Identification of disease-related events in patients with primary breast cancer using routinely available datasets in England for use in survival-related endpoints: A practical alternative to hospital-based follow-up for breast cancer clinical trials.

Supplementary material

Contents

Additional information on methods and procedure development

Data transfer between ICR-CTSU and NCRAS

A file containing explicit identifiers (project-specific pseudo-ID, NHS number, date of birth, sex) pertaining to the trial participant cohort was sent from ICR-CTSU to the NCRAS via secure file transfer (256 AES encrypted file). Once linked, pseudonymised datasets were returned to ICR-CTSU for analysis using the same file transfer method. Passwords for decryption in both cases were transferred by telephone on data receipt.

Additional information of use of cause of death

Cause of death described as the primary cause of death on the death certificate (fields 1a-c) was used to establish whether the death was due to breast cancer. If breast cancer was recorded as a secondary cause of death on the death certificate (field 2) this was not included as a breast cancer death in the analysis.

Alternative methods to identify disease recurrence

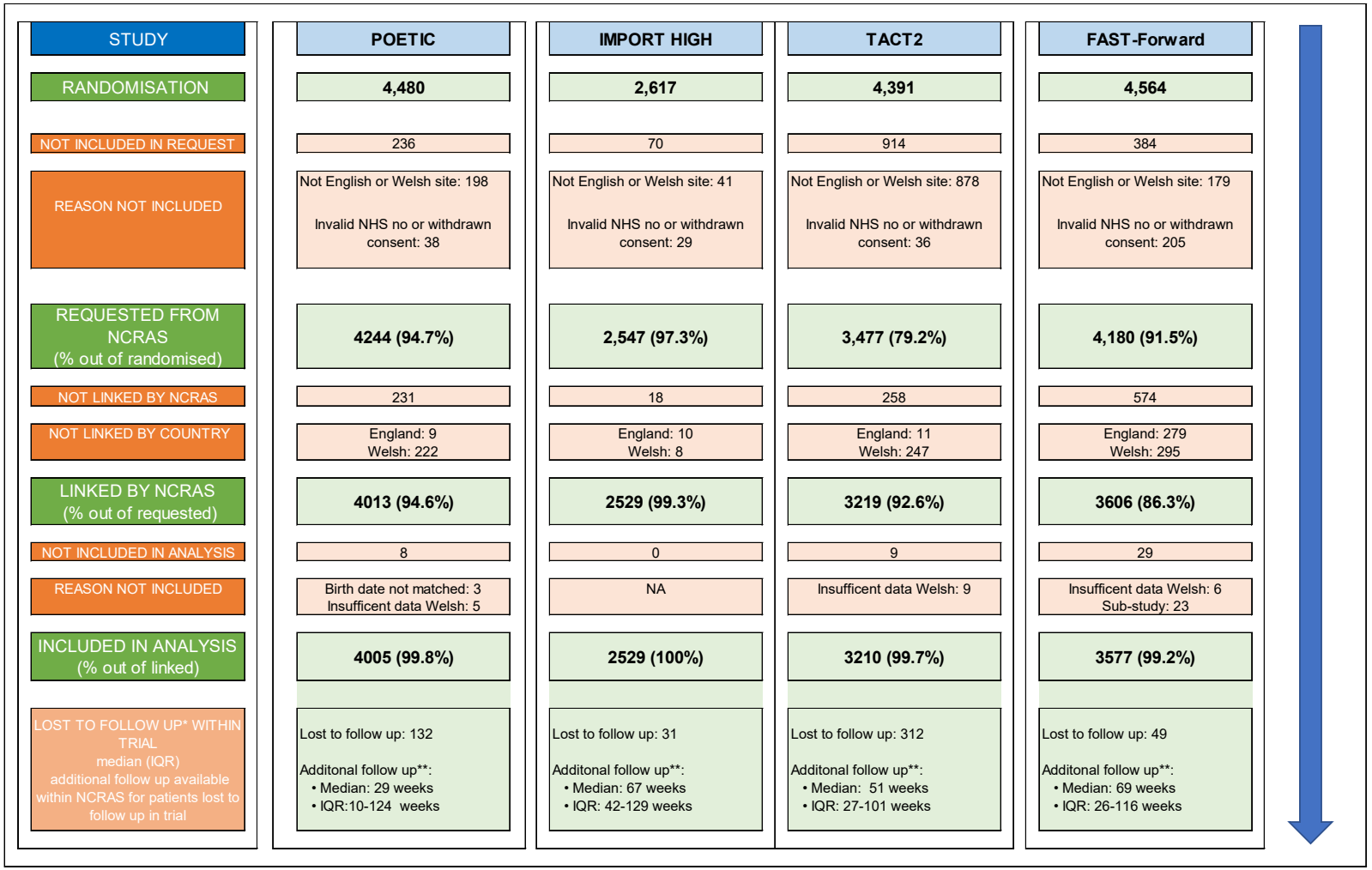
Recurrence identification was initially explored by looking at the events related to chemotherapy, radiotherapy and breast cancer surgery and assigning a weight to each activity and assessing whether treatment events were within close proximity to each other (e.g., within 5 weeks) and calculating a total score. Scores above a given value were assigned as being an event vs not being an event (but related to routine follow up). In addition, to avoid picking up the clinical management related to a patient's initial diagnosis, only activity after 64 weeks was considered (or after 90 weeks for HER2 positive patients to allow for the extended anti-HER2 treatment these patients usually receive). The first group of events was identified and compared to the trial data. This showed promise for picking up the first event in the test (POETIC) dataset. However, details relating to the reason for the hospital visit were inconsistently reported therefore an alternative route was considered. A second option considered activity in the HES OP and APC datasets predominantly by looking at frequency of medical oncology visits as this higher-level data appeared to be more complete. Again, this method showed promise however, neither method readily provided a mechanism to pick up the site of recurrence, i.e. local vs distant disease, and so, a third method was explored as described in the main body of the article.

