R680 R650 R651 R659 A410 A411 A412 A413 A414 A415 A418 A419 A020 A021 A022 A028 A029 A040 A041 A042 A043 A044 A045 A046 A047 A048 A049 A050 A051 A052 A053 A054 A058 A059 A070 A071 A072 A073 A078 A079 A080 A081 A082 A083 A084 A085 A150 A151 A152 A153 A154 A155 A156 A157 A158 A159 A170 A171 A178 A179 A180 A181 A182 A183 A184 A185 A186 A187 A188 A190 A191 A192 A198 A199 A38 A38X A390 A391 A392 A394 A395 A398 A399 A400 A401 A402 A403 A408 A409 A420 A421 A422 A427 A428 A429 A46 A46X A480 A481 A482 A483 A484 A488 A490 A491 A492 A493 A498 A499 A810 A811 A812 A818 A819 A850 A852 A858 A86X A86 A870 A871 A872 A878 A879 A880 A881 A888 A89 A89X B001 B002 B003 B004 B005 B007 B008 B009 B010 B011 B012 B018 B019 B020 B021 B022 B023 B027 B028 B029 B07X B07 B080 B081 B082 B083 B084 B085 B088 B09X B150 B159 B160 B161 B162 B169 B170 B171 B172 B178 B179 B190 B199 B250 B251 B252 B258 B259 B270 B271 B278 B279 B300 B301 B302 B303 B308 B309 B330 B331 B332 B333 B334 B338 B340 B341 B342 B343 B344 B348 B349 B371 B372 B373 B374 B375 B376 B377 B378 B379 B440 B441 B442 B447 B448 B449 B450 B451 B452 B453 B457 B458 B459 B49X B59X B950 B951 B952 B953 B954 B955 B956 B957 B958 B960 B961 B962 B963 B964 B965 B966 B967 B968 B970 B971 B972 B973 B974 B975 B976 B977 B978 B99 B99X J200 J201 J202 J203 J204 J205 J206 J207 J208 J209 J22X J120 J121 J122 J123 J128 J129 J13 J14 J13X J14X J150 J151 J152 J153 J154 J155 J156 J157 J158 J159 J160 J168 J170 J171 J172 J173 J178 J180 J181 J182 J188 J189 J09 J100 J101 J108 J110 J111 J118 J850 J851 J852 J853 J860 J869 N10X N390 N300 N308 N309 N340 N151 N450 N459 N410 N412 N413 L00X L010 L011 L020 L021 L022 L023 L024 L028 L029 L030 L031 L032 L033 L038 L039 L040 L041 L042 L043 L048 L049 L050 L059 L080 L081 L088 L089 N700 N709 N710 N72X N730 N732 N733 N735 N760 N762 N764 N61X T814 G000 G001 G002 G003 G008 G009 G01X G020 G021 G028 G030 G038 G039 G040 G041 G042 G048 G049 G050 G051 G052 G058 G060 G061 G062 G07X G08X A851 M600 I330 I339 I300 I301 I308 I309 I400 I401 I408 I409 I514 I518 H700 K052 K113 J040 J041 J042 H600 H601 H603 H660 J010 J011 J012 J013 J014 J018 J019 J020 J028 J029 J030 J038 J039 M871 K102 M860 M861 M869 M000 M001 M002 M008 M009 K750 K610 K611 K612 K613 K614 K800 K803 K804 K810 K830 K630 K65 K65X
Neutropenic sepsis	*Defined where neutropenia & infection are recorded for the same admission.*
Gastrointestinal disorder	K521 K528 K529 A090 A099 R110 R111 R112 R11X R13X K590 K564 K121 K123 B370 K710 K711 K712 K716 K719 K720 K729 R17 R17X K221 K223 K251 K253 K255 K261 K262 K263 K265 K271 K273 K275 K281 K283 K285 K291 K293 K295 K631 K914 N321 N820 N822 N823 N824 K316 K603 K605 K604
Cardiovascular	
Arrhythmia	I440* I441* I442* I443* I444* I445* I446* I447* I471* I472* I480* I483* I484* I489* I48X* I450* I451* I452* I453* I454* I455* I456* I458* I459* I490* I491* I492* I493* I494* I495* I498* I499* R000 R001 R002 R008
Hypertension	I10* I10X* I110* I119* I120* I129* I130* I131* I132* I139* I150* I151* I152* I158* I159*
Angina	I200* I201* I208* I209*
Congestive cardiac failure	I500* I501* I509*
Cerebrovascular	I630* I631* I632* I633* I634* I635* I636* I638* I639* I600* I601* I602* I603* I604* I605* I606* I607* I608* I609* I64* I64X* I610* I611* I612* I613* I614* I615* I616* I618* I619* I620* I621* I629* I690* I691* I692* I693* I694* I698* G450* G451* G452* G453* G454* G458* G459* G460* G461* G462* G463* G464* G465* G466* G467* G468*
Other	I210 I211 I212 I213 I214 I219 I220 I221 I228 I229 I230 I231 I232 I233 I234 I235 I236 I238 I950 I951 I952 I958 I959 I260 I269 I313 I319 I427 I429 I740 I741 I742 I743 I744 I745 I748 I749 I822 I823 I828 I829 I800 I801 I802 I803 I808 I809
Pain	R100 R101 R102 R103 R104 M255 M540 M541 M542 M543 M544 M545 M546 M548 M549 R07 R07X R070 R071 R072 R073 R074 R520 R529 H920 K146 H571 M796
Psychological disorder	

