Validity of chemotherapy information derived from routinely collected healthcare data: a national cohort study of colon cancer patients

Jemma M. Boyle[a,b], Angela Kuryba[b], Michael S. Braun[c] PhD, Ajay Aggarwal[d] PhD, Jan van der Meulen[a,b] PhD, Thomas E. Cowling[a,b] PhD & Kate Walker[a,b*] PhD.

[a]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK [b]Clinical Effectiveness Unit, Royal College of Surgeons of England, London, UK, [c]Department of Oncology, The Christie NHS Foundation Trust, Manchester, UK, and [d]Department of Oncology, Guy's and St. Thomas' NHS Foundation Trust, London, UK.

*Kate Walker is the senior author

Corresponding author: Dr Jemma M. Boyle, Clinical Effectiveness Unit, Royal College of Surgeons of England, London, UK. WC2A 3PE. Tel: 020 7869 6624 E-mail: jboyle@rcseng.ac.uk

A Kuryba – akuryba@rcseng.ac.uk
M Braun – Michael.braun@nhs.net
A Aggarwal – ajay.aggarwal@lshtm.ac.uk
J van der Meulen – jan.vandermeulen@lshtm.ac.uk
T Cowling – Thomas.cowling@lshtm.ac.uk
K Walker – kate.walker@lshtm.ac.uk

Declarations of interest*:* None.

Funding statement - The National Bowel Cancer 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).

T.E.C was supported by the Medical Research Council (grant number MR/S020470/1).

A.A was supported by a National Institute for Health Research (NIHR) Advanced Fellowship (NIHR300599).

The above sponsors 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.

Abstract – 271 (300 limit)

Word count – 3,009 (3000 limit)

KEYWORDS: chemotherapy, colon cancer, validation, administrative data, routinely collected data

ABSTRACT

Background: We used a structured approach to validate chemotherapy information derived from a national routinely collected chemotherapy dataset and from national administrative hospital data.

Methods: 10,280 patients who had surgical resection with stage III colon cancer were included. First, we compared information derived from the national chemotherapy dataset (SACT) and from the administrative hospital dataset (HES) in the English NHS with respect to receipt of adjuvant chemotherapy (ACT). Second, we compared regimen and number of cycles in linked patient-level records. Third, we carried out a sensitivity analysis to establish to what extent the impact of ACT receipt differed according to data source.

Results: 6,012 patients (58%) received ACT according to either dataset. Of these patients, 3,460 (58%) had ACT records in both datasets, 1,649 (27%) in SACT alone, and 903 (15%) in HES alone. Of the 3,460 patients with records in both datasets, 3,320 (96%) had matching regimens. There was good agreement on cycle number with similar proportions of patients recorded with a single cycle (6% in SACT vs. 7% in HES) and slightly fewer patients recorded with more than 8 cycles in SACT (32% in SACT vs. 35% in HES). 3-year cancer-specific mortality was similar for patients receiving ACT, regardless of whether a patient received ACT according to SACT alone (16.6%), according to HES alone (16.8%), or according to either SACT or HES (17.1%).

Conclusion: Routinely collected national chemotherapy data and administrative hospital data are highly accurate in recording regimen and number of chemotherapy cycles. However, chemotherapy information should ideally be captured from both datasets to avoid under-capture, particularly of oral chemotherapy from administrative hospital data, and to minimise bias.

1. INTRODUCTION

Chemotherapy is a critical component of oncological treatment. Evidence regarding the efficacy of chemotherapy treatment has come from high quality, large randomised controlled trials (RCTs).[1-3] RCTs, however, include highly selected patient populations under rigorously controlled conditions, generally under-representing older patients, and those who are frail or comorbid. Population-based studies, using data such as electronic healthcare records, are needed to assess outcomes in diverse non-selected populations under realistic clinical conditions, and can be used to complement RCT findings.[4-8]

All English National Health Service (NHS) providers are mandated to collect chemotherapy data for all patients in routine care via the Systemic Anti-Cancer Therapy (SACT) dataset.[9] The use of this dedicated chemotherapy dataset for research has been limited. Several studies have highlighted possible data issues, for example that older patients and those with comorbidities are not fully represented within the dataset, and that there might be limitations in the accurate recording of chemotherapy cycle numbers, particularly with oral drugs.[10-15] The only study to date that has attempted to validate SACT data was carried out in a study using general practice records of only 7% of the UK population, and no validation of chemotherapy regimens was attempted.[15]

This study aimed to validate chemotherapy data in a contemporary national cohort of patients with pathological stage III colon cancer, identified from the National Bowel Cancer Audit (NBOCA), who had undergone potentially curative surgical resection and were candidates for adjuvant chemotherapy (ACT) according to national guidelines.[16]

We used a structured four-step framework to compare national chemotherapy data with data available in the Hospital Episode Statistics, a national administrative dataset of all hospital admissions in the English NHS. First, we assessed the agreement between the two datasets for chemotherapy receipt in all patients. Second, we compared the chemotherapy regimen and cycle number in both datasets. Regimens were established in hospital administrative data using novel methodology to translate clinical coding guidelines into clinically meaningful information. Third, we identified potential biases that may originate from incomplete capture of chemotherapy in each dataset by exploring the characteristics of patients, regimens, and number of cycles

according to which dataset information was obtained from. Lastly, we carried out a sensitivity analysis to evaluate to what extent the observed impact of ACT has on 3-year colon cancer-specific mortality dependent on the type of dataset that was used to capture ACT information.

2. METHODS

2.1 Chemotherapy Data Sources

2.1.1 Systemic Anti-Cancer Therapy (SACT) dataset

The SACT is a dedicated chemotherapy dataset that includes detailed drug-level information, including administration date, drug name, dose, and administration route.[9] SACT captures chemotherapy administered in any inpatient, daycase, outpatient or community setting, and in most hospitals the data is collected via electronic prescribing systems.[10] In SACT, the drug name is a mandatory data item which is mapped to a pre-defined list of regimens.

2.1.2 Hospital Episode Statistics

The Hospital Episode Statistics database (HES) is an administrative dataset of all admissions to English NHS hospitals.[17] Inpatient and daycase chemotherapy use is captured via clinical coding, primarily through dedicated Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures, 4th revision (OPCS-4) codes[18], with chemotherapy-related International Classification of Diseases, 10th revision (ICD-10) codes also available (Appendix 1).[19]

2.2 Study population

2.2.1 National Bowel Cancer Audit

The National Bowel Cancer Audit (NBOCA) is a prospective mandatory database for all newly diagnosed colorectal cancer patients in the English NHS. Patients aged 18 years and above with a primary diagnosis of colon cancer, according to ICD-10 code C18, undergoing major resection at an English NHS hospital between 1 June 2014 and 30 April 2017 with pathological stage III colon cancer were identified in the NBOCA database. Cancers of the appendix were excluded.