Details on the programs developed for the final procedure

The procedure development was split into three stages. The first stage was where the main data processing and cleaning took place. This was done at the dataset level generally as opposed to the event level and included general formatting such as merging datasets, formatting dates, assigning text to OPCS4 codes and ICD-10 codes to help identify data of interest, assigning sites of disease and cause of death.

In addition, flags to identify key data such as invasive cancer diagnosis and metastatic cancer codes (C77X, C78X and C79X) were created and the data reduced as appropriate. The second stage required the formatted data to be used to identify the key events: initial diagnosis of breast cancer, deaths, second primary cancer diagnosis (breast and non-breast), distant and local recurrence; which was then combined to create the composite endpoints of interest. The final stage compared the NCRAS event data to the trial data. It was felt that this was the most efficient way to organise and analyse the data presented for this research project.

Figure A1: Data linkage patient flow for each of the four ICR-CTSU trials

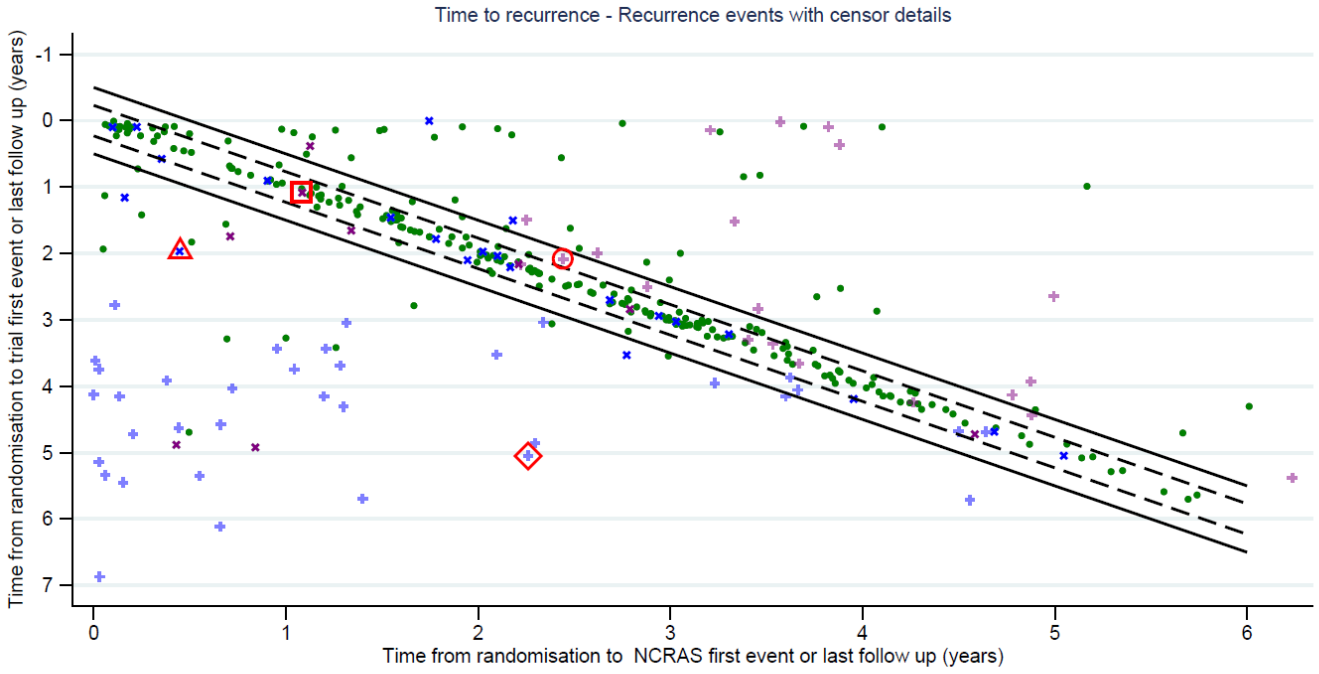


* Lost to follow up is based on patients were their follow up stopped before the end of the trial and the patient hadn't died.

** The additonal follow up is based on the patients that are classifed as lost to follow up and assumes the patients hosptial visits are still being picked up in NCRAS up to the last follow for the HES datasets (31/01/2016).

