

# Cancer survival methodology

---

## About the NDRS

The National Disease Registration Service (NDRS) is part of NHS Digital (NHSD). Its purpose is to collect high-quality, timely data on cancer, rare diseases and congenital anomalies to monitor changes in the health of the population.

The NDRS includes:

- The National Cancer Registration and Analysis Service (NCRAS); and
- The National Congenital Anomaly and Rare Disease Registration Service (NCARDRS).

Healthcare professionals, researchers and policy makers use data to better understand population health and disease. The data is provided by patients and collected by the NHS as part of their care and support. The NDRS uses the data to help:

- understand cancer, rare diseases and congenital anomalies;
- improve diagnosis;
- plan NHS services;
- improve treatment;
- evaluate policy;
- improve genetic counselling.



National Disease Registration Service  
NHS Digital (NHSD)  
The Leeds Government Hub  
7&8 Wellington Place  
Leeds  
LS1 4AP

For queries relating to this document, please contact:

[NDRSenquiries@nhs.net](mailto:NDRSenquiries@nhs.net)

---

Improving lives with data and technology – NHS Digital support NHS staff at work, help people get the best care, and use the nation's health data to drive research and transform services.



# Contents

About the NDRS	1
Contents	2
Introduction	4
Cancer survival in England	4
Adult survival	4
Childhood survival	6
Index of cancer survival	7
Background	7
Survival methods	8
Adjusting for age, sex and cancer type	8
Conditional crude probabilities of deaths in England	10
Background	10
Survival methods	10
Concepts and definitions	11
Cancer terms	11
Cancer	11
Childhood cancer	11
Primary cancer	11
Metastatic or secondary cancer	11
Cancer stage (at diagnosis)	12
Patient follow-up information	13
Follow-up	13
Censor date	13
Lost to follow-up	13
Survival methods	13
Crude survival	13
Overall survival (Kaplan-Meier estimator)	14
Relative survival	14
Net survival	14
Life tables	14
Survival analysis approaches	14
	2

Cohort approach	15
Complete approach	15
Period approach	16
Hybrid approach	17
Geography	18
NHS England Regions	18
Cancer Alliances	18
Sustainability and Transformation Partnerships	18
Clinical Commissioning Group	18

## Introduction

This document provides the methodology for the survival publications produced by the National Cancer Registration and Analysis Service (NCRAS) within NHS Digital (NHSD). The publications and statistics included are:

- Cancer survival in England:
  - 1- to 5-year net survival from diagnosis for adults (15 to 99) split by:
    - Age-group
    - Sex
    - Stage at diagnosis (if appropriate)
    - Geographic area
  - 1-, 5- and 10-year overall survival for childhood cancers
- Index of cancer survival
- Conditional crude probability of deaths

The data used in all these publications is part of the [National Cancer Registration Dataset](#) , data collected are subject to quality assurance and control described in the methodology on Data collection and Quality assurance of administrative data for cancer registrations.

For adults, cancers are grouped using the [International Statistical Classification of Diseases 10th Revision](#) (ICD-10). Cancer sites are presented where there are enough diagnoses within the time period of study.

Sites are grouped, where appropriate, by broad physiological category. For example, some cancers of the lower digestive tract are combined to form colorectal (bowel) cancer.

The same cancer site definitions are applied across all the cancer survival publications and are also used in the Cancer registrations statistics publication.

All publications consist of data tables and an accompanying bulletin. The data tables are produced in OpenDocument Spreadsheet format (.ods) and the bulletin is produced in HTML format to ensure accessibility for all technologies accessing this output.

The most recent cancer survival statistics publications can be accessed on [cancer survival for England collection page](#). All of the publications classified as National Statistics also have a visualisation tool to explore the data, these are hosted on [cancerdata.nhs.uk](#) and a link to the relevant visualisation is provided in each bulletin.

## Cancer survival in England

### Adult survival

Adults cancer survival is produced for people aged 15-99. The age is limited to 99 to match the International Cancer Survival Standard (ICSS).

Many adult cancer patients will die from causes unrelated to their cancer diagnosis. To show only the effect of cancer deaths on patient survival, adult survival estimates are net survival estimates.