Depression	F329*
Anxiety	F419*
Other	F320 F321 F322 F323 F328 F410 F411 F412 F413 F418
Constitutional	R530 R531 R538 R53X R64 R64X R630 R634 R638 E877 E860 E86X E861 E869 R600 R601 R609 R60X
Neurological	
Epilepsy	G400* G401* G402* G403* G404* G405* G406* G407* G408* G409* G410* G411* G412* G418* G419* R56* R560* R568*
Other	R55X R55 R42 R42X G620 G628 G629 R200 R201 R202 R203 R208 R209 H910 H931 J385 G250 G251 G252 G253 G258 G259 G240 G254 G256 G711 G720 R270 R260 G430 G431 G432 G433 G438 G439 G440 G441 G442 G443 G444 G448 R51 R51X
Metabolic/endocrine disorder	E870 E871 E872 E873 E874 E875 E876 E878 E833 E835 E838 E839 E883 E834 R730 R739 E15 E15X E160 E161 E162 E032 E058 E064 E273 E231
Renal disorder	N170 N171 N172 N178 N179 N19X N19 N10 N10X N12X N12 N130 N131 N132 N133 N134 N135 N136 N137 N138 N139 N141 N142 N144 N158 N159 N280
Line complications	T825 T827 T828 T829 Z452 T800 T801 T802 T808 T809
Dermatology & rheumatology	
Gout	M100* M102* M104* M109*
Other	R21X R21 L270 L271 L298 L299 L51 L510 L511 L512 L518 L519 L539 R238 R239
Respiratory disorder	R05X R05 J80X J80 J81 J81X R060
Bleeding	R040 R310 R31X N938 N939 R042 J942 K625 I850 K920 K921 K922 K250 K252 K254 K256 K260 K262 K264 K266 K270 K272 K274 K276 K280 K282 K284 K286 K290 K292 K294 K296
Ophthalmic disorder	
Retinal disorder	H300* H301* H302* H308* H309* H310* H311* H313* H314* H318* H319* H330* H332* H335* H340* H341* H342* H348* H349* H350* H352* H353* H356* H357* H358* H359*
Blindness/visual impairment	H540* H541* H542* H543* H544* H545* H546* H549*
Visual disorder	H490* H491* H492* H493* H494* H498* H499* H500* H501* H502* H503* H504* H505* H506* H508* H509* H510* H511* H512* H518* H519* H46X* H46* H470* H471* H472* H473* H474* H475* H476* H477*
Other	H320 H191 H192 H10 H100 H101 H102 H103 H105 H108 H109 H11 H111 H112 H113 B300 B301 B302 B303 B308 B309 H150 H151 H158 H159 H160 H161 H162 H163 H164 H168 H169 M350 H170 H171 H178 H179 H180 H181 H182 H183 H184 H186 H187 H188 H189 H200 H202 H208 H209 H210 H211 H212 H213 H214 H215 H218 H219 H263 H278 H279 H406 H531 H532 H533 H534 H535 H536 H538 H539 H000 H001 H010 H018 H019 H041 H042 H043 H020 H021 H050 H052 H058 H059 H578 H579 H431 H432 H433 H438 H439 H440 H441 H448 H449
Drug reaction	L500 T782 T783 T784 T886 T887 T451

*Codes excluded if present in the 12 months preceding treatment administration

Table A2: Baseline characteristics (inclusion/exclusion criteria) of patients in each adjuvant trastuzumab trial and associated routine data used to define trial eligibility within the cohort of women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer

Characteristics	Adjuvant Trastuzumab Trial						Data used to define trial eligibility (Yes/No) in routine care
	B31 AC-P vs AC-PH	**BCIRG006** AC-T vs AC-TH	**FinHer** T/V-FEC vs T/VH-FEC	**HERA** Obs vs H	**N9831** AC-P vs AC-PH	**PACS-04** FEC/ET-Obs/H	
Age (years)	22-78	22-74	25-65	18-79	19-82	22-65	Age at diagnosis
Sex	Female	Female	Female	Female	Female	Female	Gender
Performance status		Karnofsky PS >80%	WHO PS 0/1				WHO performance status 0/1
Hepatic function	Adequate	Normal	Exclude if abnormal	Adequate	Adequate		No Liver comorbidity
Renal function	Adequate	Normal		Adequate	Adequate		No Renal comorbidity
Cardiac function	LVEF met or exceeded lower limit of normal	Normal.		Exclude if baseline LVEF <55%.	LVEF met or exceeded lower limit of normal	Exclude if LVEF <50%	No MI or CCF comorbidity
Comorbidity	Exclude if history of CCF/MI/ cardiomyopathy	Exclude if any cardiac disease or diabetes	Exclude if severe hypertension or any cardiac disease, history of CCF /MI	Exclude if history of documented CCF	Exclude if history of CCF /MI/ cardiomyopathy	Exclude if signs of CCF	
Other unstable conditions incl. dementia.		Exclude if history of dementia					No dementia comorbidity
Hormone status			PR-negative				PR status record
HER2 status	HER2-positive	HER2-positive	HER2-positive	HER2-positive	HER2-positive	HER2-positive	HER2 status record
Tumour stage		Exclude if T4		Exclude if T4		Exclude if T4 or greater	Tumour stage
Nodal stage	N+	N+, high risk N0 (tumour size≥2cm). Exclude if N2/3	N+, or N0 with tumour size≥2cm	N+, or N0 with tumour size>1cm	N+, high risk N0 (tumour size>2cm OR >1cm & HR-negative)	N+	N stage (& tumour size for high risk)
Metastatic disease	Exclude if evidence of metastatic disease	Exclude if M1	Exclude if distant metastases	Exclude if distant metastases	Exclude if evidence of metastatic disease	Exclude if suspected metastases	Stage 4 record
Surgery	Complete resection of primary tumour	Definitive surgery of the breast cancer	Breast surgery	Complete excision of the cancer	Complete resection of primary tumour	Cancer completely surgically removed	BCS/mastectomy record
Other surgical procedure	Axillary-node dissection	Axillary lymph node assessment	Axillary-node dissection or sentinel-node biopsy		Axillary-node dissection		AND/SNB record
Prior		No prior systemic therapy		No prior anti-HER2 therapy			Neoadjuvant chemo or HER2 therapy recorded
Other unstable conditions incl. dementia.		No history of dementia					Dementia comorbidity recorded

Table A3: 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	

Table A4: Comparison of patient, tumour and treatment characteristics among women receiving adjuvant treatment for early invasive breast cancer, by HER2 status