Figure A2: Scatter plots for comparison of time to recurrence event and other between NCRAS and trial data for each ICR-CTSU trial

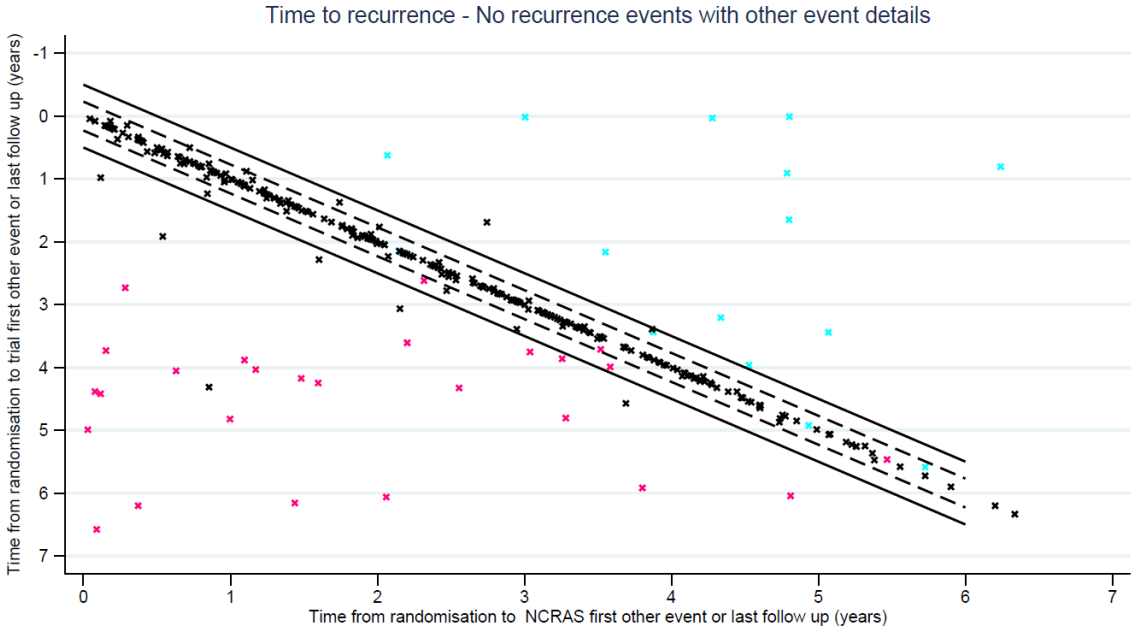
(A) POETIC TTR events



Other event = other second primary including breast, death not from breast cancer
NCRAS only event: last trial date based on last trial follow up or other event
Trial only event: last NCRAS date based on the date the registry date is complete to or other event

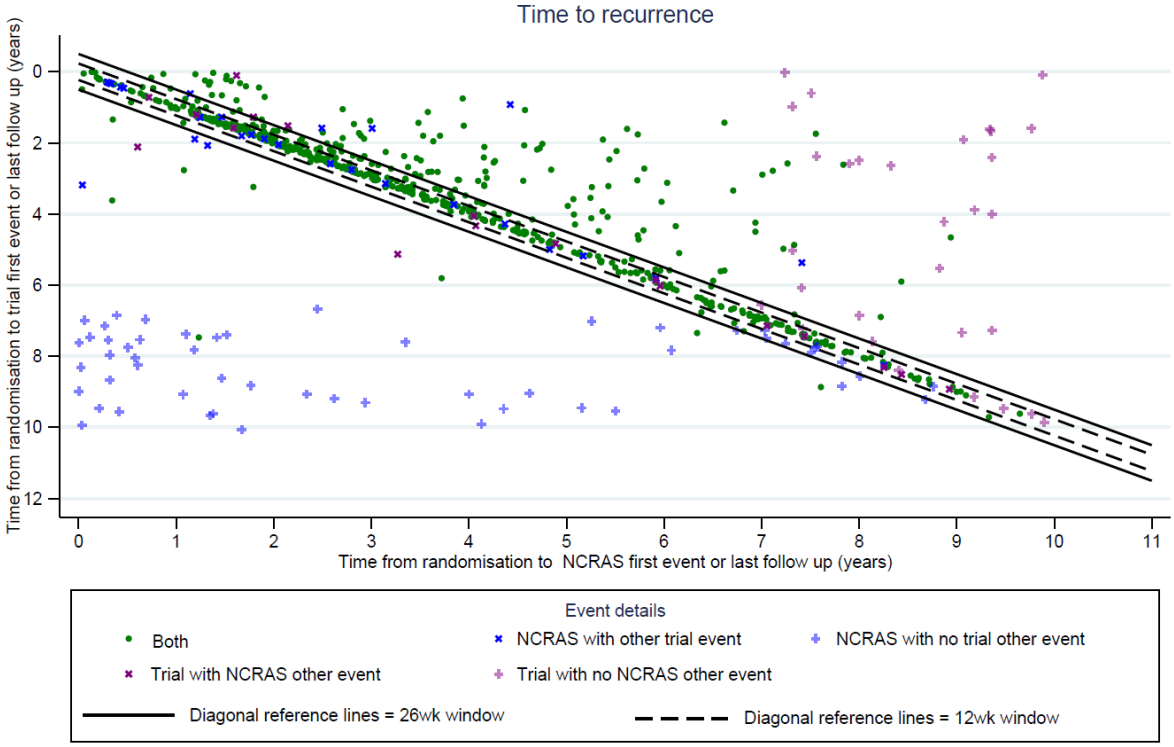
Examples of data highlighting possible misclassifications or not	
○	Trial event and no NCRAS event, trial event close to last follow up for NCRAS, possible available in later data
□	Trial event and other NCRAS event, the same date indicates misclassification of event
◇	NCRAS event and no trial event, last trial event >2 years after NCRAS event indicating that limited follow up is not causing there to be no trial event
△	NCRAS event and other trial event over 1.5 years after the NCRAS event indicating misclassification is not reason for different event types