Net survival is an unbiased estimator that reflects the performance of the healthcare system for cancer patients. Net survival estimates are calculated by comparing the survival of cancer patients with that expected based on the general population of the same profile of age, sex and socio-economic status.

Survival estimates are produced for 1- to 5-year survival. The approach used to define the diagnosis years is the complete approach. The complete approach is used to estimate survival for 2-, 3-, 4- and 5-year estimates where time since diagnosis does not allow the full follow-up. Full follow-up is possible for 1-year survival because follow-up is obtained for all patients for the calendar year following the final diagnosis year included in each output.

Stage at diagnosis information is used to produce net survival estimate for cancer sites split by stage since it has been shown that tumours with earlier stage at diagnosis have better survival. [TNM](#) is the staging system used for most cancer sites and has 4 broad stages (for more detail see [Cancer stage \(at diagnosis\)](#)).

Estimates are presented for England and for two subnational geographies: Cancer Alliances (CAs) and Sustainability and Transformation Partnerships (STPs). These are the smallest geographical units for which reliable 5-year cancer survival estimates can be published.

The latest geographical health boundaries are always used. Health boundaries are routinely updated in the April of each year. Details on changes to geographical boundaries are available from the [Office for National Statistics \(ONS\) geoportal website](#).

After applying the inclusion and exclusion criteria outlined in the Data Collection and Quality assurance of administrative datasets section net survival is calculated using the [stns function](#) in Stata. Age-standardised net survival estimates for adults, stage and geographic areas are presented.

Age-standardisation allows for fair comparisons to be made between populations and over time. The estimates are standardised using the [International Cancer Survival Standard \(ICSS\)](#) weightings with 5 age-groups.

All age-groups must pass the statistical tests below:

- a minimum of 10 patients should be alive at the beginning of the survival period being estimated (for example, first year of follow-up for a 1-year estimate, fifth year of follow-up for a 5-year estimate and tenth year of follow up for a 10-year estimate)
- at least two deaths registered in the years before or after the duration(s) being estimated
- the level of the survival estimates should not increase with duration (for example, the survival estimated at five years following diagnosis should be lower than the survival estimated at one year following diagnosis)
- the standard error of the survival estimates should be lower than 20%

If any of these criteria are failed the age-groupings are reduced to 4 age-groups with a single pair of adjacent age-groups being combined and the same tests are performed. If the combined

groups still fail the tests then the estimate is suppressed and displayed as missing in the data tables.

Trends in cancer survival are presented for all geographical areas presented above which have been consistently defined for at least 9 years before the last diagnosis date included. A geographic area is defined consistently if it has been made up of the same Lower Layer Super Output Areas (LSOAs) during this time.

Trends in cancer survival are estimated as the change in net survival over the 5-year aggregated diagnosis periods starting from the 5 years that finish at least 9 years before the last diagnosis date included up to the most present aggregation i.e. for 2018 latest diagnosis date the first 5-year period would be 2006 to 2010.

Non-standardised estimates are used to estimate a trend to maximise the number of trends available since this figure is least likely to be suppressed. The trends are presented with an assessment of statistical significance only if the absolute difference in survival between two consecutive 5-year aggregated diagnosis periods does not exceed 20%.

These trends are not calculated for adult cancer survival by stage at diagnosis in England due to insufficient staging data available before 2012. A trend for stage will be added to the bulletin when there is a long enough time series available.

On completion of data preparation and analysis, consistency checks are applied to ensure result are valid and ready for publication. These can be classified into two categories:

- raw data checks; these include checking counts between each cancer site and that these are consistent with estimates provided in the Cancer registrations in England release
- sensitivity checks; these include checking outputs against previous years data

## Childhood survival

Overall survival is appropriate to use in the cancer survival for children in England publication because there is an extremely low level of mortality in children (excluding mortality in the first few months after birth) and almost all deaths in children diagnosed with cancer would be caused by their cancer.

The analyses were carried out using the cohort, period and hybrid approaches. 1-, 5- and 10-year survival is estimated for each publication. These different approaches are explained in more detail in the Concepts and Definitions section.