		Total N = 18664		HER2-positive adjuvant trastuzumab N = 5087		HER2-negative adjuvant chemotherapy N = 13557	
Characteristic		N	%	N	%	N	%
Age group	50-59 yrs	8726	46.8%	2153	42.3%	6573	48.4%
	60-69 yrs	6943	37.2%	1906	37.5%	5037	37.1%
	70-79 yrs	2859	15.3%	941	18.5%	1918	14.1%
	80+ yrs	136	0.7%	87	1.7%	49	0.4%
IMD	1 - Most deprived	2957	15.8%	772	15.2%	2185	16.1%
	2	3296	17.7%	852	16.7%	2444	18.0%
	3	3935	21.1%	1088	21.4%	2847	21.0%
	4	4206	22.5%	1162	22.8%	3044	22.4%
	5 - Least deprived	4270	22.9%	1213	23.8%	3057	22.5%
Charlson Comorbidity Index	**0**	17100	91.6%	4609	90.6%	12491	92.0%
	1	1241	6.6%	381	7.5%	860	6.3%
	2+	323	1.7%	97	1.9%	226	1.7%
SCARF Index	Fit	16083	86.2%	4358	85.7%	11725	86.4%
	Mild frailty	1759	9.4%	474	9.3%	1285	9.5%
	Moderate - severe frailty	822	4.4%	255	5.0%	567	4.2%
Stage grouping	1	5455	29.2%	2365	46.5%	3090	22.8%
	2	10598	56.8%	2361	46.4%	8237	60.7%
	3A	2611	14.0%	361	7.1%	2250	16.6%
Grade of disease	G1	527	2.8%	98	1.9%	429	3.2%
	G2	7419	39.8%	1868	36.7%	5551	40.9%
	G3	10718	57.4%	3121	61.4%	7597	56.0%
Tumour stage	T1	8092	43.4%	2863	56.3%	5229	38.5%
	T2	9232	49.5%	2056	40.4%	7176	52.9%
	T3	1340	7.2%	168	3.3%	1172	8.6%
Nodal stage	N0	9487	50.8%	3499	68.8%	5988	44.1%
	N1	7187	38.5%	1291	25.4%	5896	43.4%
	N2	1990	10.7%	297	5.8%	1693	12.5%
Positive hormone-receptor status	Yes	12864	68.9%	3504	68.9%	9360	68.9%
	No/Unknown	5800	31.1%	1583	31.1%	4217	31.1%
Surgery type	BCS	12552	67.3%	3477	68.4%	9075	66.8%
	Mastectomy	6112	32.7%	1610	31.6%	4502	33.2%
Radiotherapy reported/setting	No	2777	14.9%	1055	20.7%	1722	12.7%
	Yes – before treatment	335	1.8%	301	5.9%	34	0.3%
	Yes – during treatment	2958	15.8%	2958	58.1%	0	0.0%
	Yes – after treatment	12594	67.5%	773	15.2%	11821	87.1%
Hormone therapy prescribed (Y)		12825	68.7%	3508	69.0%	9317	68.6%
Chemotherapy reported (Y)		18592	99.6%	5015	98.6%	13577	100.0%
	Anthracycline only (Y)	5079	27.3%	911	18.2%	4168	30.7%
	Taxane only (Y)	3463	18.6%	2070	41.3%	1393	10.3%
	Docetaxel (Y)	9626	51.8%	2484	49.5%	7142	52.6%
	Paclitaxel (Y)	4046	21.8%	1639	32.7%	2407	17.7%
	Anthracycline and taxane (Y)	9904	53.3%	1988	39.6%	7916	58.3%

Note: Anthracyclines = doxorubicin, epirubicin, mitoxantrone recorded in SACT.

Taxanes = docetaxel, cabazitaxel, paclitaxel, nab-paclitaxel recorded in SACT.

Table A5: Frequency of severe acute toxicity events (SATE), overall / by individual SATE type, among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer, by sequential or concurrent use of chemotherapy *(ordered by the most frequently recorded; only individual SATE with >5% presented)*

Event	Total N = 5087		Sequential chemotherapy N = 1160		Concurrent chemotherapy N = 3927	
	N	%	N	%	N	%
Any	1670	32.8%	402	34.7%	1268	32.3%
Haematological	774	15.2%	213	18.4%	561	14.3%
Neutropenia	*714*	*14.0%*	*205*	*17.7%*	*509*	*13.0%*
Anaemia	*103*	*2.0%*	*26*	*2.2%*	*77*	*2.0%*
Thrombocytopenia	*15*	*0.3%*	*2*	*0.2%*	*13*	*0.3%*
Infection	773	15.2%	151	13.0%	622	15.8%
Neutropenic sepsis	659	13.0%	188	16.2%	471	12.0%
Gastrointestinal	507	10.0%	100	8.6%	407	10.4%
Cardiovascular	346	6.8%	102	8.8%	244	6.2%
Arrhythmia	*106*	*2.1%*	*29*	*2.5%*	*77*	*2.0%*
Hypertension	*73*	*1.4%*	*16*	*1.4%*	*57*	*1.5%*
Angina	*7*	*0.1%*	*0*	*0.0%*	*7*	*0.2%*
Congestive cardiac failure	*25*	*0.5%*	*7*	*0.6%*	*18*	*0.5%*
Cerebrovascular	*14*	*0.3%*	*7*	*0.6%*	*7*	*0.2%*
Other	*175*	*3.4%*	*64*	*5.5%*	*111*	*2.8%*

Table A6: Frequency of severe acute toxicity events (SATE), overall / by individual SATE type, among women receiving adjuvant treatment for early invasive breast cancer, by HER2 status *(ordered by the most frequently recorded; only individual SATE with >5% presented)*

Event	Total N = 18664		HER2-positive adjuvant trastuzumab N = 5087		HER2-negative adjuvant chemotherapy N = 13557	
	N	%	N	%	N	%
Any	6784	36.3%	1670	32.8%	5115	37.7%
Haematological	3430	18.4%	774	15.2%	2656	19.6%
Neutropenia	*3212*	*17.2%*	*714*	*14.0%*	*2498*	*18.4%*
Anaemia	*415*	*2.2%*	*103*	*2.0%*	*312*	*2.3%*
Thrombocytopenia	*98*	*0.5%*	*15*	*0.3%*	*83*	*0.6%*
Infection	3052	16.4%	773	15.2%	2279	16.8%
Neutropenic sepsis	2918	15.6%	659	13.0%	2259	16.6%
Gastrointestinal disorder	2171	11.6%	507	10.0%	1668	12.3%
Cardiovascular	1505	8.1%	346	6.8%	1159	8.5%
Arrhythmia	*492*	*2.6%*	*106*	*2.1%*	*386*	*2.8%*
Hypertension	*309*	*1.7%*	*73*	*1.4%*	*236*	*1.7%*
Angina	*42*	*0.2%*	*7*	*0.1%*	*35*	*0.3%*
Congestive cardiac failure	*109*	*0.6%*	*25*	*0.5%*	*84*	*0.6%*
Cerebrovascular	*66*	*0.4%*	*14*	*0.3%*	*52*	*0.4%*
Other	*751*	*4.0%*	*175*	*3.4%*	*576*	*4.2%*
Pain	954	5.1%	220	4.3%	734	5.4%

Table A7: Frequency of severe acute toxicity events (SATE), overall / by individual SATE type, among women receiving adjuvant trastuzumab-based treatment for HER2-positive early invasive breast cancer, by comorbidity burden *(ordered by the most frequently recorded; only individual SATE with >5% presented)*