(B) POETIC other events



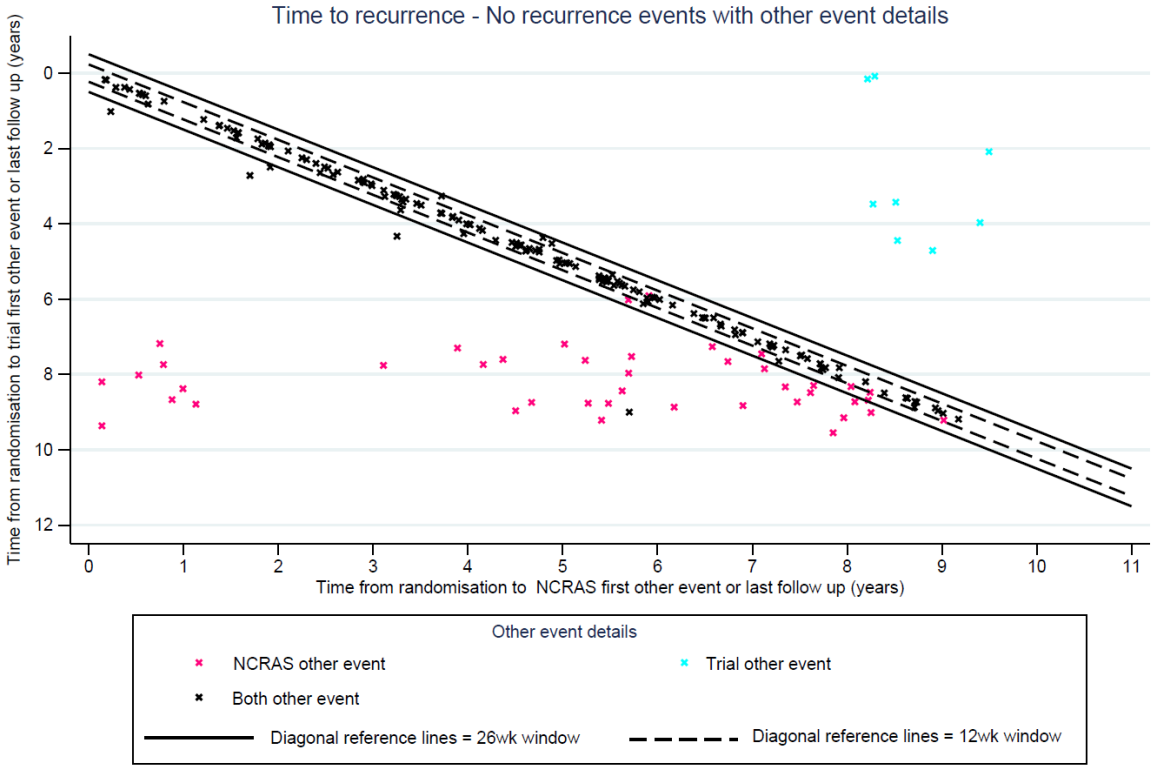
Other event = other second primary including breast, death not from breast cancer