10-year survival estimates from 2002 to 2019 are based on the following approaches:

- cohort from 2002 to 2009
- period from 2010 to 2018
- hybrid for 2019

5-year survival estimates from 2002 to 2019 have been based on the following methods:

- cohort from 2002 to 2014
- period from 2015 to 2018

- hybrid for 2019

1-year survival estimates from 2002 to 2019 have been based on the following methods:

- cohort from 2002 to 2018
- hybrid for 2019

Age-group specific estimates are presented for children aged 0 to 4, 5 to 9 and 10 to 14 years, the estimates are age-standardised by giving equal weight to all 3 age-groups.

To reduce the volatility of the reported estimates (mainly as a result of relatively small numbers of diagnoses each year in children), [locally weighted regression smoothing](#) is applied to highlight underlying trends over time. This is presented as 1-, 5-, and 10-year smoothed survival for children.

Since overall survival is calculated in a simpler way compared to net survival, statistical inclusion criteria do not need to be applied after the estimations are calculated.

## Index of cancer survival

### Background

This publication includes adults only and presents 1-year age-(sex-cancer)-standardised net cancer survival for tumours diagnosed in England for individual years, followed up for at least one whole calendar year.

The analyses for the bulletin are estimated by use of flexible parametric models to produce survival estimates for adults for breast (women), colorectal and lung cancer separately and all cancers combined excluding non-melanoma skin cancer (NMSC) and prostate cancer.

NMSC is excluded because it is known to be underestimated and therefore unreliable for comparison purposes.

Prostate cancer is excluded from the index, because the widespread introduction of prostate-specific antigen (PSA) testing since the early 1990s has led to difficulty in the interpretation of short-term survival trends in local geographies, as explained in [excess cases of prostate cancer and estimated over diagnosis associated with PSA testing in East Anglia](#).

The estimates produced are:

- 1-year net cancer survival index for Clinical Commissioning Groups (CCGs) in England
- 1-, 5- and 10-year net cancer survival index for England, STPs and CAs in England

Note: although survival for CAs and STPs is published in both the index for cancer survival and cancer survival for England, due to differences in methodology they are not directly comparable since outputs are created using a flexible parametric model unlike other outputs mentioned above.



For geographic areas with small populations, like most CCGs, some fluctuation in survival estimates between consecutive years can be expected, as found in the following studies; [Cancer survival indicators for Clinical Commissioning Groups in England](#) and [Cancer survival indicators for primary care organisations in England](#).

Fluctuations in cancer survival by CCG can occur due to the small numbers of cancer diagnoses and deaths each year within the population. Therefore, a low survival figure for a single calendar year should not be over-interpreted. However, if the survival estimates in a given CCG are consistently low "outliers" for several years in a row, possible explanations should be considered.

These survival estimates must be interpreted with care. They do not reflect the survival prospects for any individual cancer patient; they represent the survival for all cancer patients in each area in a given period of time. The survival estimates should not be compared across geographies because the estimates are taken from independently run statistical models.

## Survival methods

Survival is estimated using a publicly available program [stpm2](#) in [Stata](#). The program calculates net survival using flexible parametric survival models, with age and year of diagnosis as main effects and an interaction between age and year of diagnosis.

Several models are fitted to allow up to five degrees of freedom for both the baseline hazard function and time-dependent effects. The best-fitting statistical model is selected by assessing the relative goodness of fit using the [Akaike Information Criterion \(AIC\)](#) and [Bayesian Information Criterion \(BIC\)](#), with scaling tests to check for oversensitivity and a likelihood ratio test to compare the best-fitting models according to AIC and BIC. A separate model is fitted for each combination of CCG, type of cancer and sex that are run independently of one another.

For each type of cancer, CCG and sex, the fitted models were used to estimate survival for 5 age-groups at diagnosis (15 to 44, 45 to 54, 55 to 64, 65 to 74 and 75 to 99 years) and each diagnosis year. For each CCG and diagnosis year, the all-cancers survival index was then calculated as weighted averages of the net survival estimates for each type of cancer, sex and age-group.

