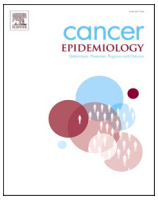
Cancer Epidemiology 83 (2023) 102337

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.

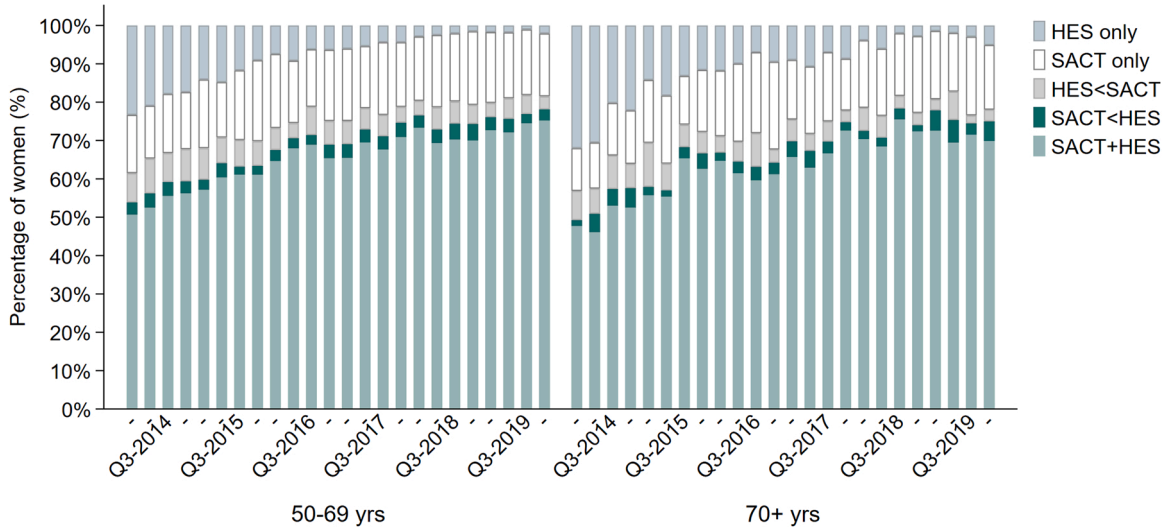


Fig. 1. Percentage with CDT details recorded in SACT or HES-APC (and agreement on first cycle date), among women receiving surgery for early invasive breast cancer, by age and date of diagnosis. **Key**: CDT = Cancer Drug Therapy; SACT = Systemic Anti-Cancer Therapy data; HES-APC = Hospital Episodes 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.

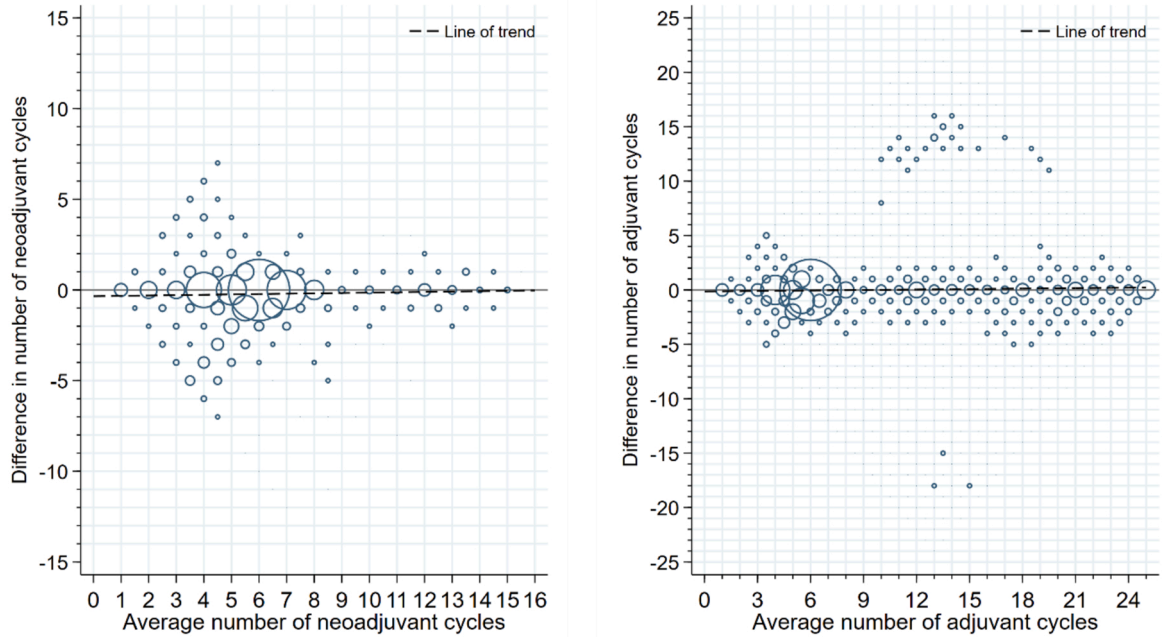


Fig. 3. Weighted scatter plot of agreement between the average numbers of CDT cycles recorded in SACT and HES-APC, and the difference in number of cycles recorded in each source, among women receiving CDT for early invasive breast cancer, by CDT setting. Key: CDT = Cancer Drug Therapy; SACT = Systemic 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.