(C) TACT2 TTR events



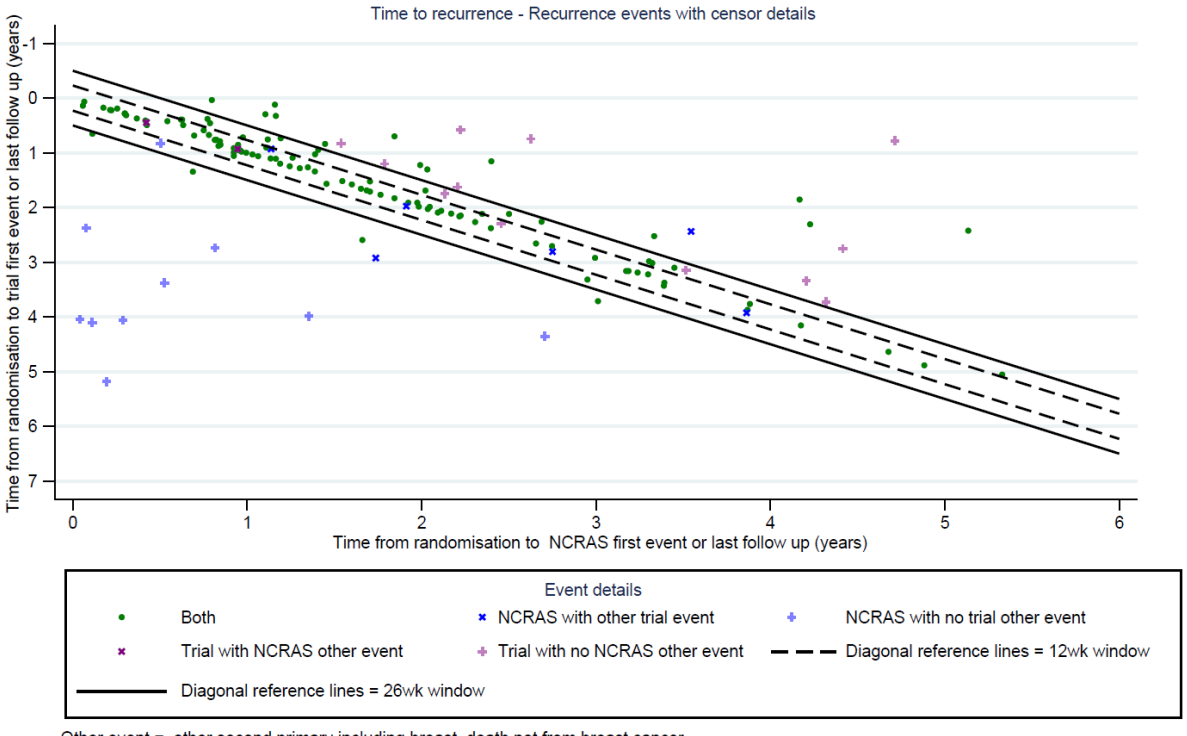
Other event = other second primary including breast, death not from breast cancer
NCRAS only event: last trial date based on last trial follow up or other event
Trial only event: last NCRAS date based on the date the registry date is complete to or other event

(D) TACT2 other events



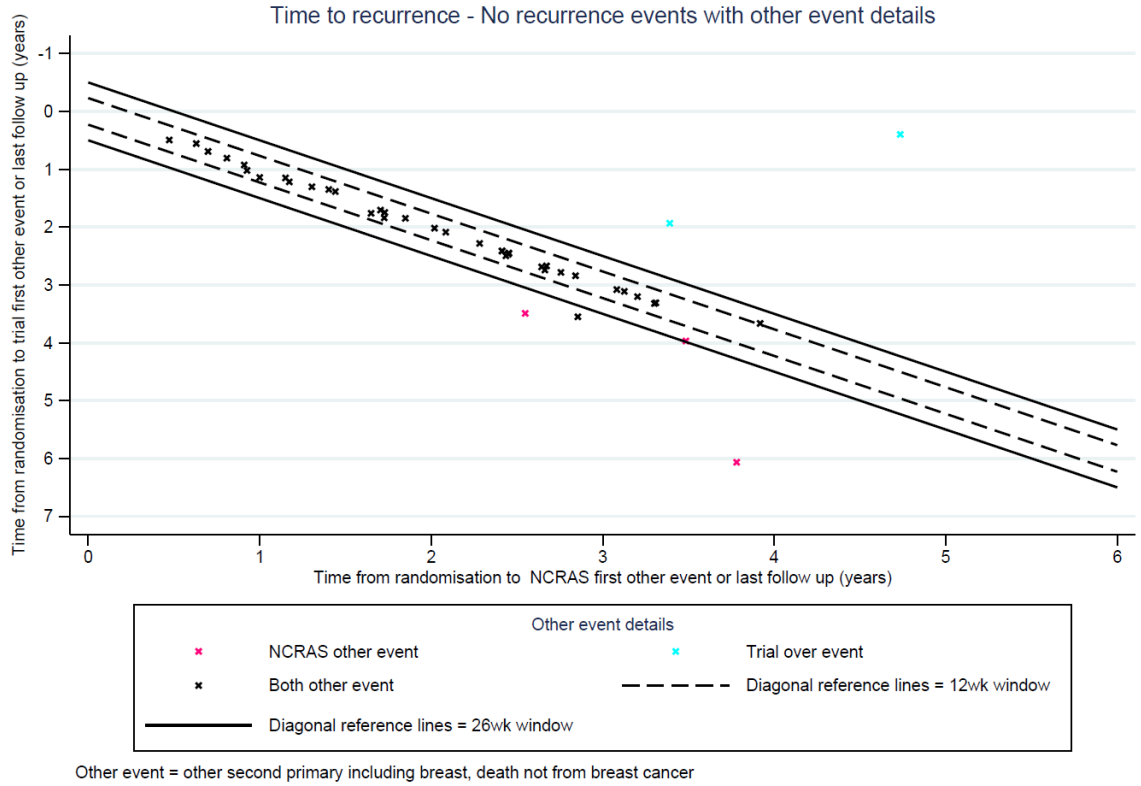
Other event = other second primary including breast, death not from breast cancer