10,280 NBOCA records were identified. These records were linked at patient-level to HES records, and to SACT records containing a colorectal cancer ICD-10 diagnosis code (C18-C20). Only SACT records between 30 June 2014 and 30 April 2018 were used because not all English NHS chemotherapy providers were submitting SACT data before 1 July 2014.[10] This ensured that all patients were followed up for at least 12 months from the date of surgery, allowing sufficient time for ACT completion. Linkage to SACT included all chemotherapy for each patient regardless of treatment intent (e.g. curative or palliative).

2.3 Measuring adjuvant chemotherapy

According to clinical guidelines for stage III colon cancer, standard ACT is considered to be fluoropyrimidine monotherapy (5-fluorouracil (5-FU) or capecitabine), or combination therapy as either 5-FU with oxaliplatin (FOLFOX), or capecitabine with oxaliplatin (CAPOX).[16]

For both datasets, the same rules were applied to determine which chemotherapy had been given in the adjuvant setting. First, restriction to the four standard regimens above was applied. Second, chemotherapy needed to have been started within 4 months of the NBOCA date of surgery and completed within 9 months of the first chemotherapy cycle, with gaps no larger than 3 months between consecutive cycles. Third, any patients who switched regimens partway through treatment to a non-standard regimen were assumed to have switched to palliative chemotherapy (Figure 1).

2.4 Establishing chemotherapy regimens within HES

HES does not provide regimen names. However, the National Tariff Chemotherapy Regimens List provides guidance on which OPCS-4 procurement and delivery codes should be used in HES, according to whether the chemotherapy is recorded as an inpatient or daycase administration, and which regimen is administered (Appendix 2).[18] Chemotherapy regimens were therefore indirectly captured in HES using novel methodology involving these codes.

2.5 Clinical characteristics

Data regarding sex, age, performance status, pathological T- and N-staging, surgical urgency and surgical access were obtained from NBOCA records. The Royal College of Surgeons' (RCS) Charlson comorbidity score was derived from ICD-10 codes recorded in the HES dataset in the year preceding colon cancer diagnosis.[20]

The hospital where the surgery was performed was identified according to NBOCA data. University teaching hospital status was determined according to the hospitals' membership of the University Hospital Association of United Kingdom University Hospitals.[21] Information about on-site chemotherapy facilities were collected in a national NBOCA survey of colorectal cancer services.[22]

Dates and causes of death were obtained from linked Office for National Statistics (ONS) mortality data.[23]

2.6 Stepwise validation framework

2.6.1 Patient-level agreement of receipt, regimen and number of cycles of ACT

First, patient-level agreement between SACT and HES with respect to ACT receipt and regimen was explored using a contingency table. Second, in patients who had linked SACT and HES records, agreement between the number of chemotherapy cycles recorded in each dataset was evaluated using Bland-Altman analysis with a line of best fit.[24] The distribution of recorded number of cycles according to each dataset was compared using bar charts.

2.6.2 Evaluating potential biases from incomplete capture within each dataset

Third, clinical characteristics, regimens and numbers of cycles were compared between patients with ACT captured in SACT alone, HES alone, or both datasets, using chi-squared tests to calculate *p*-values, using 0.05 as the statistical significance level.

2.6.3 Sensitivity of findings to the data source

Fourth, the 3 -year colon cancer-specific mortality from the NBOCA date of surgery was estimated separately for patients according to receipt of ACT. This was estimated using a competing risks method in which death

from other causes was the competing event.[25] Survival times were censored at 3 years after surgery or, if earlier, on the date of the last available death record, which was 10th February 2020. These mortality estimates were compared between analyses in which ACT receipt was identified in SACT alone, HES alone, or in either SACT or HES.

3. RESULTS

3.1 Patient-level agreement of receipt of adjuvant chemotherapy

10,280 patients were identified who had undergone surgical resection with pathological stage III colon cancer. 6,012 (58%) were identified as having received ACT according to either SACT or HES (Table 1). Of these 6,012 patients, 3,460 patients (58%) had ACT according to both datasets, 1,649 patients (27%) had ACT according to SACT alone, and 903 patients (15%) according to HES alone. Overall, there was 75% agreement between the two datasets (concordant cells / total number of patients). 68% of patients with ACT identified in SACT had ACT according to HES, and 79% of patients with ACT identified in HES had ACT according to SACT.

3.2 Patient-level agreement of recorded regimen and cycle number

Of the 3,460 patients with ACT recorded in both datasets, 3,320 (96%) had matching regimens in HES and SACT (Table 2). The Bland-Altman plot demonstrated reasonable agreement between the numbers of cycles recorded in each dataset for patients with ACT records in both (Figure 2). The 95% limits of agreement were -6.84 to 6.52. The line of best fit was very close to a zero mean difference in cycles which demonstrated good overall agreement across the range of mean number of cycles.

For the 3,460 patients with ACT recorded in both datasets, the overall distribution of the number of recorded cycles was similar regardless of data source, including the proportion of patients with a single cycle of chemotherapy recorded (6% in SACT vs. 7% in HES) (Figure 3a). HES captured slightly more patients having more than 8 cycles of chemotherapy (32% in SACT vs. 35% in HES).

3.3 Evaluating potential biases from incomplete capture within each dataset

Patients identified as having ACT in only one dataset were significantly more likely to be older, more comorbid, and less fit, compared to patients captured in both datasets (Table 3).

Patients identified as having ACT in SACT alone also tended to have less advanced nodal disease, and were more likely to have undergone major resection in a hospital without chemotherapy facilities on-site. Patients identified as having ACT in HES alone were more likely to have undergone major resection in a hospital that was not a university teaching hospital, and more likely to have had their major resection in a hospital with on-site chemotherapy facilities (Table 3).

There were statistically significant differences in capture of regimen according to dataset (P<0.001) (Table 4). Patients with ACT recorded in SACT alone were more likely to have received capecitabine, compared to patients in HES alone who were more likely to have received CAPOX.

Patients with ACT identified within a single dataset had a higher proportion of patients with a single cycle of chemotherapy recorded (11% for SACT alone and 12% for HES alone compared to 6% and 7% for patients with ACT identified in both datasets) (Figure 3a and 3b).

3.4 Sensitivity of findings to the data source

For patients classified as receiving ACT in either dataset (n=6,012), the 3-year colon cancer-specific mortality was 17.1% (95% CI: 16.1% to 18.0%). This was 16.8% (95% CI: 15.7% to 17.9%) in patients classified as receiving ACT according to HES alone (n=4,363), and 16.6% (95% CI: 15.6% to 17.7%) in patients classified as receiving ACT according to SACT alone (n=5,109).