Event	CCI 0 N = 4609		CCI 1 N = 387		CCI 2+ N = 97	
	N	%	N	%	N	%
Any	1461	31.7%	161	42.3%	48	49.5%
Haematological	694	15.1%	64	16.8%	16	16.5%
Neutropenia	645	14.0%	56	14.7%	13	13.4%
Anaemia	84	1.8%	15	3.9%	4	4.1%
Thrombocytopenia	11	0.2%	3	0.8%	1	1.0%
Infection	666	14.4%	79	20.7%	28	28.9%
Neutropenic sepsis	596	12.9%	50	13.1%	13	13.4%
Gastrointestinal disorder	433	9.4%	55	14.4%	19	19.6%
Cardiovascular	291	6.3%	38	10.0%	17	17.5%
Arrhythmia	90	2.0%	11	2.9%	5	5.2%
Hypertension	64	1.4%	8	2.1%	1	1.0%
Angina	1	0.0%	5	1.3%	1	1.0%
Congestive cardiac failure	15	0.3%	6	1.6%	4	4.1%
Cerebrovascular	12	0.3%	2	0.5%	0	0.0%
Other	152	3.3%	14	3.7%	9	9.3%
Pain	182	3.9%	28	7.3%	10	10.3%
Psychological disorder	102	2.2%	12	3.1%	3	3.1%
Depression	59	1.3%	8	2.1%	3	3.1%
Anxiety	41	0.9%	2	0.5%	0	0.0%
Other	19	0.4%	4	1.0%	0	0.0%
Constitutional	128	2.8%	25	6.6%	10	10.3%
Neurological	125	2.7%	18	4.7%	6	6.2%
Epilepsy	7	0.2%	1	0.3%	0	0.0%
Other	121	2.6%	17	4.5%	6	6.2%
Metabolic/endocrine disorder	104	2.3%	23	6.0%	7	7.2%
Renal disorder	64	1.4%	16	4.2%	11	11.3%

Key: CCI = Charlson Comorbidity Index.

Table A8: Distribution of patient, tumour and treatment characteristics among women receiving adjuvant trastuzumab-based treatment for HER2-positive early invasive breast cancer, and associated odds of any severe acute toxicity events (SATE) or cardiovascular SATE.

		All	Any SATE			Cardiovascular SATE		
Characteristic		**N**	**N (%)**	**OR (95% CI)**	**Overall p-value**	**N (%)**	**OR (95% CI)**	**Overall p-value**
Total		5087	1670 (32.8%)	-		346 (6.8%)	-	
Age group	50-59 years	2153	710 (33.0%)	1.00	0.615	132 (6.1%)	1.00	0.024
	60-69 years	1906	625 (32.8%)	1.00 (0.87-1.14)		125 (6.6%)	1.04 (0.80-1.35)	
	70-79 years	1028	335 (32.6%)	1.08 (0.91-1.28)		89 (8.7%)	1.46 (1.09-1.95)	
IMD	1 - Most deprived	772	302 (39.1%)	1.38 (1.13-1.69)	0.012	68 (8.8%)	1.53 (1.07-2.2)	0.077
	4	852	290 (34.0%)	1.12 (0.93-1.37)		63 (7.4%)	1.26 (0.88-1.82)	
	2	1088	349 (32.1%)	1.05 (0.87-1.26)		62 (5.7%)	1.01 (0.71-1.45)	
	3	1162	359 (30.9%)	1.01 (0.84-1.20)		87 (7.5%)	1.35 (0.97-1.89)	
	5 - Least deprived	1213	370 (30.5%)	1.00		66 (5.4%)	1.00	
CCI	0	4609	1461 (31.7%)	1.00	0.938	291 (6.3%)	1.00	0.680
	1	381	161 (42.3%)	1.08 (0.68-1.70)		38 (10%)	0.69 (0.29-1.60)	
	2+	97	48 (49.5%)	1.05 (0.45-2.44)		17 (17.5%)	0.55 (0.13-2.44)	
History of MI	No	5070	1660 (32.7%)	1.00	0.091	340 (6.7%)	1.00	0.002
	Yes	17	10 (58.8%)	2.53 (0.86-7.43)		6 (35.3%)	7.84 (2.17-28.34)	
History of CCF	No	5070	1662 (32.8%)	*No crude association*		343 (6.8%)	1.00	0.330
	Yes	17	8 (47.1%)	-		3 (17.6%)	2.16 (0.46-10.23)	
History of diabetes	No	4925	1597 (32.4%)	1.00	0.575	328 (6.7%)	1.00	0.278
	Yes	162	73 (45.1%)	1.16 (0.70-1.91)		18 (11.1%)	1.62 (0.68-3.88)	
History of liver disease	No	5055	1652 (32.7%)	1.00	0.150	339 (6.7%)	1.00	0.009
	Yes	32	18 (56.3%)	1.82 (0.81-4.11)		7 (21.9%)	1.62 (0.68-3.88)	
History of CPD	No	4858	1570 (32.3%)	1.00	0.418	315 (6.5%)	1.00	0.024
	Yes	229	100 (43.7%)	1.23 (0.75-2.02)		31 (13.5%)	2.67 (1.14-6.25)	
History of renal disease	No	5049	1651 (32.7%)	1.00	0.204	343 (6.8%)	*No crude association*	
	Yes	38	19 (50%)	1.65 (0.76-3.58)		3 (7.9%)	-	
SCARF Index	Fit	4358	1364 (31.3%)	1.00	0.005	278 (6.4%)	1.00	0.490
	Mild frailty	474	193 (40.7%)	1.42 (1.13-1.79)		35 (7.4%)	0.92 (0.59-1.42)	
	Moderate - severe frailty	255	113 (44.3%)	1.47 (1.03-2.10)		33 (12.9%)	1.30 (0.73-2.34)	
Anthracycline chemotherapy	No	2116	593 (28%)	1.00	<0.0001	135 (6.4%)	*No crude association*	
	Yes	2899	1065 (36.7%)	1.55 (1.36-1.78)		207 (7.1%)	-	
	No chemotherapy	72	12 (16.7%)	0.47 (0.24-0.89)		4 (5.6%)	-	
Sequential chemotherapy	No	3916	1263 (32.3%)	*No crude association*		242 (6.2%)	1.00	0.003
	Yes	1099	395 (35.9%)	-		100 (9.1%)	1.54 (1.19-1.98)	
	No chemotherapy	72	12 (16.7%)	-		4 (5.6%)	0.69 (0.24-2.03)	

Key: IMD = Index of Multiple Deprivation; CCI = Charlson Comorbidity Index; MI = myocardial infarction; CCF = congestive cardiac failure; CPD = Chronic pulmonary disease; SCARF = Secondary Care Administrative Records Frailty; chemo = chemotherapy.
Note: ORs adjusted for other factors and year of diagnosis.

Table A9: Relationship between adjuvant trastuzumab for HER2-positive early invasive breast cancer and any severe acute toxicity events (SATE).