(E) IMPORT HIGH TTR events



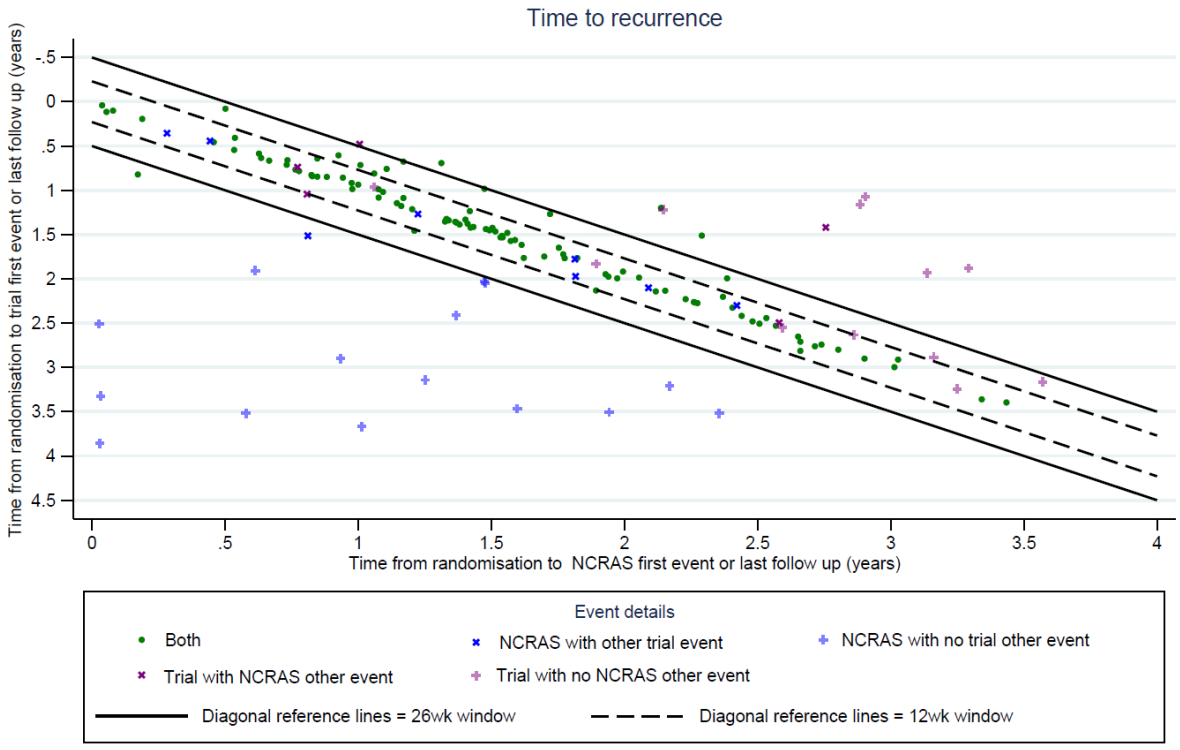
Other event = other second primary including breast, death not from breast cancer
NCRAS only event: last trial date based on last trial follow up or other event
Trial only event: last NCRAS date based on the date the registry date is complete to or other event

(F) IMPORT HIGH other events



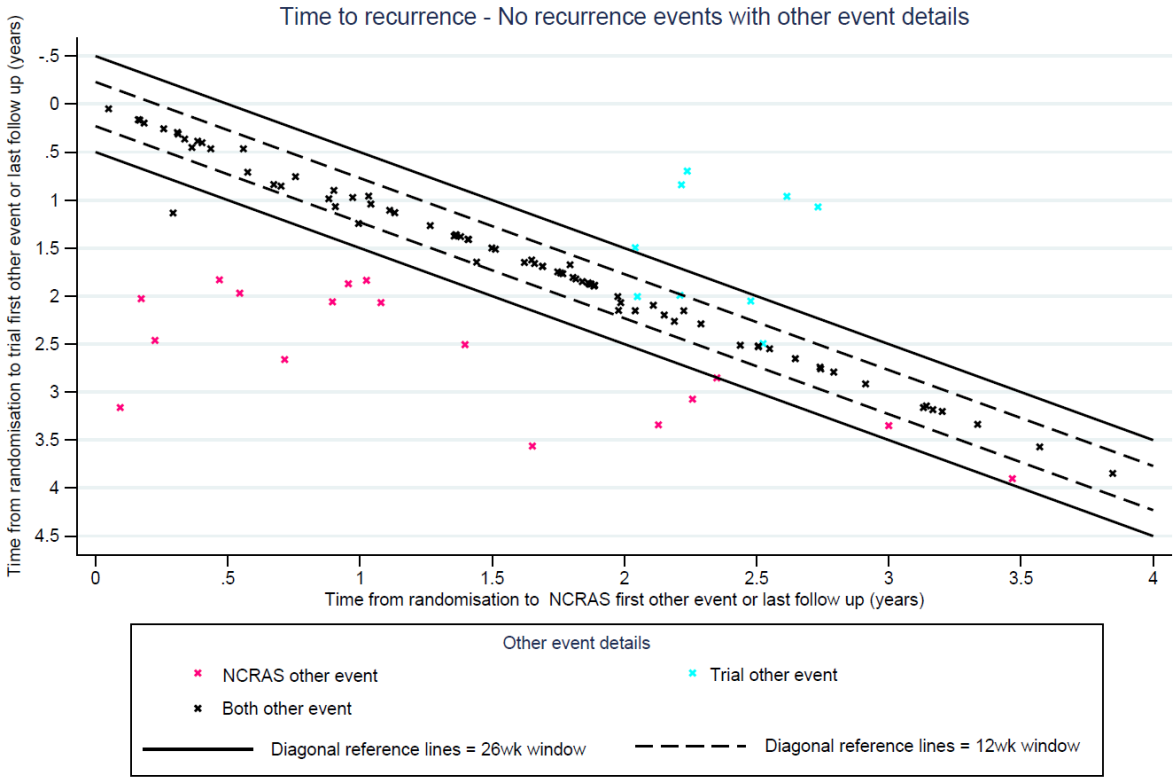
Other event = other second primary including breast, death not from breast cancer