The precision values presented in the bulletin are calculated as the inverse of the variance of each survival estimate. More information on the models can be found in the following articles:

- [flexible parametric models](#)
- [further development of flexible parametric models](#)

More information on the AIC and BIC is also available:

- [Applications and developments of flexible parametric survival models](#)
- [Using model selection criteria to determine the optimal number of B-spline control points for areal deformation modelling](#)

## Adjusting for age, sex and cancer type

Survival estimates are age-standardised, to improve the comparability between population groups and over time. To produce the all-cancers combined index, the data also need to be standardised by sex and cancer type to allow for comparisons across the different populations.

Estimates are calculated using weights based on the [International Classification of Survival Standard \(ICSS\)](#) for age-standardisation, with additional weighting applied to standardise for sex and cancer type. Table 1 shows the weights used for standardisation.

For the all-cancers survival index all values were adjusted using the same set of standard weights. This means that the survival index can be compared over time, because the index is adjusted for any changes in the profile of cancer patients by age, sex or type of cancer. This adjustment is necessary because, without standardisation, changes in survival could result from changes in the profile of cancer patients.

For example, overall cancer survival in each CCG could change simply because of changes in the profile of its cancer patients, even if survival at each age, for each cancer and in each sex did not change.

**Table 1: Weights to adjust for age, sex and cancer type taken from International Classification of Survival Standard (ICSS)**

Cancer type	Age-group	Age weight	Male weight	Female weight	Cancer type weight	Final ICSS-based weight
Breast	15 to 44	0.070	-	1.000	0.167	0.012
	45 to 54	0.120	-	1.000	0.167	0.020
	55 to 64	0.230	-	1.000	0.167	0.038
	65 to 74	0.290	-	1.000	0.167	0.048
	75 to 99	0.290	-	1.000	0.167	0.048
Colorectal	15 to 44	0.070	0.500	0.500	0.167	0.006
	45 to 54	0.120	0.500	0.500	0.167	0.010
	55 to 64	0.230	0.500	0.500	0.167	0.019
	65 to 74	0.290	0.500	0.500	0.167	0.024
	75 to 99	0.290	0.500	0.500	0.167	0.024
Lung	15 to 44	0.070	0.500	0.500	0.167	0.006
	45 to 54	0.120	0.500	0.500	0.167	0.010
	55 to 64	0.230	0.500	0.500	0.167	0.019
	65 to 74	0.290	0.500	0.500	0.167	0.024
	75 to 99	0.290	0.500	0.500	0.167	0.024
Other	15 to 44	0.070	0.500	0.500	0.500	0.018
	45 to 54	0.120	0.500	0.500	0.500	0.030
	55 to 64	0.230	0.500	0.500	0.500	0.058
	65 to 74	0.290	0.500	0.500	0.500	0.073
	75 to 99	0.290	0.500	0.500	0.500	0.073

In the survival analyses, for any sub-group of age, sex, cancer type, geography and diagnosis year where the estimates do not meet the following quality criteria, the result is suppressed.

The quality criteria used are:

- for 1-year survival, a minimum of 3 patients within the sub-group or cumulative survival for those within the sub-group of at least 1 person-year
- for 5-year survival, a minimum of 5 patients within the sub-group or cumulative survival for those within the sub-group of at least 5 person-years
- for 10-year survival, a minimum of 5 patients within the sub-group or cumulative survival for those within the sub-group of at least 10 person-years

Normally, to age-standardise survival estimates for cancer of the breast (women) and age-sex-standardise for colorectal and lung cancers, a weighted average of all the survival estimates of the 5 age-groups (15–44, 45–54, 55–64, 65–74, 75–99) is used.

Since the number of patients in a CCG can be quite small, even for these very common cancers, it is sometimes impossible to produce robust estimates of survival for one or more of the age-groups. This happened most often for patients in the age-group 15 to 44 years.

In this situation, the missing values are replaced by the equivalent value for its parent geography i.e. if a CCG has a missing value it is replaced by the appropriate estimate from the CA it is a part of. This replacement of values is used for the individual cancer sites as well as the all-cancer combined index.