	Total	50-59 years	60-69 years	70-79 years	80+ years
Cycle-based analysis	**N = 5087**	**N = 2153**	**N = 1906**	**N = 941**	**N = 87**
Experienced SATE AND					
No further cycles (discontinued)	67 (1.3%)	21 (1.0%)	14 (0.7%)	25 (2.7%)	7 (8.0%)
Delay/break before next cycle	402 (7.9%)	147 (6.8%)	157 (8.2%)	91 (9.7%)	7 (8.0%)
Delay/break after any subsequent cycle	25 (0.5%)	9 (0.4%)	12 (0.6%)	4 (0.4%)	0 (0.0%)
No delay/break before end of treatment	4 (0.1%)	1 (0.0%)	1 (0.1%)	1 (0.1%)	1 (1.1%)
No SATE recorded AND					
Delay/break before next cycle	3323 (65.3%)	1414 (65.7%)	1258 (66.0%)	590 (62.7%)	61 (70.1%)
Patient-based analysis					
Experienced SATE AND					
Treatment completed	586 (11.5%)	256 (11.9%)	221 (11.6%)	102 (10.8%)	7 (8.0%)
Treatment included delays/breaks	866 (17.0%)	344 (16.0%)	330 (17.3%)	176 (18.7%)	16 (18.4%)
Treatment continued* after SATE but stopped early	1 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Treatment stopped after SATE	67 (1.3%)	21 (1.0%)	14 (0.7%)	25 (2.7%)	7 (8.0%)
No SATE recorded AND					
Treatment completed	1823 (35.8%)	783 (36.4%)	689 (36.1%)	324 (34.4%)	27 (31.0%)
Treatment included delays/breaks	2629 (51.7%)	1132 (52.6%)	989 (51.9%)	461 (49.0%)	47 (54.0%)
Treatment discontinued	2107 (41.4%)	897 (41.7%)	792 (41.6%)	379 (40.3%)	39 (44.8%)

*With no delays/breaks

Table A10: Distribution of patient, tumour and treatment characteristics among women receiving adjuvant trastuzumab for HER2-positive early invasive breast cancer, diagnosed in NHS trusts in England between January 2014 and December 2019, by trial eligibility.

		Total N = 5087		Not trial eligible N = 3274		Trial eligible N = 1813	
Characteristic		N	%	N	Row %	N	Row %
Age group	50-59 years	2153	42.3%	1365	63.4%	788	36.6%
	60-69 years	1906	37.5%	1262	66.2%	644	33.8%
	70-79 years	941	18.5%	585	62.2%	356	37.8%
	80+ years	87	1.7%	62	71.3%	25	28.7%
IMD	1 - Most deprived	772	15.2%	477	61.8%	295	38.2%
	2	852	16.7%	529	62.1%	323	37.9%
	3	1088	21.4%	702	64.5%	386	35.5%
	4	1162	22.8%	756	65.1%	406	34.9%
	5 - Least deprived	1213	23.8%	810	66.8%	403	33.2%
Charlson Comorbidity Index	**0**	4609	90.6%	2925	63.5%	1684	36.5%
	1	381	7.5%	268	70.3%	113	29.7%
	2+	97	1.9%	81	83.5%	16	16.5%
SCARF Index	Fit	4358	85.7%	2764	63.4%	1594	36.6%
	Mild frailty	474	9.3%	316	66.7%	158	33.3%
	Moderate - severe frailty	255	5.0%	194	76.1%	61	23.9%
Stage grouping	1	2365	46.5%	1973	83.4%	392	16.6%
	2	2361	46.4%	1258	53.3%	1103	46.7%
	3A	361	7.1%	43	11.9%	318	88.1%
Grade of disease	G1	98	1.9%	74	75.5%	24	24.5%
	G2	1868	36.7%	1260	67.5%	608	32.5%
	G3	3121	61.4%	1940	62.2%	1181	37.8%
Tumour stage	T1	2863	56.3%	2110	73.7%	753	26.3%
	T2	2056	40.4%	1104	53.7%	952	46.3%
	T3	168	3.3%	60	35.7%	108	64.3%
Nodal stage	N0	3499	68.8%	2790	79.7%	709	20.3%
	N1	1291	25.4%	456	35.3%	835	64.7%
	N2	297	5.8%	28	9.4%	269	90.6%
Positive hormone-receptor status	Yes	3504	68.9%	2306	65.8%	1198	34.2%
	No/Unknown	1583	31.1%	968	61.1%	615	38.9%
Surgery type	BCS	3477	68.4%	2414	69.4%	1063	30.6%
	Mastectomy	1610	31.6%	860	53.4%	750	46.6%
Radiotherapy reported/setting	No	1055	20.7%	725	68.7%	330	31.3%
	Yes – before treatment	301	5.9%	76	25.2%	225	74.8%
	Yes – during treatment	2958	58.1%	1951	66.0%	1007	34.0%
	Yes – after treatment	773	15.2%	522	67.5%	251	32.5%
Hormone therapy prescribed	No	1579	31.0%	981	62.1%	598	37.9%
	Yes	3508	69.0%	2293	65.4%	1215	34.6%
Chemotherapy reported	No	72	1.4%	54	75.0%	18	25.0%
	Yes	5015	98.6%	3220	64.2%	1795	35.8%
	Anthracycline only – No	4104	81.8%	2956	72.0%	1148	28.0%
	Anthracycline only – Yes	911	18.2%	264	29.0%	647	71.0%
	Taxane only – No	2945	58.7%	1519	51.6%	1426	48.4%
	Taxane only – Yes	2070	41.3%	1701	82.2%	369	17.8%
	Docetaxel – No	2531	50.5%	1573	62.1%	958	37.9%
	Docetaxel – Yes	2484	49.5%	1647	66.3%	837	33.7%
	Paclitaxel – No	3376	67.3%	1899	56.3%	1477	43.8%
	Paclitaxel – Yes	1639	32.7%	1321	80.6%	318	19.4%
	Anthracycline and taxane – No	3027	60.4%	1989	65.7%	1038	34.3%
	Anthracycline and taxane – Yes	1988	39.6%	1231	61.9%	757	38.1%

Note: Anthracyclines = doxorubicin, epirubicin, mitoxantrone recorded in SACT.
Taxanes = docetaxel, cabazitaxel, paclitaxel, nab-paclitaxel recorded in SACT.

9.7. Appendix 7 – Supplementary material for Research Paper 6

Figure S1: Details of patient cohort selection from women aged 50 and over, diagnosed with breast cancer in a NHS organisation in England between 2014 and 2019.