For patients classified as not receiving ACT in either dataset (n=4,268), the 3-year colon cancer-specific mortality was 34.8% (95% CI: 33.4% to 36.2%). This was only slightly higher than the 30.1% (95% CI: 28.9% to 31.2%) observed in patients classified as not receiving ACT according to HES alone (n=5,917), and the 32.1% (95% CI; 30.9% to 33.4%) observed in patients classified according to SACT alone (n=5,171).

4. DISCUSSION

This study used a structured validation framework to examine the capture of ACT receipt and the accuracy of recording of regimen and cycle number in routinely collected national chemotherapy data (SACT) and administrative hospital data within the English NHS.[9] These datasets can be used in isolation or linked together, as well as linked at patient-level to other data sources, in order to inform improvements in service provision, clinical practice, and patient outcomes.[10]

Both datasets were found to be accurate in recording regimen type and cycle number. To our knowledge, the use of detailed coding within HES to assign chemotherapy regimens has not previously been explored. National guidelines exist which explicitly instruct on the recording of chemotherapy codes for financial reimbursement within hospital administrative data, meaning that their use should be standardised and this novel methodology transferable to other cancer types.[18]

This study did, however, highlight issues of incomplete capture of ACT in both datasets, particularly within hospital administrative data. Differences were demonstrated in clinical characteristics and regimens captured, although mortality rates remained comparable regardless of data source. Ideally, both sources of chemotherapy data should be used together to maximise capture of ACT and, in that way, reduce the potential for bias.

The main limitation of this study was that we did not consider HES outpatient data which might have captured more patients, in particular those receiving oral drugs, although capture of diagnosis and procedure coding within HES outpatient data is known to be very incomplete.[26]

Another limitation was that cycles were not matched between the two datasets according to dates. However, the same algorithm for determining ACT was applied to each dataset, and 90% of patients with ACT records in both datasets had a first chemotherapy date that matched within one week, in line with a previous study.[15]

The proportion of patients identified as receiving ACT according to either dataset is similar to other population-based studies.[27] In addition, the proportion of all patients recorded as receiving ACT according to hospital administrative data alone was 15%, comparable to previous work showing 12.5%.[15]

Reasons for patients being captured in only one dataset are likely multifactorial. However, the most important reason appears to be differences in the capture of chemotherapy delivered in outpatient and community settings, as demonstrated by 60% of all capecitabine being captured in the dedicated chemotherapy dataset alone.

Poor submission of SACT data may explain why some ACT is captured in HES alone. SACT case ascertainment has been linked to the availability of electronic prescribing (e-prescribing).[10] This is supported by our results showing that fewer patients in HES alone were managed in a university teaching hospital. It might be expected that the uptake of e-prescribing is higher in larger, academic oncology units.

The novel methodology used to assign chemotherapy regimens within hospital administrative data is important. First, it reduces overestimation of ACT cycles by limiting the inclusion of chemotherapy given outside the adjuvant setting. Second, it facilitates more clinically meaningful interpretation of hospital administrative data and can be adapted for other cancer types.

This study showed that for patients with records in both datasets, the proportion recorded as having only one cycle was similar in each dataset. In addition, the higher proportion of single cycles recorded when ACT was captured in a single dataset were consistent regardless of data source. This is therefore more likely to reflect clinical characteristics rather than a data quality issue.[10] Older, less fit patients are more likely to discontinue chemotherapy early due to toxicity, and these are the patients more likely to be captured in only one dataset.

Concerns about the capture of oral chemotherapy within SACT have been raised.[28] However, our results showed that reporting of capecitabine was considerably more complete in SACT, and the higher proportion of

patients with just one cycle of capecitabine recorded in one dataset alone could again be explained by the older, less fit population.

Patients captured in SACT alone tended to be older, less fit, have less advanced disease, and were more likely to receive fewer than 8 cycles. Capecitabine therapy consisted of 8 cycles as standard practice during the timeframe of this study.[29 30] Capecitabine monotherapy is also often favoured in the elderly, as well as those with low-risk cancer (T1-T3 and N1 disease), due to uncertain survival advantages and neurotoxicity associated with combination therapy.[31 32]

Hospital administrative data was more likely to have ACT missing if surgery was performed at a hospital which did not have on-site chemotherapy. This is likely explained by several tertiary oncology centres notably not recording chemotherapy within this dataset. Patients receiving ACT in tertiary centres are usually referred from hospitals which are different to those in which they underwent surgery.

Our findings are in line with those previously reported in lung cancer patients which suggested older, more comorbid patients might be under-represented in SACT, and raised concerns that mortality in those receiving chemotherapy may then be underestimated.[15] However, our study showed similar survival outcomes for those receiving chemotherapy regardless of which data source was used to classify ACT receipt. A higher mortality and larger absolute difference between those receiving and those not receiving ACT according to both datasets, compared to those classified by one dataset, supports the interpretation that the classification of ACT receipt is more accurate when both datasets are in agreement.

This study highlights the importance of validating routinely collected data, either national chemotherapy or administrative hospital data, on real world chemotherapy practice within specific cancer types. For example, some of the biases demonstrated were due to differential capture of oral chemotherapy which may not be applicable to all cancers. However, the transparent structured validation can be applied to other cancer types as well as different lines and types of chemotherapy, such as hormonal and biological agents.

The dedicated chemotherapy dataset is the first national dataset of its kind, relying largely on capture of data from e-prescribing systems. This data can be linked at patient-level to other national datasets, providing invaluable opportunities for research.[12] Many European countries, the United States, Australia and New Zealand, have been expanding e-prescribing within primary care[33], and this study adds further rationale for implementing e-prescribing in secondary care to reduce data collection burden if not already available.

5. CONCLUSION

This study has demonstrated the accuracy of data from a national chemotherapy dataset (SACT) and administrative hospital dataset (HES) for patients with stage III colon cancer receiving chemotherapy in the English NHS when records are present from both sources. However, chemotherapy information should ideally be captured from both datasets to avoid under-capture, particularly of oral chemotherapy in administrative hospital data, and to minimise bias.

This methodology should facilitate more accurate and robust national reporting of chemotherapy use and outcomes, with applicability across different cancer types. Other countries should consider the feasibility of e-prescribing for the routine collection of national dedicated chemotherapy data which can be linked to other data sources in order to inform healthcare quality improvement.

Acknowledgements*:* This study was based on data collected by the National Bowel Cancer Audit linked to the Systemic Anti-cancer Therapy database (https://www.chemodataset.nhs.uk/home) made available by the National Cancer Analysis and Registration Service and Hospital Episode Statistics Admitted Patient Care made available by NHS Digital (https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics).