## Conditional crude probabilities of deaths in England

### Background

Probabilities of death were produced for adults only, split by age-group, cancer site and cause of mortality. The cause of mortality was split into cancer-mortality and other-cause mortality (non-cancer-cause mortality). The probabilities were also calculated based on whether the person had just been diagnosed with the cancer, survived 1 year or survived 3 years. The probability of death was calculated for 1-, 2-, 3-, 4- and 5-years post diagnosis.

The values are presented as cumulative estimated probabilities.

### Survival methods

Adult cancer mortality estimates are based on conditional crude probabilities, which are calculated using the methods presented by [Cronin and Feuer](#), in the complete approach. Briefly, in each time interval, the probability of death due to other causes, and the probability of death due to cancer, are applied to the probability of being alive at the beginning of the interval.

A correction is applied in order to account for events that could occur from both competing causes for a subsample of the patients in each interval. Conditional crude probabilities are also calculated for patients who survive a one and three years from diagnosis. Time since diagnosis was split into yearly intervals for calculations. The estimates are produced using a publicly available program [strs](#) function in [Stata](#).

Like the cancer survival in England publication, results are also presented as age-standardised estimates. Age-standardisation allows for fair comparisons to be made between population groups and over time. To age-standardise, the adult estimates use the [International Cancer Survival Standard](#) weightings with 5 age-groups.

To ensure high-quality results the whole series of mortality estimates were suppressed for each cancer site, age-group, sex-specific combination if the cumulative mortality estimates were not monotonically increasing at individual yearly time points.

## Concepts and definitions

### Cancer terms

#### Cancer

For adults, cancers are coded using the [International Statistical Classification of Diseases 10th Revision](#) (ICD-10). ICD-10 coding for cancer is based on the nature and anatomical site of the cancer.

Morphology and behaviour codes used can be found in the International Classification of Diseases for Oncology, Second Edition (ICD-O-2). Morphology codes denote the cell types in the cancer and behaviour codes say if the tumour is invasive (tumours that invade into surrounding tissues) or not.

For the purposes of adult cancer survival, the term “cancer” includes all tumours that are invasive, these are listed under site code numbers C00 to C97 in ICD-10 excluding non-melanoma skin cancer (C44).

Further details of the eligibility and exclusion criteria have been published in [Control of data quality for population-based cancer survival analysis](#).

#### Childhood cancer

All children (aged 0 to 14 years) resident in England who were diagnosed with a primary malignant neoplasm of any organ, or a non-malignant neoplasm of the brain and central nervous system (CNS), as defined in the third edition of the [International Classification of Childhood Cancer](#), are considered eligible for inclusion in the survival analyses.

Cancers of the skin other than melanoma and secondary and unspecified malignant neoplasms were excluded. These conditions equate to the ICD-10 site codes C00-C43, C45-C76, C80-97, D33 and D43.

Further details of the eligibility and exclusion criteria have been published in [Control of data quality for population-based cancer survival analysis](#).

#### Primary cancer

A primary cancer is the tumour that first develops in an identifiable part of the body, for example, the stomach, and usually gives the name to the type of cancer with which a patient is diagnosed.

#### Metastatic or secondary cancer

A metastatic or secondary cancer is a cancer that has spread from the primary cancer, which may be located within the same site as the primary cancer (local metastasis) or spread beyond the site of the primary cancer (distant metastasis).

The metastatic cancer should have the same underlying cell biology and morphology as the primary cancer. A spread of primary tumour cells within the system of lymph nodes is not usually considered to be metastatic cancer.

In the Cancer survival for England publication, cancers diagnosed at a metastatic stage are classified as stage 4.

## Cancer stage (at diagnosis)

Many common cancers have a staging system that aims to give an indication of how far the disease has progressed; stage is usually recorded at diagnosis although the stage of disease in a patient will vary over time. This is not true for all cancers; for instance, most brain cancers do not have a staging system.