(G)FAST Forward TTR events



Other event = other second primary including breast, death not from breast cancer
NCRAS only event: last trial date based on last trial follow up or other event
Trial only event: last NCRAS date based on the date the registry date is complete to or other event

(H) FAST Forward other events



Other event = other second primary including breast, death not from breast cancer

Table A1: 3 year survival estimates and hazard ratios for treatment effects for TTR, iDFS and OS for each trial by NCRAS and trial data.

Event	Measure	POETIC		IMPORT HIGH		TACT2		FAST-FORWARD	
		NCRAS	Trial	NCRAS	Trial	NCRAS	Trial	NCRAS	Trial
TTR									
estimate (95% CI)	Treatment 1	94.6 (93.7 to 95.4)	95.5 (94.6 to 96.2)	94.2 (91.9 to 95.8)	94.3 (92.1 to 96.0)	89.8 (88.2 to 91.2)	90.0 (88.4 to 91.4)	96.3 (95.0 to 97.3)	96.9 (95.6 to 97.8)
	Treatment 2			94.1 (91.8 to 95.7)	94.2 (92.0 to 95.9)	90.6 (89.1 to 92.0)	90.5 (89.0 to 91.9)	95.6 (94.0 to 96.8)	95.7 (94.0 to 96.9)
	Control 1	94.4 (93.0 to 95.5)	94.6 (93.3 to 95.7)	94.8 (92.6 to 96.3)	94.6 (92.5 to 96.1)	91.5 (90.0 to 92.7)	90.2 (88.6 to 91.6)	96.2 (94.8 to 97.3)	95.6 (93.9 to 96.8)
	Control 2					90.7 (89.1 to 92.0)	89.7 (88.1 to 91.1)		
HR (95%CI), p-value (log rank)	Treatment 1 vs control 1	1.08 (0.86-1.37) p= 0.51	1.10 (0.86-1.41) p=0.45	1.09 (0.71-1.69) p=0.69	0.96 (0.62-1.50) p=0.87	0.97 (0.83-1.13) p= 0.67	1.00 (0.85-1.17) p= 0.98	1.07 (0.69-1.65) p= 0.76	1.19 (0.76-1.85) p= 0.45
	Treatment 2 vs control 1			1.22 (0.80-1.87) p=0.36	1.13 (0.74-1.72) p=0.56			0.98 (0.64-1.50) p= 0.92	1.03 (0.67-1.58) p= 0.90
	Treatment 2 vs control 2					1.00 (0.86-1.17) p= 0.96	1.04 (0.88-1.21) p= 0.67		
iDFS									
estimate (95% CI)	Treatment 1	90.0 (88.8 to 91.1)	91.0 (89.9 to 92.1)	91.9 (89.2 to 93.9)	91.6 (88.9 to 93.7)	87.6 (85.9 to 89.1)	87.8 (86.1 to 89.3)	93.2 (91.4 to 94.6)	93.4 (91.7 to 94.7)
	Treatment 2			91.1 (88.3 to 93.2)	91.0 (88.3 to 93.2)	88.5 (86.9 to 90.0)	88.1 (86.5 to 89.6)	92.9 (91.1 to 94.4)	93.6 (91.8 to 95.0)
	Control 1	89.2 (87.4 to 90.7)	89.3 (87.5 to 90.8)	93.6 (91.2 to 95.3)	93.3 (91.0 to 95.1)	90.1 (88.5 to 91.4)	88.4 (86.7 to 89.9)	93.1 (91.2 to 94.6)	92.8 (90.8 to 94.4)
	Control 2					89.2 (87.6 to 90.6)	88.1 (86.4 to 89.6)		
HR (95%CI), p-value (log rank)	Treatment 1 vs control 1	1.13 (0.96-1.33) p=0.16	1.17 (0.99-1.39) p=0.068	1.10 (0.75-1.61) p=0.63	1.02 (0.69-1.49) p=0.93	0.94 (0.82-1.07) p= 0.34	0.98 (0.85-1.12) p= 0.74	1.02 (0.74-1.40) p= 0.90	0.99 (0.72-1.37) p= 0.96
	Treatment 2 vs control 1			1.29 (0.90-1.86) p=0.17	1.26 (0.88-1.81) p=0.21			0.99 (0.72-1.36) p= 0.95	1.10 (0.79-1.53) p= 0.59
	Treatment 2 vs control 2					0.98 (0.86-1.12) p= 0.76	0.99 (0.86-1.13) p= 0.85		

Table A1: 3 year survival estimates and hazard ratios for treatment effects for TTR, iDFS and OS for each trial by NCRAS and trial data (continued).