Table 1 - Numbers of patients identified as commencing ACT within 4 months of surgical resection with pathological stage III colon cancer, according to either SACT or HES datasets

	ACT according to SACT		
ACT according to HES	**Yes**	**No**	**Total**
Yes	3,460*	903*	4,363
No	1,649*	4,268	5,917
Total	5,109	5,171	10,280

*6,012 patients (58%) were identified as receiving ACT according to SACT and/or HES

Table 2 – Numbers of patients receiving each standard ACT regimen according to SACT and HES, for patients with ACT in both datasets (n=3,460)

	First adjuvant HES regimen			
First adjuvant SACT regimen	5-FU	FOLFOX	Capecitabine	CAPOX
5-FU	**252**	31	1	1
FOLFOX	7	**1,097**	0	18
Capecitabine	6	2	**391**	63
CAPOX	0	8	3	**1,580**

Table 3 – Patient, tumour and hospital-level characteristics according to the dataset capturing ACT use

	SACT and HES (n=3,460)		HES alone (n=903)		SACT alone (n=1,649)		χ^2 *P* values
	No.	%	No.	%	No.	%	
Sex							0.827
Male	1,832	52.9	483	53.5	862	52.3	
Female	1,628	47.1	420	46.5	787	47.7	
Age							**<0.001**
<60	992	28.7	250	27.7	357	21.6	
60-69	1,250	36.1	296	32.8	463	28.1	
70-79	1,024	29.6	287	31.8	703	42.6	
=>80	194	5.6	70	7.8	126	7.6	
RCS Charlson Score							**0.001**
0	2,242	64.8	549	60.8	996	60.4	
1	944	27.3	276	30.6	475	28.8	
≥2	274	7.9	78	8.6	178	10.8	
Performance Status							**<0.001**
0	1,854	62.8	522	64.3	765	52.5	
1	871	29.5	215	26.5	534	36.7	
≥2	228	7.7	75	9.2	157	10.8	
Missing	507	14.7	91	10.1	193	11.7	
Pathological T-stage							0.072
T1/T2	294	8.5	75	8.3	158	9.6	
T3	1,756	50.8	445	49.3	877	53.2	
T4	1,409	40.7	382	42.4	614	37.2	
Missing	1	0	1	0.1	0	0	
Pathological N-stage							**0.005**
N1	2,158	62.4	534	59.1	1,080	65.5	
N2	1,302	37.6	369	40.9	569	34.5	
Surgical Urgency							**0.001**
Elective/Scheduled	2,802	81.1	680	75.4	1,315	79.9	
Emergency/Urgent	653	18.9	222	24.6	330	20.1	
Missing	5	0.1	1	0.1	4	0.2	
Surgical Access							**0.004**
Open operation	1,179	34.2	361	40.0	582	35.4	
Laparoscopic converted	258	7.5	71	7.9	148	9.0	
Laparoscopic	2,008	58.3	470	52.1	916	55.7	
Missing	15	0.4	1	0.1	3	0.2	
Chemotherapy on-site							**<0.001**
No	224	6.5	46	5.1	434	26.3	
Yes	3,236	93.5	857	94.9	1,215	73.7	

University Teaching Hospital							**<0.001**
	No	2,547	73.6	744	82.4	1,234	74.8
	Yes	913	26.4	159	17.6	415	25.2

Table 4 – Distribution of ACT regimen according to whether the patient has ACT in both HES and SACT, SACT only, or HES only. χ^2 test for association: *P* value=<0.001

Regimen	SACT and HES (n=3,640)	SACT only (n=1,649)	HES only (n=903)	Overall (n=6,012)
5-FU	285 (63.1)	76 (16.8)	91 (20.1)	452
(Column %)	*8.2*	*4.6*	*10.1*	*7.5*
FOLFOX	1,122 (63.2)	404 (22.7)	250 (14.1)	1,776
(Column %)	*32.4*	*24.5*	*27.7*	*29.5*
Capecitabine	462 (32.7)	854 (60.4)	97 (6.9)	1,413
(Column %)	*13.4*	*51.8*	*10.7*	*23.5*
CAPOX	1,591 (67.1)	315 (13.3)	465 (19.6)	2,371
(Column %)	*46.0*	*19.1*	*51.5*	*39.4*

Figure 1 – Algorithms applied to SACT and HES records to establish ACT and resulting final patient cohorts

