SCOTTISH WIDOWS

In partnership with MACMILLAN CANCER SUPPORT

# Cancer Prevalence Projections to 2045: Technical Methodology

An explanation of the data sources, modelling approach and key assumptions underpinning the projected estimates.

Macmillan Cancer Support in partnership with Scottish Widows

## Contents



## 1. Background

This document explains how the cancer prevalence statistics were calculated in the report Living with and beyond cancer in 2045: see the future, support the people. It is intended to be read alongside

this report. This methodology document is aimed at those with an interest in understanding the calculation approach and assumes a basic understanding of statistical or actuarial principles.

Cancer prevalence statistics are published by the respective cancer registries in each of the four UK nations. These statistics are considered high quality, 'gold standard' estimates of the **number of people living with cancer at a point in time.** The statistics are calculated using purpose-built registers of cancers diagnosed, which themselves draw from NHS electronic patient records.

Between them, the national cancer registries collect information on every malignant cancer diagnosed in the UK. This includes details such as cancer type, age at diagnosis and survival outcomes. Because these systems are comprehensive and carefully validated, they are widely regarded as the most authoritative source of cancer statistics in the UK.

While these statistics are robust, they are not designed to provide a complete or forward-looking picture of cancer prevalence.

- They are time limited, meaning that only cancers diagnosed from the mid-1990s onwards are counted. This means that individuals diagnosed prior to this period are not included in their counts, leading to an underestimate of the total number of people living with or beyond cancer.
- The four nations also publish statistics using different breakdowns of age, sex and cancer type. This makes it difficult to produce a consistent UK-wide view of prevalence by detailed population group (for example, women aged 40–49 living with brain cancer).
- Published prevalence statistics are typically retrospective, reporting on the number of people living with cancer at a date two or more years before publication. This reflects the time required for cancer registry teams to collect, validate and quality assure the data.
- Importantly, no UK nation routinely publishes projections of future cancer prevalence for all cancer types.

This project was undertaken by Macmillan Cancer Support in partnership with Scottish Widows to extend the range of publicly available cancer prevalence statistics. Among the project's ambitions, we wanted to:

- Produce prevalence estimates using a consistent set of age groups and cancer type definitions across England, Scotland, Wales and Northern Ireland.
- Estimate complete prevalence - the number of people alive who have ever been diagnosed with cancer.
- Develop forward-looking projections of prevalence up to 2045, based on observed trends in cancer incidence, survival and population change.
- Construct the model exclusively using publicly available, aggregated and non-identifiable data. This approach complements the outputs produced by cancer registries, promotes transparency and demonstrates that individual-level data is not always required to produce robust population statistics.

The scope includes all malignant cancers (ICD-10 codes C00–C97), excluding non-melanoma skin cancer (C44). No Macmillan data was shared with Scottish Widows for this analysis.

This project constitutes technical actuarial work as defined by the Financial Reporting Council. This methodology and the actuarial information it contains have been prepared in compliance with TAS 100: Principles for Technical Actuarial Work. The data, assumptions, methods, and judgements used have been selected and applied in accordance with the Reliability Objective, ensuring that the

resulting actuarial information is relevant, transparent, complete and comprehensible, including the communication of material uncertainty. Limitations to this work are set out and justified within this document, in accordance with TAS requirements.

# 2. Methodology overview

The modelling process was split into two phases, which we have first summarised and then described in more detail:

**Phase one: Building the prevalence model framework**

This involved constructing a time series of observed cancer incidence, cancer survival and background mortality statistics, and defining the computational rules required to combine these components to estimate prevalence.

The initial model was developed and validated using England data, reflecting the greater depth and historical coverage available. The objective was to ensure that the modelling framework could accurately reproduce published prevalence statistics before extending it to projections and to the other nations.

**Phase two: Creating projections**

This extended the validated prevalence model framework by projecting forward age-specific incidence rates survival probabilities and population structure to 2045. These projected inputs were then combined using the same model structure to produce future prevalence estimates.

The projections are based on the continuation of observed trends in incidence and survival, subject to modelled tapering where appropriate. They do not incorporate explicit assumptions about future policy interventions, system capacity, innovation breakthroughs in diagnosis and treatment or societal behavioural changes that could influence incidence or survival. These results should therefore be interpreted as conditional projections rather than precise forecasts or predictions.

The fundamental design of the prevalence model in both phases was based on understanding two key trends:

1. **The number of cancer diagnoses occurring over time**

   In phase one, observed diagnosis counts were used directly, after applying cleaning and transformation processes

In phase two, diagnoses were estimated by projecting age-specific incidence rates using age-period models and applying these to projected population estimates from the Office for National Statistics (ONS). This approach separates the changes in underlying cancer risk from changes in population size and age structure.

Age-specific incidence rates describe the probability of a cancer diagnosis occurring to a specific population group defined by their age, sex and country. Applying these rates to projected population structures allows diagnosis counts to reflect demographic change independently of risk trends.