Cancer stage at diagnosis is a measure of how far the primary tumour has grown when the patient first presents in hospital. It is measured and recorded according to internationally agreed standards, often as agreed by the [Union for International Cancer Control](#). The most common staging standard is sometimes called the TNM staging method and is based on three components:

- tumour size (the T component)
- nodal involvement of the lymphatic system (N)
- metastatic spread (M)

The [International Federation of Gynaecology and Obstetrics \(FIGO\)](#) system is used to stage gynaecological cancers, the [Ann Arbor](#) system or [Murphy St Jude system](#) are used to stage Hodgkin lymphomas, the [International Neuroblastoma Risk Group Staging System \(INRGSS\)](#) is used to stage neuroblastoma, the [Chang's staging system](#) is used to stage medulloblastoma; these all have 4 broad stages. The [International Staging System \(ISS\)](#) is used to stage myelomas, the Binet staging system is used to stage chronic lymphocytic leukaemia; these both have 3 stages.

NCRAS uses the TNM to complement the staging information provided by the other systems, except for cervix where only FIGO is used. This is because the previous version of FIGO for cervix did not include nodal status (N component of TNM). The new version of FIGO for cervix does now include nodal status, but is slightly different to current TNM staging.

Although the combinations of tumour size, nodal involvement and metastatic spread change by tumour type, generally there are four broad stages of cancer progression:

- stage 1: the primary tumour is usually small and is contained within the body organ in which the tumour started growing
- stage 2: although larger, the primary tumour has not spread to other parts of the body; spread to the lymphatic system may be included depending on the primary tumour site
- stage 3: the primary tumour is larger and may have spread into neighbouring parts of the body and into the lymphatic system
- stage 4: the primary tumour has spread to at least one other part of the body, creating a secondary or metastatic tumour

There are several reasons why a tumour cannot be staged, for example, some samples taken do not produce clear results and some patients are too unwell to undergo the surgery required to obtain sufficient tissue sampling for staging.

In the Cancer survival for England publication, missing stage is treated as a separate category and survival estimates are produced for patients with “unknown” stage alongside the other categories of known stage.

In the Cancer survival for England publication A cancer is considered unstageable if a staging system does not exist for its morphology and site (topography) combination. For example, a diagnosis of primary malignant melanoma of the ileum is not considered stageable.

Multiple myeloma has a separate staging system, the [International Staging System](#), which has three levels of disease progression. In common with some other cancer sites that are presented in the Adult cancer survival in England publication, the staging data for myeloma is not complete enough to be considered reliable enough for publication by stage.

## Patient follow-up information

### Follow-up

A measure of the patient’s time at risk of death following diagnosis. For example, the time from when a patient is diagnosed with cancer, until their date of death, embarkation (to a country outside of the NHS system) or the latest date they were known to be alive (censor date).

### Censor date

The censor date is the date a patient was last known to be alive, which may be the last time checks against medical and deaths records were undertaken. The publications covered by this document have a censor date of 31 December in the year following the most recent diagnosis year included i.e. for data going up to 2018 follow up would be taken to 31 December 2019. Where a patient cannot be determined to be alive or dead on the censor date using these checks, a patient is said to be lost to follow-up (or censored) on the last date where they were known to be alive.

### Lost to follow-up

If a patient cannot be determined to be alive or dead on the censor date, for example, because they have emigrated or because important identifiers to link datasets (such as NHS number, date of birth) contain an error that prevents automatic linkage, a patient is lost to follow-up on the date where they are last known to be alive that precedes the censor date. If a particular group of patients is lost to follow-up for reasons related to their cancer, then this is said to be informative censoring.

## Survival methods

### Crude survival

This is the simplest method for calculating cancer survival, by calculating the proportion of a group of cancer patients who are still alive at time(s) of interest following a diagnosis of cancer.

This method produces biased estimates of survival because it does not consider:

- the total amount of time a group of patients are living with cancer until death occurs or the censor date being reached
- how to deal with patients who are lost to follow-up

### Overall survival (Kaplan-Meier estimator)

To allow for the total amount of time for which patients are alive and also for those patients who are lost to follow-up, a more sophisticated and unbiased estimator is overall survival (more formally known as a Kaplan-Meier estimator). The [Kaplan-Meier estimator](#) is a (non-parametric) method, which calculates the cumulative probability of “all-cause” (any cause) survival.

### Relative survival

[Relative survival](#) is an estimate of the probability of survival from the cancer alone excluding other potential causes of death.