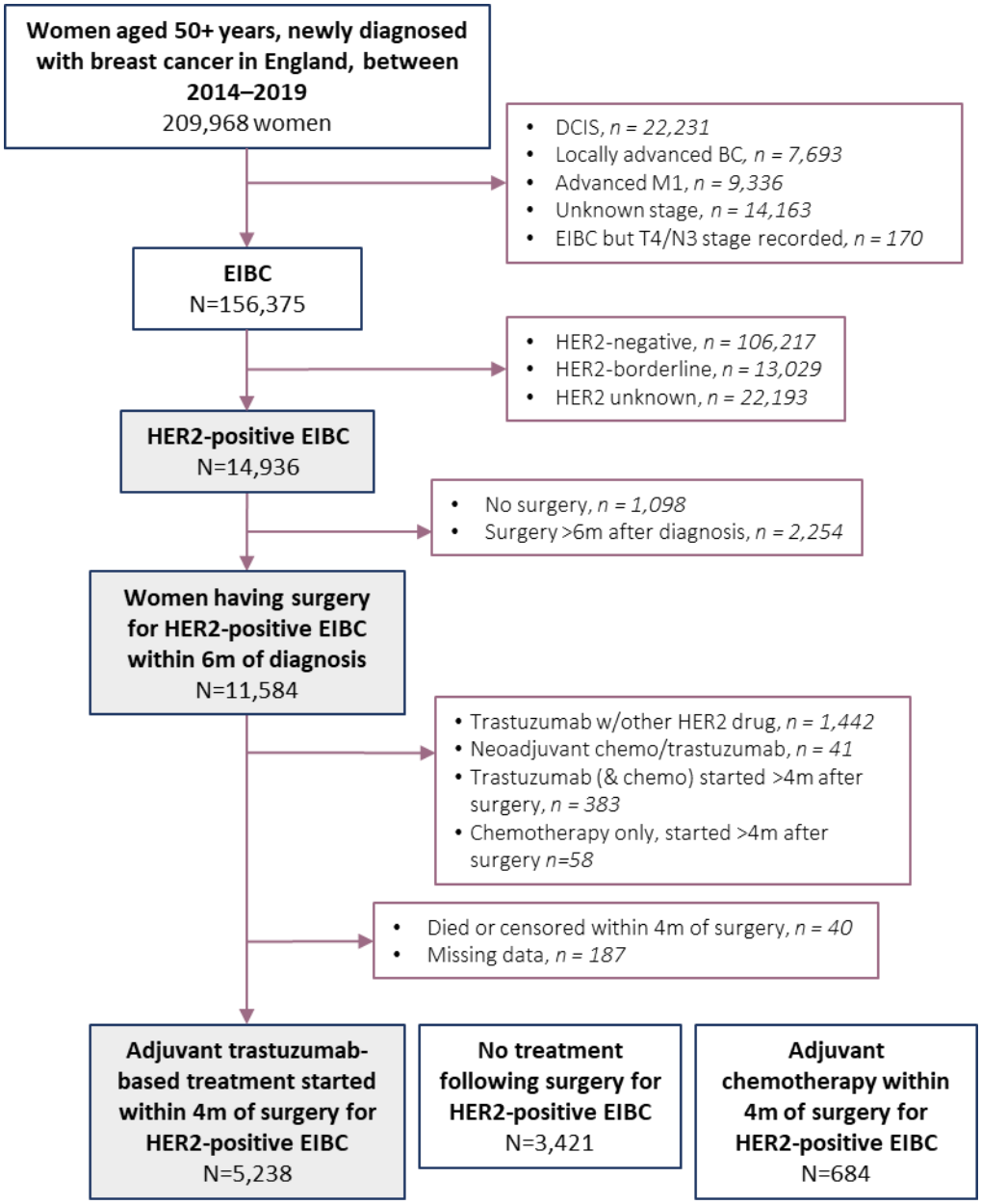


Figure S2: Kaplan-Meier survival curves (including 95% confidence intervals) for overall survival following initiation of adjuvant trastuzumab-based treatment, overall and by patient and tumour characteristics.

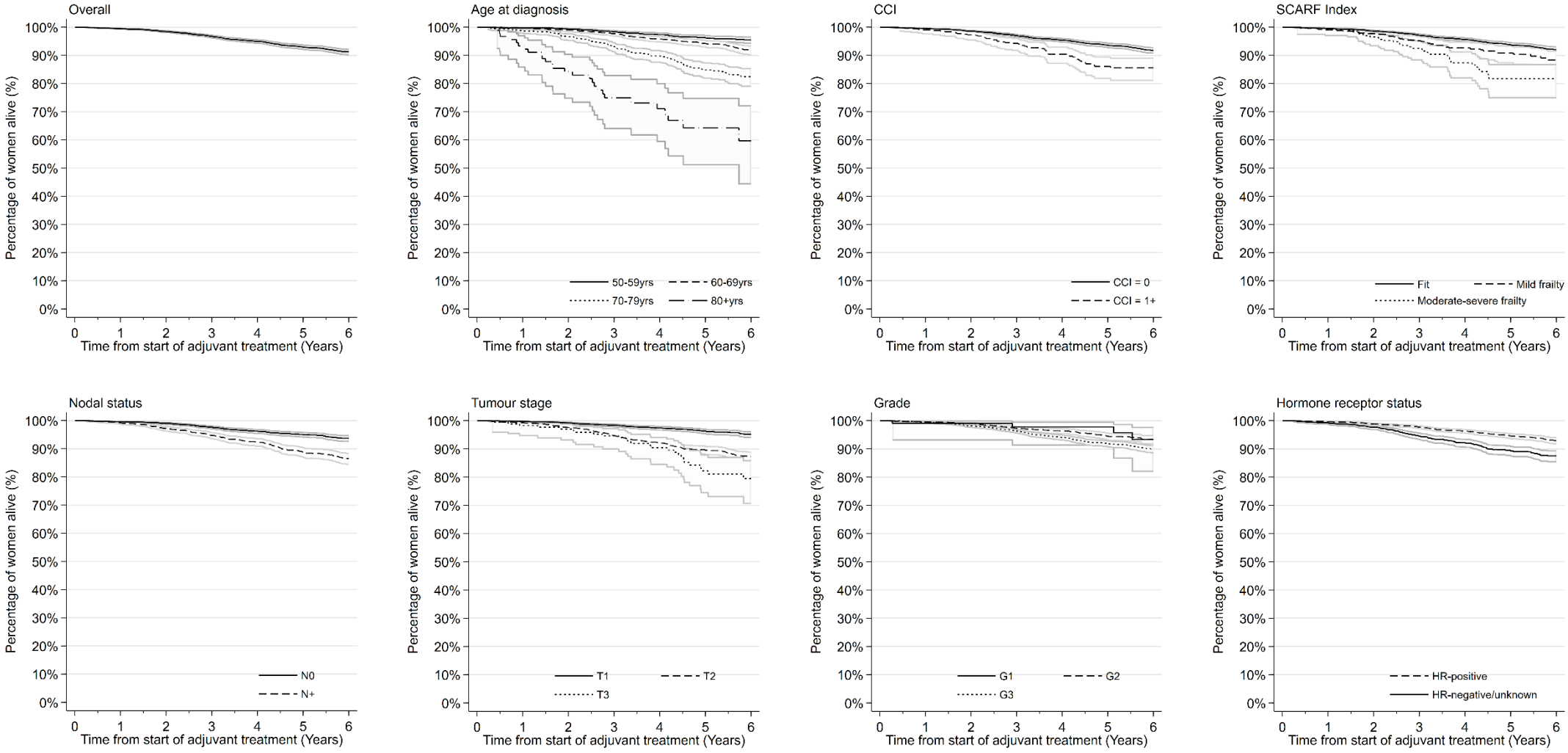


Figure S3: Balance obtained with standardised means difference, while accounting, or not, for selection bias (i.e. weighted and unweighted, respectively)