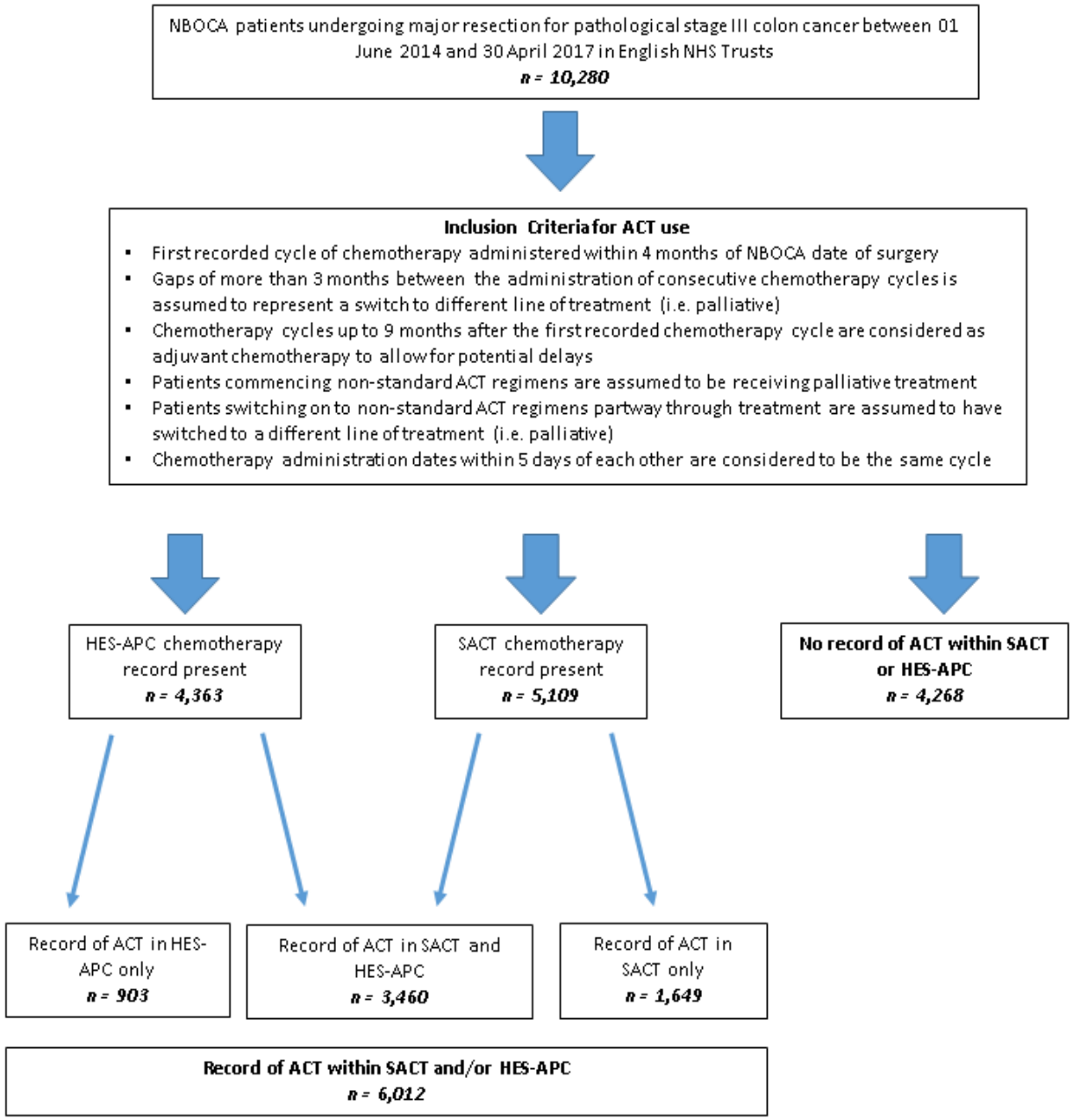


Figure 2 – Bland-Altman plot demonstrating agreement between the mean number of cycles of chemotherapy according to SACT and HES at patient-level, and the difference between the number of cycles recorded in HES and SACT at patient-level, for patients with ACT in both datasets (n=3,460)

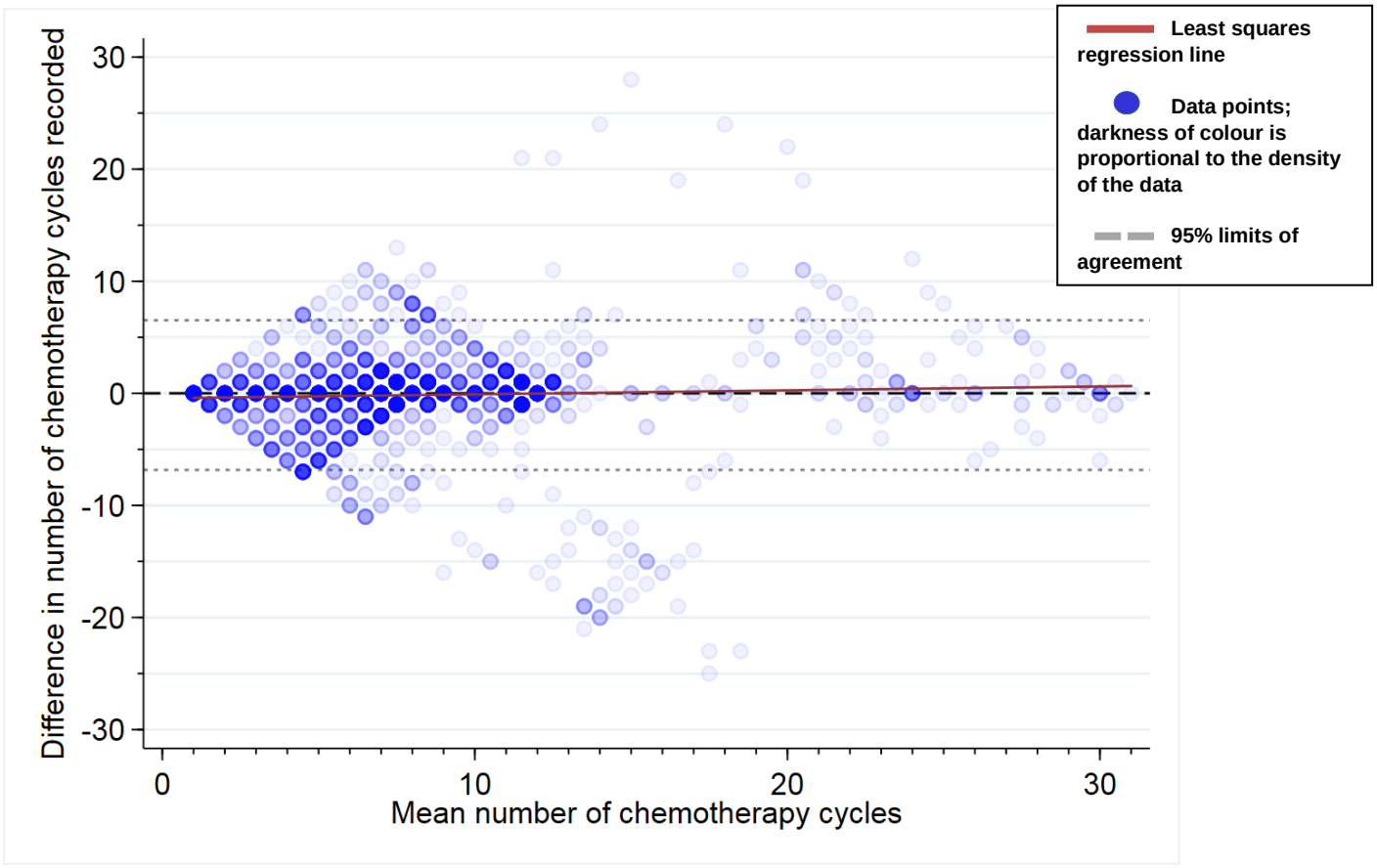
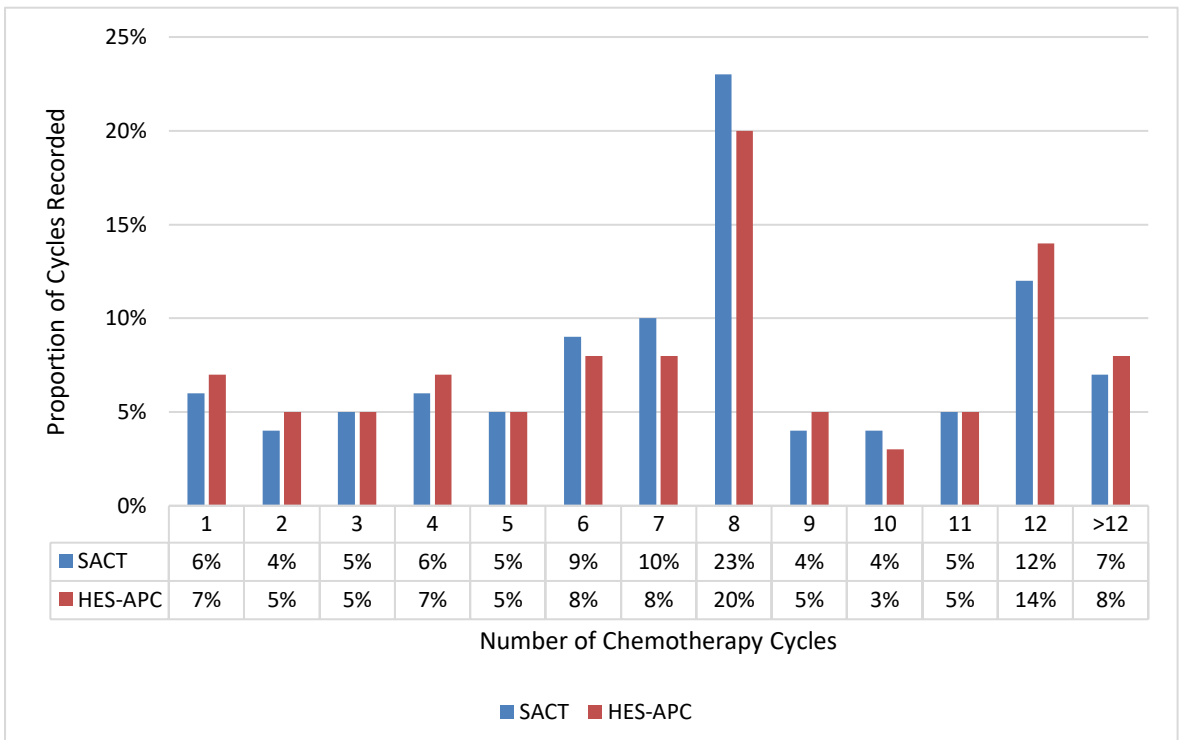
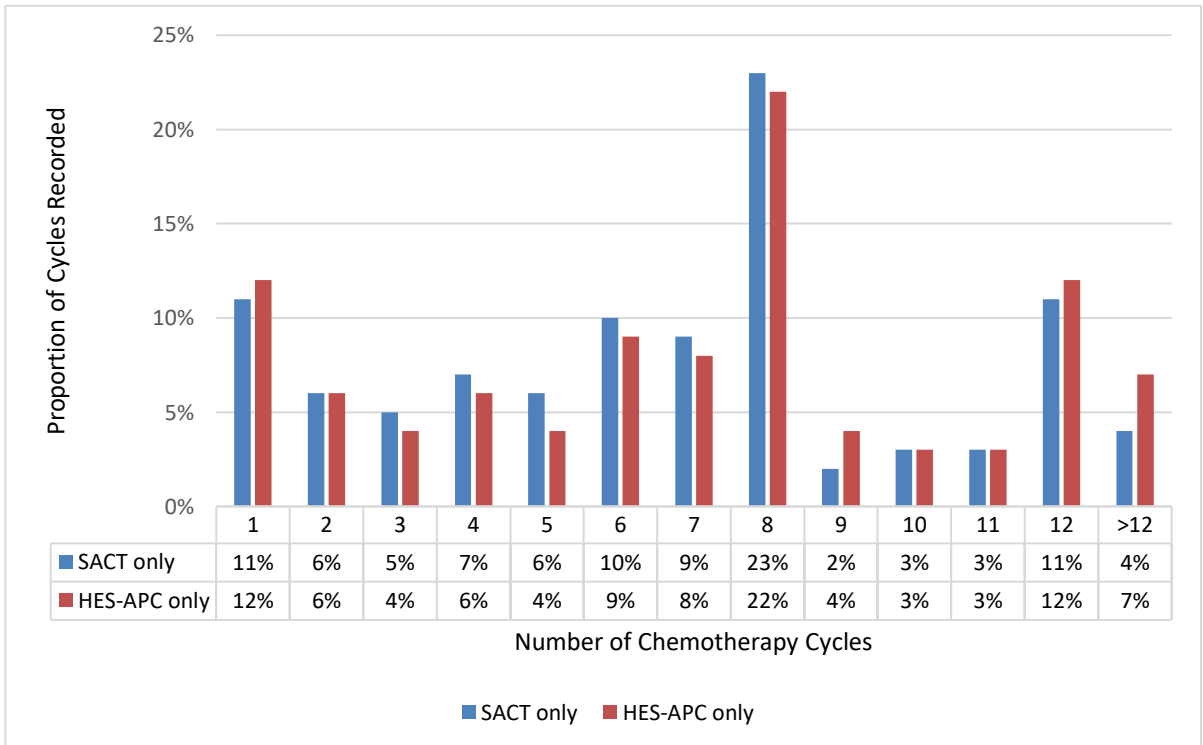