In relative survival, it is assumed that for a group of cancer patients:

$$\text{total mortality (1)} = \text{mortality from cancer (2)} + \text{mortality from other causes (3)}$$

This is saying that a cancer patient may die because of their cancer or another cause but not from both their cancer and another cause.

Measuring the total mortality (1) for a group of cancer patients can be calculated by applying the overall survival method. The mortality in the general population or from other causes (3) is calculated in [life tables](#). The mortality from cancer (2) can then be obtained from (1) and (3).

### Net survival

[Net survival](#) is a variant of relative survival that is preferred as a measure of cancer survival in adults because it is an unbiased estimator. Net survival estimates the survival of cancer patients compared with the background mortality that patients would have experienced if they had not been diagnosed with cancer.

The Pohar-Perme estimator of net survival is an unbiased version of relative survival that accounts for informative censoring bias.

### Life tables

Mortality for the general population is derived from [population life tables](#) published by NCRAS. Using these life tables, the mortality of cancer patients is compared with that of individuals in the general population who belong to the same single year of age (0 to 99 years), sex, population weighted quintile of the index of multiple deprivation (IMD) and region. Age is capped at 99 in order to align with the ICSS ages.

## Survival analysis approaches

In this section, the various approaches to forming groups of patients for estimating survival are illustrated. These situations cover the scenarios where all outcomes at the estimation time of interest are known and those where outcomes at the estimation time of interest are only partially known.

Tables 1 to 4 are survival approach diagrams that highlight the diagnosis year for the demonstrated approach and the patient years of follow-up included in that approach. A patient pathway begins with diagnosis in year zero when there are no years of follow-up, this continues right across the diagram increasing for each year of follow-up.

For example, patients diagnosed in 2012 with follow-up until 2019 have at least seven years of follow-up. These tables focus on 5-year survival, but the principles are also applicable to 10-year survival.

## Cohort approach

When follow-up information is available for each patient for at least one year, 1-year survival can be estimated using the (classical) cohort approach. For example, once follow-up information is available for each patient over the entire calendar year following their diagnosis, the cohort approach can be used to estimate 1-year survival by combining the conditional probabilities of survival to the end of each successive sub-period of the analysis.

Table 1 highlights 5-year survival using the cohort approach. This approach requires that at least five complete years' worth of potential follow-up are available for each patient considered. It is the simplest approach as all patients could be diagnosed in the same year and potentially followed up for the same length of time.

However, it could also be used for patients diagnosed in different years, for example, to calculate survival for patients diagnosed in 2012 to 2014 if all patients have full follow-up available for the latest year. The restriction with this approach is that it cannot be calculated until at least five years have passed.

**Table 1: Cohort approach for the most recent year with follow-up to 2019**

Diagnosis year	2012	2013	2014	2015	2016	2017	2018	2019
2012	0	1	2	3	4	5	6	7
2013		0	1	2	3	4	5	6
2014†			0†	1†	2†	3†	4†	5†
2015				0	1	2	3	4
2016					0	1	2	3
2017						0	1	2
2018							0	1

Notes

- † = 5-year survival for the most recent year.

## Complete approach

The complete approach to survival analysis, a variant of the classical cohort approach, is used when some patients may have been followed up for less than the full period. This approach uses all potential years of follow-up for patients diagnosed in a 5-year period.



The advantage to this approach is it combines timeliness and efficiency using all the available follow-up. A disadvantage is that this approach cannot be used to give an estimate for a single diagnosis year.

Table 2 highlights the complete approach for 5-year survival for the diagnosis years from 2012 to 2018 with follow-up to 2019. For example, it is viable to use the complete approach to produce 5-year survival estimates for patients diagnosed during 2014 to 2018 with follow-up until 31 December 2019, even though not every patient has had the opportunity to be followed up for the full five years.

In this example, the potential follow-up time varies between a single year and five years, depending on the year of diagnosis.

**Table 2: Complete approach for the most recent years with follow-up to 2019**

Diagnosis year	Follow-up							
	2012	2013	2014	2015	2016	2017	2018	2019
2012	0	1	2	3	4	5	6	7
2013		0	1	2	3	4	5	6
2014†			0†	1†	2†	3†	4†	5†
2015†				0†	1†	2†	3†	4†
2016†					0†	1†	2†	3†
2017†						0†	1†	2†
2018†							0†	1†

Notes

1. † = 5-year survival for the most recent year.