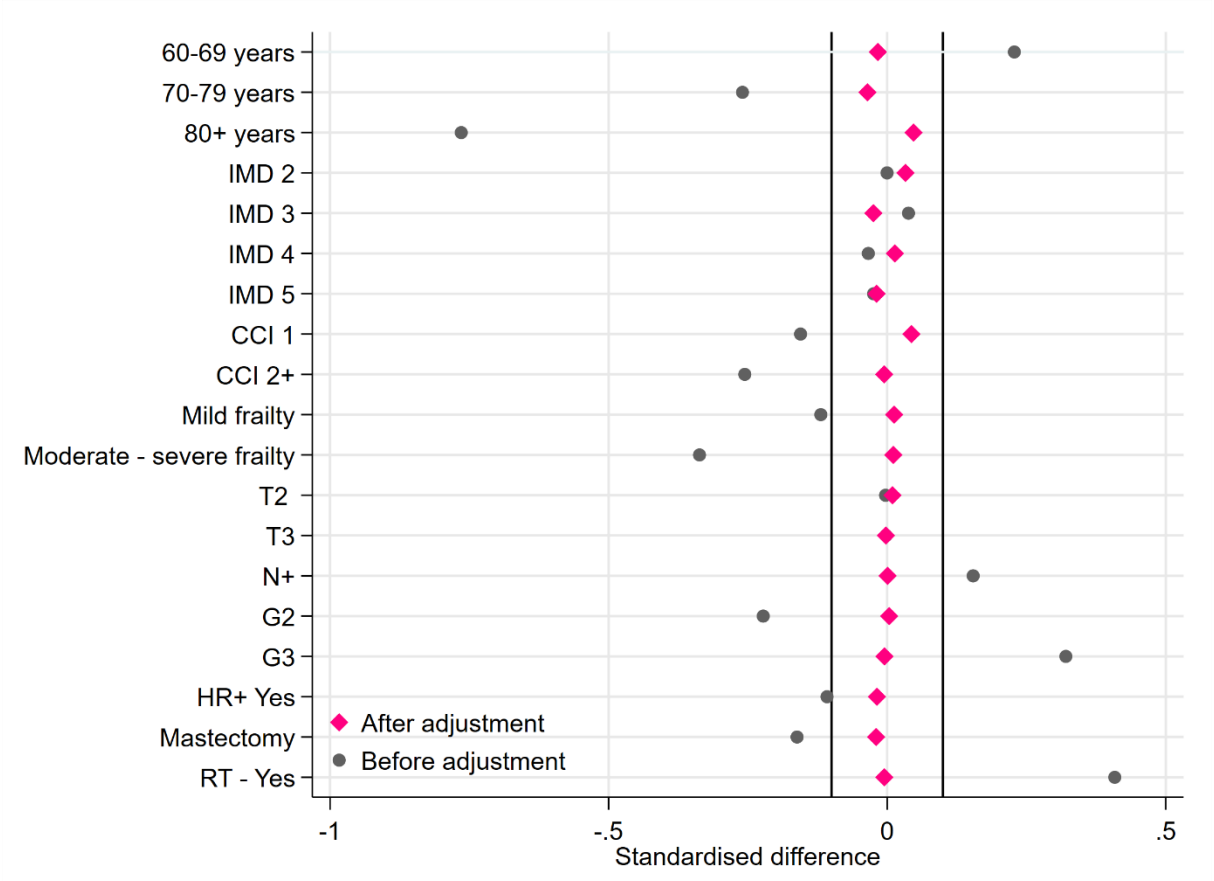


Figure S4: Density plot of propensity score by treatment group

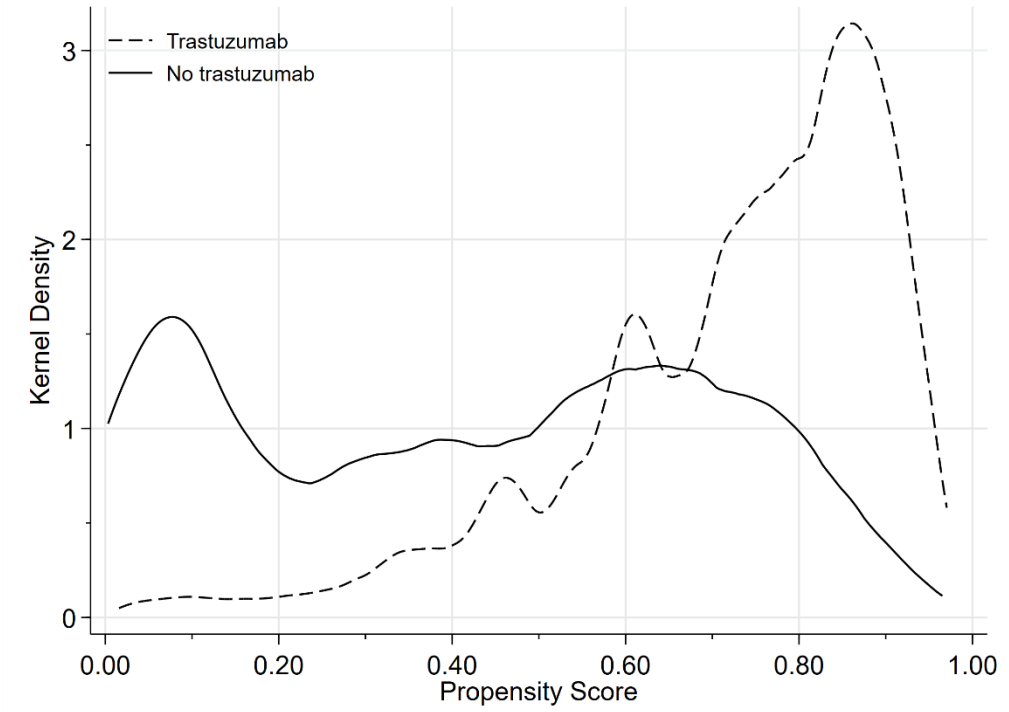


Table S1: Specification and emulation of a target trial of adjuvant trastuzumab versus no treatment among patients with HER2-positive EIBC between 2014 and 2019 in England