Figure 3a - Bar chart demonstrating the distribution of total ACT cycles recorded in SACT compared to HES, for patients with ACT in both datasets (n=3,460)



	1	2	3	4	5	6	7	8	9	10	11	12	>12
SACT	6%	4%	5%	6%	5%	9%	10%	23%	4%	4%	5%	12%	7%
HES-APC	7%	5%	5%	7%	5%	8%	8%	20%	5%	3%	5%	14%	8%

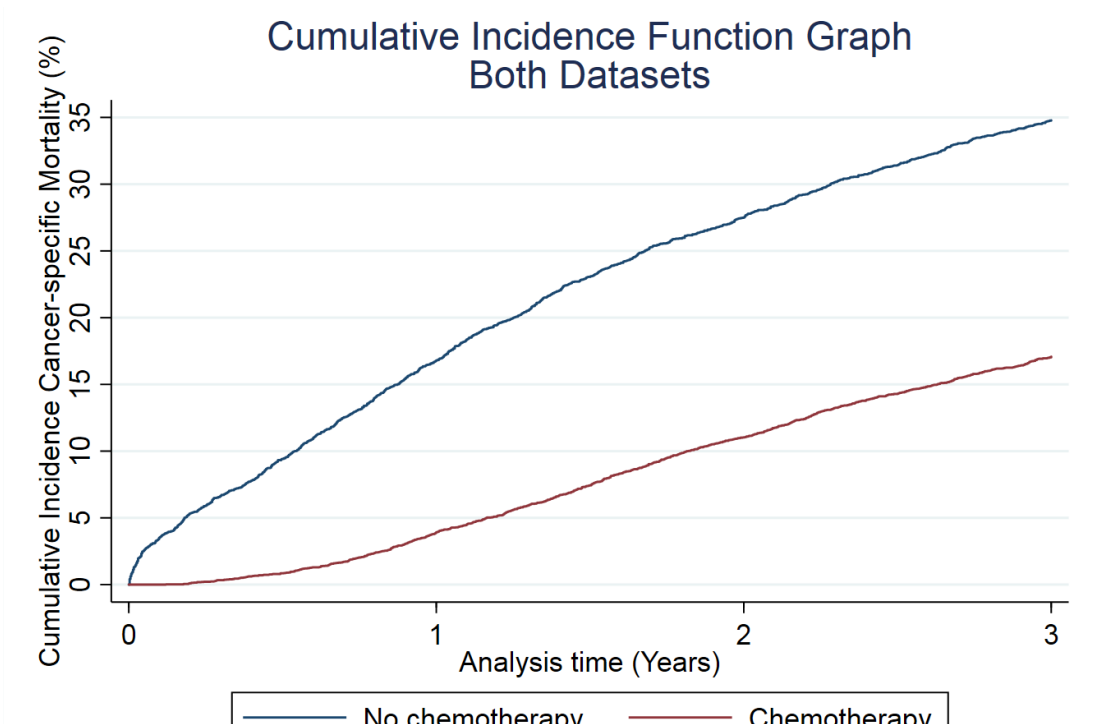
Figure 3b - Bar chart demonstrating the distribution of total ACT cycles recorded in SACT compared to HES for those patients with ACT in SACT only (1,649) and HES only (n=903)



