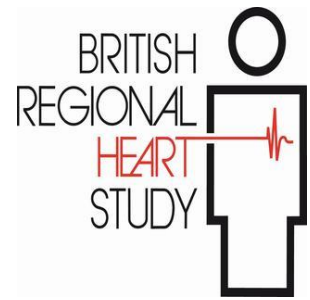


## BRHS Cancer follow-up data

Follow-up period Baseline 1978 (Q1) to 2024



Data file: BRHS CANCER FILE 7735 2025 P14.dat

Data dictionary: BRHS CANCER FILE 7735 2025 P14 Data Dictionary.xlsx

### Cancer Follow-up

All BRHS cohort participants ( $n = 7,735$ ) have been flagged for cancer notification through the Cancer Registry since the study baseline (1978–1980). Information has been received on the date of a cancer diagnosis registration and the type of cancer. Cancer type is provided using the International Classification of Diseases, 10th Revision (ICD-10).

These data were supplied by NHS England, NHS Scotland, or their predecessor organisations.

**To protect participant confidentiality, the following steps were taken:**

- Original 4-digit ICD-10 codes were recoded into less granular 3-digit ICD-9 codes.
- Dates of cancer diagnosis registration were perturbed by randomly shifting them by 1–14 days.

#### 1. Follow-up period

Baseline: 1978–1980 (Q1)

End of follow-up: 2024

This dataset includes all cancer events registered with the Cancer Registry during the follow-up period. Although events are recorded up to 9 August 2024, the data are likely to be complete for all participants only up to 30 June 2024, as some events may not yet have been reported.

**\*Recommended cutoff date for cancer events: 30 June 2024**

## 2. Dataset description

### 2.1 Data file structure

Each BRHS cohort participant may have up to seven cancer diagnoses. Accordingly, each participant record has seven lines of data, one for each potential cancer event. Events are listed in chronological order, with the earliest event first. For example, Line 1 contains the date and type of the participant's first cancer diagnosis.

The data file is in ASCII format and contains fixed-length records, with seven lines per participant for all 7,735 BRHS cohort participants.

### 2.2 Variables included in the BRHS Cancer Dataset

#### 2.2.1 Event number

Events are numbered from 1 to 7, indicating the order in which they occurred (e.g. 1st, 2nd, ..., 7th). Events are sorted chronologically within each participant's record.

#### 2.2.2 Event date – Cancer Registration Date

This is the cancer registration date for a cancer diagnosis (day, month, and year), provided as three separate variables, and is taken to be the date of diagnosis. The date has been perturbed by adding a random number of days (1–14) to the actual diagnosis date.

#### 2.2.3 Type of cancer

The type of cancer is classified using the International Classification of Diseases, 9th Revision (ICD-9), as an ICD-9 3-digit code.

### 2.3 Missing values

When no event was present, missing values were generated for the **event date** (day, month, year) and **cancer type**. These missing values are represented as dots (".") in the ASCII dataset.

A complete list of variables included in the data file is provided in **Table 1** below.

**TABLE 1: List of Variables included in the Cancer Registry Events Dataset**

<b>BRHS CANCER DATASET</b>			
<b>Variable description</b>	<b>Units/category labels</b>	<b>Methods (in sections above)</b>	<b>BRHS variable name</b>
Study ID	serial		serial
<b>1<sup>st</sup> Cancer event (Event 1)</b>	1-1	2.2.1	event_num1
Perturbed date of 1 <sup>st</sup> Cancer Registry diagnosis			
Date of 1 <sup>st</sup> Cancer Registry diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d1
Date of 1 <sup>st</sup> Cancer Registry diagnosis (Month)	1-12, ·=no event		pCancerRDate_m1
Date of 1 <sup>st</sup> Cancer Registry diagnosis (Year)	Year:1960 – 2024, ·=no event		pCancerRDate_y1
Type of 1 <sup>st</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_1
Study ID (repeated)	serial2		serial2
<b>2<sup>nd</sup> Cancer event (Event 2)</b>	2-2	2.2.1	event_num2
Perturbed date of 2 <sup>nd</sup> Cancer Registry diagnosis			
Date of 2 <sup>nd</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d2
Date of 2 <sup>nd</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m2
Date of 2 <sup>nd</sup> Cancer diagnosis (Year)	Year:1971 -2024, ·=no event		pCancerRDate_y2
Type of 2 <sup>nd</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_2
Study ID (repeated)	serial3		serial3
<b>3<sup>rd</sup> Cancer event (Event 3)</b>	3-3	2.2.1	event_num3
Perturbed date of 3 <sup>rd</sup> Cancer Registry diagnosis			
Date of 3 <sup>rd</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d3
Date of 3 <sup>rd</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m3
Date of 3 <sup>rd</sup> Cancer diagnosis (Year)	Year:1979 -2024, ·=no event		pCancerRDate_y3
Type of 3 <sup>rd</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_3
Study ID (repeated)	serial4		serial4
<b>4<sup>th</sup> Cancer event (Event 4)</b>	4-4	2.2.1	event_num4
Perturbed date of 4 <sup>th</sup> Cancer Registry diagnosis			
Date of 4 <sup>th</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d4
Date of 4 <sup>th</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m4
Date of 4 <sup>th</sup> Cancer diagnosis (Year)	Year:1980 -2024, ·=no event		pCancerRDate_y4
Type of 4 <sup>th</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_4
Study ID (repeated)	serial5		serial5
<b>5<sup>th</sup> Cancer event (Event 5)</b>	5-5	2.2.1	event_num5
Perturbed date of 5 <sup>th</sup> Cancer Registry diagnosis			
Date of 5 <sup>th</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d5
Date of 5 <sup>th</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m5
Date of 5 <sup>th</sup> Cancer diagnosis (Year)	Year:1991 -2023, ·=no event		pCancerRDate_y5
Type of 5 <sup>th</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_5
Study ID (repeated)	serial6		serial6
<b>6<sup>th</sup> Cancer event (Event 6)</b>	6-6	2.2.1	event_num6
Perturbed date of 6 <sup>th</sup> Cancer Registry diagnosis			
Date of 6 <sup>th</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d6
Date of 6 <sup>th</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m6
Date of 6 <sup>th</sup> Cancer diagnosis (Year)	Year:1992 -2021, ·=no event		pCancerRDate_y6
Type of 6 <sup>th</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_6
Study ID (repeated)	serial7		serial7
<b>7<sup>th</sup> Cancer event (Event 7)</b>	7-7	2.2.1	event_num7
Perturbed date of 7 <sup>th</sup> Cancer Registry diagnosis			
Date of 7 <sup>th</sup> Cancer diagnosis (Day)	1-31, ·=no event	2.2.2, 2.3	pCancerRDate_d7
Date of 7 <sup>th</sup> Cancer diagnosis (Month)	1-12, ·=no event		pCancerRDate_m7
Date of 7 <sup>th</sup> Cancer diagnosis (Year)	Year:2012 -2012, ·=no event		pCancerRDate_y7
Type of 7 <sup>th</sup> Cancer diagnosis	ICD-9 (3-digit code), ·=no event	2.2.3	CancerICD9_7