Event	Measure	POETIC		IMPORT HIGH		TACT2		FAST-FORWARD	
		NCRAS	Trial	NCRAS	Trial	NCRAS	Trial	NCRAS	Trial
OS									
estimate (95% CI)	Treatment 1	95.2 (94.3 to 95.9)	95.2 (94.3 to 95.9)	96.8 (95.1 to 97.8)	96.6 (94.9 to 97.7)	93.8 (92.5 to 94.9)	93.9 (92.6 to 95.0)	96.8 (95.6 to 97.7)	96.7 (95.5 to 97.6)
	Treatment 2			95.6 (95.1 to 97.8)	95.6 (93.8 to 96.9)	94.3 (93.1 to 95.3)	94.3 (93.1 to 95.3)	97.2 (96.0 to 98.0)	97.2 (96.0 to 98.0)
	Control 1	94.3 (92.9 to 95.4)	94.3 (92.9 to 95.4)	96.0 (93.8 to 96.9)	96.0 (94.2 to 97.2)	95.4 (94.2 to 96.3)	95.4 (94.2 to 96.3)	97.1 (95.9 to 97.9)	97.0 (95.8 to 97.8)
	Control 2					94.9 (93.7 to 95.9)	95.0 (93.8 to 96.0)		
HR (95%CI), p-value (log rank)	Treatment 1 vs control 1	1.03 (0.85-1.26) p=0.75	1.05 (0.87-1.28) p=0.61	0.85 (0.54-1.34) p=0.49	0.93 (0.60-1.44) p=0.73	0.94 (0.80-1.11) p= 0.47	0.96 (0.81-1.13) p= 0.62	0.94 (0.63-1.41) p= 0.76	0.92 (0.62-1.37) p= 0.69
	Treatment 2 vs control 1			0.97 (0.63-1.50) p=0.90	1.02 (0.66-1.57) p=0.92			1.06 (0.70-1.61) p= 0.80	1.03 (0.69-1.55) p= 0.88
	Treatment 2 vs control 2					1.00 (0.84-1.18) p= 0.97	1.02 (0.86-1.20) p= 0.86		

POETIC Treatments:
Treatment 1:Peri AI
Control 1: No Peri AI

IMPORT-HIGH Treatments:
Treatment 1: 48 Gy/15 Fr concomitant dose esc
Treatment 2: 53 Gy/15 Fr concomitant dose esc
Control 1: 40 Gy/15 +16 Gy/8Fr sequential dose esc

TACT2 Treatments:
Treatment 1: Accelerated Epirubicin
Control 1: Standard Epirubicin
Control 1: 40.05 Gy in 15Fr
Control 2: CMF

FAST-Forward Treatments:
Treatment 1: 27.0 Gy in 5Fr
Treatment 2: 26.0 Gy in 5Fr
Control 1: 40.05 Gy in 15Fr

TACT2 Treatments:
Treatment 1: Accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of Classical/Bonadonna CMF & Accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of oral Capecitabine/Xeloda
Control 1: Non-accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of Classical/Bonadonna CMF & Non-accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of oral Capecitabine/Xeloda
Treatment 2: Non-accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of oral Capecitabine/Xeloda & Accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of oral Capecitabine/Xeloda
Control 2: Non-accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of Classical/Bonadonna CMF & Accelerated 4 cycles of Epirubicin every 2 weeks followed by 4 cycles of Classical/Bonadonna CMF

Table A2: 7 year survival estimates for TTR, iDFS and OS for TACT2 by NCRAS and trial data

Event	Measure	TACT2	
		NCRAS	Trial
TTR			
estimate (95% CI)	Treatment 1	81.9 (79.8 to 83.7)	83.2 (81.2 to 84.9)
	Treatment 2	82.2 (80.2 to 84.0)	83.2 (81.2 to 85.0)
	Control 1	81.9 (79.9 to 83.7)	82.5 (80.5 to 84.2)
	Control 2	81.6 (79.6 to 83.4)	82.4 (80.4 to 84.2)
iDFS			
estimate (95% CI)	Treatment 1	76.6 (74.4 to 78.6)	78.2 (76.0 to 80.1)
	Treatment 2	77.0 (74.8 to 78.9)	77.9 (75.8 to 79.9)
	Control 1	77.3 (75.2 to 79.3)	78.0 (75.9 to 79.9)
	Control 2	77.0 (74.8 to 79.0)	78.2 (76.1 to 80.2)
OS			
estimate (95% CI)	Treatment 1	85.4 (83.5 to 87.0)	85.4 (83.6 to 87.1)
	Treatment 2	86.1 (84.3 to 87.7)	86.2 (84.4 to 87.8)
	Control 1	86.9 (85.1 to 88.4)	86.9 (85.1 to 88.4)
	Control 2	86.2 (84.4 to 87.8)	86.1 (84.3 to 87.7)

TACT2 Treatments
Treatment 1: Accelerated Epirubicin
Control 1: Standard Epirubicin
Treatment 2: Capecitabine/Xeloda
Control 2: CMF