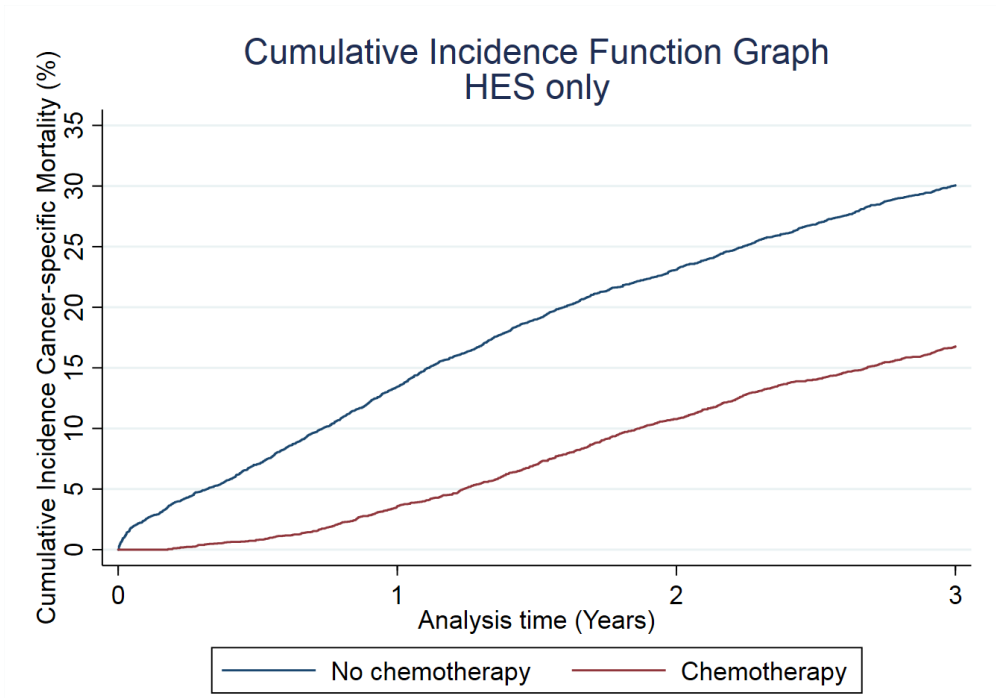
	1	2	3	4	5	6	7	8	9	10	11	12	>12
SACT only	11%	6%	5%	7%	6%	10%	9%	23%	2%	3%	3%	11%	4%
HES-APC only	12%	6%	4%	6%	4%	9%	8%	22%	4%	3%	3%	12%	7%

Figure 4 – Cumulative incidence function graphs for 3-year colon cancer-specific mortality with competing risk of other causes of death stratified by receipt of ACT, according to classification within a) HES only b) SACT only and c) both datasets

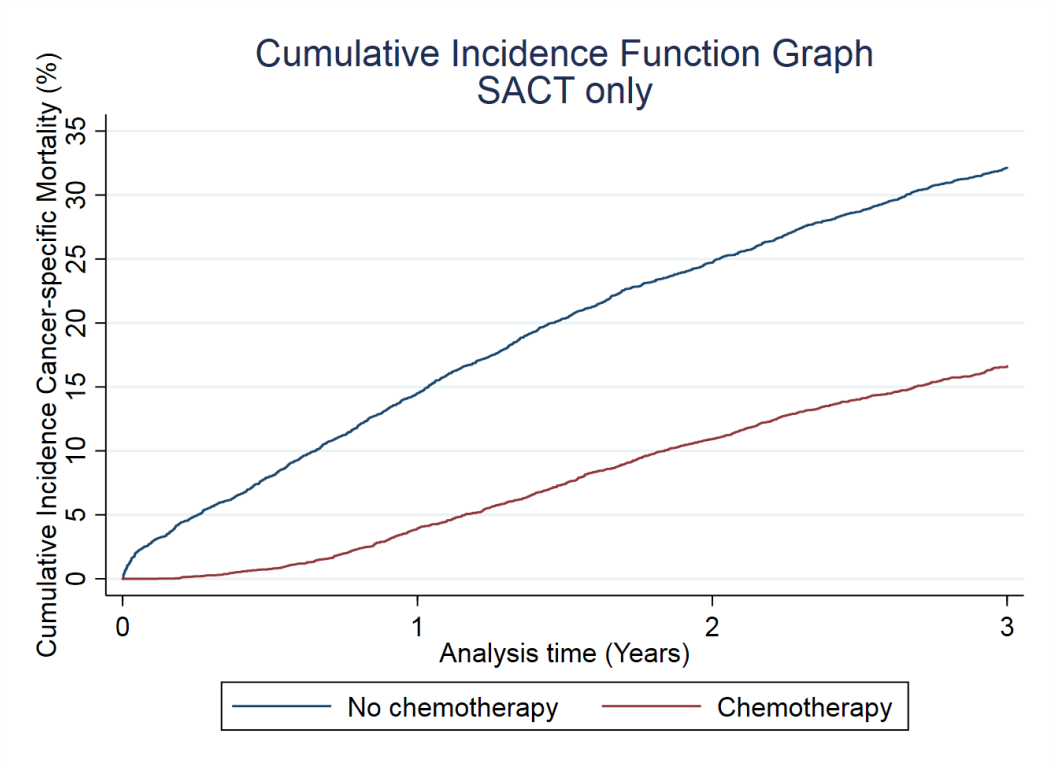
a)



b)



c)


Cumulative Incidence Function Graph
SACT only
Cumulative Incidence Cancer-specific Mortality (%)
0
5
10
15
20
25
30
35
0
1
2
3
Analysis time (Years)
No chemotherapy
Chemotherapy