Component	Description of Target trial	Description of Emulated trial using routine healthcare data
Eligibility criteria	Patients with EIBC who receive surgery, with no prior use of chemotherapy or trastuzumab.	Same as target trial
Treatment strategies	1. Initiate trastuzumab-based treatment. 2. Don't initiate trastuzumab-based treatment.	Same as target trial
Treatment assignment	Patients are randomly assigned to either strategy.	We classified individuals according to the strategy their data were compatible with. Randomisation was assumed conditional on baseline covariates, using propensity scores. A 4-month grace period from date of surgery was specified to allow for decision-making (landmark approach).
Follow up	Starts at randomisation and ends at the point of death or administrative censoring.	Starts at landmark time point and ends at the point of death or administrative censoring
Outcome	Death from any cause; Death from breast cancer.	Same as target trial
Causal contrast of interest	Intention-to-treat.	Same as target trial. To be analogous to the target trial, comparison will be of treatment initiation. Some patients allocated to strategy 1 may have been prescribed treatment but never initiated it.
Analysis plan	Intention-to-treat effect estimated via standard survival methods.	Same as target trial. Propensity scores used for balance of baseline prognostic factors.

Table S2: Overall survival by patient, tumour and treatment characteristics among women receiving adjuvant trastuzumab-based treatment for HER2-positive, early invasive breast cancer.

	N	% died	5 year OS (95% CI)	Unadjusted HR	Adjusted HR*	Grouped p-value
Overall	**5238**	**6.5%**	**92.9% (92.1-93.7)**	-	-	-
Age						
50-59 years	2226	3.4%	96.2% (95.2-97.0)	1.00	1.00	<0.0001
60-69 years	1952	5.8%	94.2% (92.9-95.3)	1.72 (1.28-2.3)	1.62 (1.21-2.17)	
70-79 years	970	12.6%	84.8% (81.9-87.3)	4.17 (3.13-5.56)	2.99 (2.18-4.11)	
80+ years	90	31.1%	64.3% (51.1-74.8)	13.76 (8.75-21.63)	6.41 (3.89-10.57)	
IMD						
1 - Most deprived	794	8.2%	91.4% (88.8-93.4)	1.00	*No crude association*	
4	886	7.2%	92.2% (89.9-94.0)	0.93 (0.66-1.31)		
2	1119	5.4%	94.1% (92.3-95.5)	0.67 (0.47-0.96)		
3	1200	6.7%	92.6% (90.7-94.1)	0.84 (0.60-1.16)		
5 - Least deprived	1239	5.6%	93.7% (91.8-95.1)	0.71 (0.50-0.99)		
CCI						
0	4744	6.0%	93.6% (92.7-94.3)	1.00	1.00	0.0053
1+	494	10.3%	86.1% (81.8-89.4)	1.98 (1.47-2.67)	1.55 (1.14-2.11)	
SCARF Index						
Fit	4478	5.9%	93.7% (92.8-94.5)	1.00	*No adjusted association*	
Mild frailty	492	8.3%	90.8% (87.3-93.3)	1.54 (1.11-2.15)		
Mod - severe frailty	268	12.7%	81.7% (74.9-86.8)	2.74 (1.91-3.92)		
Grade						
G1	100	4.0%	97.8% (91.3-99.4)	1.00	*No adjusted association*	
G2	1925	4.7%	94.5% (93.1-95.6)	1.24 (0.46-3.36)		
G3	3213	7.6%	91.8% (90.7-92.9)	1.92 (0.72-5.13)		
Tumour stage						
T1	2954	3.5%	96.3% (95.4-97.0)	1.00	1.00	<0.0001
T2	2113	9.9%	89.5% (87.9-90.9)	2.75 (2.17-3.48)	2.07 (1.62-2.65)	
T3	171	15.2%	82.2% (74.4-87.9)	4.17 (2.72-6.39)	2.60 (1.64-4.12)	
Nodal stage						
N0	3633	4.4%	95.0% (94.1-95.8)	1.00	1.00	<0.0001
N+	1605	11.1%	88.8% (86.9-90.4)	2.26 (1.82-2.79)	2.00 (1.58-2.53)	
HR-positive						
Yes	3601	5.0%	94.6% (93.6-95.4)	0.53 (0.43-0.66)	0.61 (0.49-0.75)	<0.0001
No/Unknown	1637	9.6%	89.3% (87.5-90.9)	1.00	1.00	
Surgery type						
BCS	3577	4.5%	95.0% (94.1-95.8)	1.00	*No adjusted association*	
Mastectomy	1661	10.6%	88.7% (86.9-90.3)	2.25 (1.82-2.79)		
Radiotherapy						
No	1082	8.7%	90.8% (88.7-92.6)	1.00	1.00	0.0012
Yes	4156	5.9%	93.5% (92.5-94.3)	0.67 (0.53-0.85)	0.67 (0.52-0.85)	
Endocrine therapy						
No	1627	9.8%	89.3% (87.4-90.9)	1.00	*No adjusted association*	
Yes	3611	4.9%	94.5% (93.6-95.4)	0.48 (0.39-0.60)		
Chemotherapy						
No	72	18.1%	80.6% (67.9-88.7)	1.00	1.00	0.0017
Other chemo therapy	51	9.8%	87.9% (72.8-94.9)	0.45 (0.16-1.27)	0.69 (0.25-1.89)	
Taxanes	2120	6.7%	91.4% (89.8-92.8)	0.37 (0.20-0.66)	0.49 (0.28-0.86)	
Anthracyclines	969	8.0%	92.5% (90.5-94.1)	0.33 (0.18-0.61)	0.63 (0.35-1.14)	
Taxane & anthracycline	2026	4.9%	94.9% (93.7-95.9)	0.22 (0.12-0.41)	0.37 (0.21-0.67)	

Key: OS = overall survival; CI = confidence interval; HR = hazard ratio; IMD = Index of Multiple Deprivation; CCI = Charlson Comorbidity Index; SCARF = Secondary Care Administrative Records Frailty; HR = hormone receptor; BCS = breast conserving surgery.

Note: Anthracyclines = doxorubicin, epirubicin, mitoxantrone recorded in SACT.

Taxanes = docetaxel, cabazitaxel, paclitaxel, nab-paclitaxel recorded in SACT.

* HRs adjusted for other factors