## Period approach

A period estimate of 5-year survival is a short-term prediction of survival for patients diagnosed in that period, on the assumption that they will experience the most recently observed conditional probabilities of survival in each year up to five years since diagnosis.

Table 3 shows that for each year of potential follow-up included they were from patients diagnosed in different years.

In this example, patients diagnosed in 2018 potentially have one year of follow-up, then patients diagnosed in 2017 potentially have two years of follow-up given that they survived the first year. This is then true for each successive year until patients diagnosed in 2013 are potentially followed up for the fifth year given they survived the fourth year.

**Table 3: Period approach for the most recent year with follow-up to 2019**

Diagnosis year	2012	2013	2014	2015	2016	2017	2018	2019
2012	0	1	2	3	4	5	6	7
2013		0	1	2	3	4	5†	6
2014			0	1	2	3	4†	5
2015				0	1	2	3†	4
2016					0	1	2†	3
2017						0	1†	2
2018							0†	1

Notes

1. † = 5-year survival for the most recent year.

## Hybrid approach

The hybrid approach, a variant of the period approach, is used for short-term predictions when the follow-up data are more recent than the incidence data. This short-term delay arises from the quality assurance processes applied in registering a cancer diagnosis.

These estimates assume that the probability of survival from the patients included would remain stable for the following five years. Since survival is generally improving over time, the hybrid estimate of survival will be lower than that which we can expect to observe five years from now, when the full cohort-wise estimates will be available. It has the advantage of being available several years sooner.

Table 4 highlights 5-year survival using the hybrid approach for the diagnosis years from 2012 to 2018 with follow-up to 2019. It is like the period approach but the first year of survival based on patients with follow-up from the year before. This is because typically registries only have registrations data up to one year behind potential follow-up. This method uses the cohort approach for the first year of follow-up for patients diagnosed in 2018, then uses the period approach for the remaining four years of potential follow-up for patients diagnosed between 2014 and 2017.

**Table 4: Hybrid approach for the diagnosis years from 2012 to 2018 with follow-up to 2019**

Diagnosis year	2012	2013	2014	2015	2016	2017	2018	2019
2012	0	1	2	3	4	5	6	7
2013		0	1	2	3	4	5	6
2014			0	1	2	3	4	5†
2015				0	1	2	3	4†

2016					0	1	2	3†
2017						0	1	2†
2018							0†	1†

## Notes

1. † = 5-year survival for the most recent year.

The hybrid approach can be used to predict estimates of 10-year survival, if it can be assumed that the conditional probabilities of surviving for patients diagnosed in current year are equal to those diagnosed over the full 10-year period of available data.

More information on the differences between survival approaches can be found in the article [estimating and modelling relative survival](#).

## Geography

In all our publications we use the latest geographical health boundaries as published, which are routinely updated in the April of each year. Details on [changes to geographical boundaries](#) are available.

## NHS England Regions

[NHS England Regions](#) cover healthcare commissioning and delivery in their area and provide professional leadership on finance, nursing, medical, specialised commissioning, patients and information, human resources, organisational development, assurance and delivery. Regional teams work closely with organisations such as [CCGs](#), local authorities, [Health and Wellbeing Boards](#) as well as General Practitioner (GP) practices.

## Cancer Alliances

[CAs](#) were established in late 2016 to bring together local senior clinical and managerial leaders representing the whole cancer patient pathway across a specific geography. CAs lead the local delivery of the Independent Cancer Taskforce's ambitions for improving services, care and outcomes for everyone with cancer.

## Sustainability and Transformation Partnerships

[STPs](#) were established in late 2016 as local partnerships between NHS organisations and councils. They set out practical ways to improve health and care services. They are built around the needs of the local population across whole areas, not just those of the individual organisations involved.

## Clinical Commissioning Group

[CCGs](#) were established as part of the Health and Social Care Act in 2012 and replaced Primary Care Trusts on 1 April 2013. CCGs are groups of general practices (GPs) which come together in each area to commission the best services for their patients and population.