References

1. Laurie JA, Moertel CG, Fleming TR, et al. Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. The North Central Cancer Treatment Group and the Mayo Clinic. J Clin Oncol 1989;**7**(10):1447-56.
2. Andre T, Boni C, Mounedji-Boudiaf L, et al. Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. N Engl J Med 2004;**350**(23):2343-51.
3. Twelves C, Wong A, Nowacki MP, et al. Capecitabine as adjuvant treatment for stage III colon cancer. N Engl J Med 2005;**352**(26):2696-704.
4. McKee M, Britton A, Black N, et al. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. Bmj 1999;**319**(7205):312-5.
5. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. Jama 2004;**291**(22):2720-6.
6. Sørensen HT, Lash TL, Rothman KJ. Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. Hepatology (Baltimore, Md.) 2006;**44**(5):1075-82.
7. Mokhles S, Takkenberg JJ, Treasure T. Evidence-Based and Personalized Medicine. It's [AND] not [OR]. The Annals of thoracic surgery 2017;**103**(1):351-60.
8. Treasure T, Takkenberg JJM. Randomized trials and big data analysis: we need the best of both worlds. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery 2018;**53**(5):910-14.
9. Systemic Anti-Cancer Therapy (SACT) Chemotherapy Dataset. National Cancer Registration and Analysis Service. Public Health England.
10. Bright CJ, Lawton S, Benson S, et al. Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) Dataset. International journal of epidemiology 2019.
11. Pathak R, Wallington M, Saunders C, et al. Rapid Analysis of Outcomes Using the Systemic Anti-Cancer Therapy (SACT) Dataset. Clin Oncol (R Coll Radiol) 2017;**29**(7):e134-e36.
12. Wallington M, Saxon EB, Bomb M, et al. 30-day mortality after systemic anticancer treatment for breast and lung cancer in England: a population-based, observational study. The Lancet. Oncology 2016;**17**(9):1203-16.
13. Jones GS, McKeever TM, Hubbard RB, Khakwani A, Baldwin DR. Factors influencing treatment selection and 30-day mortality after chemotherapy for people with small-cell lung cancer: An analysis of national audit data. European journal of cancer (Oxford, England : 1990) 2018;**103**:176-83.
14. Henson KE, Fry A, Lyratzopoulos G, et al. 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. British journal of cancer 2018;**118**(10):1382-90.
15. McDonald L, Sammon C, Carroll R, 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 oncology (London, England) 2020;**16**(3):4455-60.
16. National Institute for Health and Care Excellence. Colorectal cancer. NICE guideline [NG151]. Available: https://www.nice.org.uk/guidance/ng151
[Accessed: 31st March 2021]
17. Hospital Episode Statistics. NHS Digital. Available: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
[Accessed: 15/08/18]
18. The Health and Social Care Information Centre. Chemotherapy regimens clinical coding standards and guidance OPCS-4 April 2017. (2017). Available: [Accessed: 10th February 2020]
19. NHS Digital TRUD. NHS Classifications ICD-10.

20. 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.
21. Affiliate Groups of The Association of UK University Hospitals. Available: www.universityhospitals.org.uk
[Accessed: 31st March 2021]
22. Organisational Survey. Available: https://www.nboca.org.uk/reports/organisational-survey-results-2018/
[Accessed: 23rd December 2020]
23. Office for National Statistics. Deaths. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths
[Accessed: 31st March 2021]
24. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986;**327**(8476):307-10.
25. Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. Stata Journal 2004;**4**(2):103-12.
26. Marques E, Noble S, Blom AW, Hollingworth W. Disclosing total waiting times for joing replacement: evidence from the English NHS using linked HES data. Health Economics 2014;**23**(7):806-20.
27. Babaei M, Balavarca Y, Jansen L, et al. Administration of adjuvant chemotherapy for stage II-III colon cancer patients: An European population-based study. International journal of cancer 2018;**142**(7):1480-89.
28. Public Health England. Calculating Treatment Duration for Oral Drugs. Cancer Drugs Methodology Document., 2019.
29. Papamichael D, Audisio RA, Glimelius B, et al. Treatment of colorectal cancer in older patients: International Society of Geriatric Oncology (SIOG) consensus recommendations 2013. Annals of oncology : official journal of the European Society for Medical Oncology 2015;**26**(3):463-76.
30. Grothey A, Sobrero AF, Shields AF, et al. Duration of Adjuvant Chemotherapy for Stage III Colon Cancer. New England Journal of Medicine 2018;**378**(13):1177-88.
31. Tournigand C, Andre T, Bonnetain F, et al. Adjuvant therapy with fluorouracil and oxaliplatin in stage II and elderly patients (between ages 70 and 75 years) with colon cancer: subgroup analyses of the Multicenter International Study of Oxaliplatin, Fluorouracil, and Leucovorin in the Adjuvant Treatment of Colon Cancer trial. J Clin Oncol 2012;**30**(27):3353-60.
32. McCleary NJ, Meyerhardt JA, Green E, et al. Impact of age on the efficacy of newer adjuvant therapies in patients with stage II/III colon cancer: findings from the ACCENT database. J Clin Oncol 2013;**31**(20):2600-6.
33. Health Information and Standards. ePrescribing: An International Review. May 2018. Available: https://www.hiqa.ie/sites/default/files/2018-05/ePrescribing-An-Intl-Review.pdf
[Accessed: 31st March 2021]

Appendix 1 - OPCS-4 and ICD-10 codes for chemotherapy use in HES

OPCS-4 code	**Classification**
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
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
X352	Intravenous chemotherapy
X373	Intramuscular chemotherapy
X384	Subcutaneous chemotherapy
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

Appendix 2 – OPCS-4 delivery and procurement codes used to determine ACT regimen within HES according to the National Tariff Chemotherapy Regimens List

	Inpatient chemotherapy	Daycase chemotherapy		
ACT Regimen	**Procurement Code**	**Procurement Code**	**Delivery Code**	**Overall Code**
5-FU	X701	X701	X721 or X723	X701 and X721/X723
FOLFOX	X704	X704	X721	X704 and X721
Capecitabine	X702	X702	X731	X702 and X731
CAPOX	X711	X711	X722	X711 and X722

The code combinations for other potential colorectal chemotherapy regimens were checked to ensure that the same codes were not being used for other regimens. Inpatient chemotherapy is coded with a procurement code only, in comparison to daycase chemotherapy which has both procurement and delivery codes.

For each of CAPOX and 5-FU, two procurement codes were available which could potentially impact on the recording of their inpatient delivery. For 5-FU, X702 or X701 could be coded, and for CAPOX this could be X704 or X711. However, less than 1% of HES records had chemotherapy recorded as an inpatient and, when comparing linked SACT-HES data, >80% of recorded codes were X701 and X711 respectively. These factors meant that the chances of 5-FU and capecitabine, or CAPOX and FOLFOX, being misclassified when recorded as an inpatient were very low.

Similarly, 5-FU had two delivery codes available which were X721 and X723. Any possible combinations of procurement and delivery codes for 5-FU remained unique, and therefore this was not an issue for the coding of daycase 5-FU.