2. **The length of time people are living with cancer after their diagnosis**
   In both phases, survival was modelled using two components:
   - **Cancer-specific survival**: the probability of surviving cancer a given duration following diagnosis
   - **Background mortality**: the probability of death from other causes, based on mortality statistics for England & Wales published by the ONS.

   Combining these components ensures that long-term survival is not overstated, particularly at older ages.

In simple terms, prevalence at a given index date is calculated by summing, across all past diagnosis years, the number of individuals diagnosed multiplied by the probability that they remain alive at the index date. These calculations are performed at a granular level by age, sex, cancer type and country.

## Phase one: Building the prevalence model framework

To build the prevalence model, cancer incidence and survival statistics were assembled from official publications across the UK nations (see data sources section). These were cleaned and transformed into consistent time series structures suitable for modelling. Wherever possible, the most granular published tables were used.

Where data were unavailable or incomplete, limited imputation, interpolation and extrapolation were applied to create a consistent set of inputs across countries, cancers, ages and years. Examples are outlined later in this document.

All inputs were aggregated and non-identifiable, and no individual-level records were used.

Phase one then involved implementing the computational structure required to combine incidence, survival and background mortality to produce prevalence estimates.

Prevalence estimates were generated for England for an index date of 31 December 2021 and compared against published England prevalence statistics on a comparable basis. This validation process tested whether the model structure and assumptions produced credible prevalence outputs when fed observed incidence and survival data. Where discrepancies were identified, the modelling assumptions were iteratively refined until the model closely aligned with published prevalence. Only then was it extended to projections and to cover the other UK nations.

The purpose of this validation was to check that incidence and survival were being appropriately combined and that the model produced credible prevalence estimates. It was not intended to achieve an exact match for every cancer type and population group.

### Phase two: Creating projections

To project cancer prevalence, the model first generated projections of:

- Age-specific cancer incidence rates
- Cancer survival probabilities
- Background mortality

These projections were derived from the time series assembled in Phase one using statistical modelling approaches designed to extrapolate recent trends into the future, with tapering applied where appropriate to avoid implausible long-run drift.

The projected incidence, survival and mortality inputs were then applied within the validated prevalence framework developed in Phase one. This combined projected diagnoses with projected survival to estimate the number of people expected to be living with cancer in future years through to 2045.

## 3. Modelling cancer incidence

This section describes how historic and projected cancer diagnoses were estimated for each cancer type, age group, sex and UK nation.

**Purpose**

The objective of the incidence modelling was to estimate:

- The number of cancer diagnoses occurring historically, and
- The number expected to occur in future years under projected trends.

Incidence is a primary driver of prevalence. Accurate modelling of diagnosis counts is therefore essential to generating credible prevalence estimates.

**Data used**

Published cancer incidence statistics were sourced for England, Scotland, Wales and Northern Ireland. These datasets were cleaned, standardised and restructured into a consistent format across:

- Cancer type (mapped to ICD-10 codes)
- Sex
- Age group
- Year of diagnosis.

A list of data sources is included at the end of this report.

Where published tables were incomplete or varied in structure across years or nations, limited imputation and proportional allocation methods were used to create consistent time series inputs.

Time series were constructed separately for each country. England's longer historical coverage and greater level of detail were used to inform aspects of the modelling framework, including proportional allocation where data for other nations were less granular.

**Data adjustments**

Where age breakdowns were incomplete in published data, diagnosis counts were apportioned across age groups using the closest available years with detailed data.

For Northern Ireland, where age distributions were not published at sufficient granularity, diagnosis totals were distributed across age groups using Scottish age patterns for the same cancer and sex over a similar time period.

Cancer type definitions varied across publications and over time. Some datasets reported grouped cancer categories, while others reported individual ICD-10 codes. To achieve consistency:

- Cancer types were mapped to a common ICD-10 framework (C00–C97 excluding C44)
- Where grouped ICD-10 codes were reported (e.g. C00-C14) incidence was recorded at the group level. The most common individual ICD-10 code in the English data was used as the representative cancer type for the group for the purposes of later joining to survival statistics
- Where cancer types included were non-exhaustive, residual incidence was proportionally allocated across ICD codes using distributions from adjacent years with fuller detail

This harmonisation ensured a consistent cancer classification structure across nations and time periods.

**Projection of incidence rates**

Future incidence was projected by modelling age-specific incidence rates rather than raw diagnosis counts. Projecting rates (rather than counts) ensures that changes in population size and age structure do not artificially inflate or suppress underlying cancer risk trends.

The projection fits a Lee-Carter age-period model to historic incidence rates for each sex & cancer type and uses this to estimate future incidence, as described in a mortality context by Cairns [2009] [1]. Projected trends were subject to tapering to prevent implausible long-term divergence. In age groups or cancer types where modelling was not feasible due to sparse data, a simpler five-year average rate method was used as a fallback.

**Key assumptions**

Recent age-specific incidence trends provide a reasonable basis for short-to-medium-term extrapolation.

Cohort effects were assessed and judged not to materially alter projections relative to an age-period specification.

Tapering of projected trends reduces the risk of unrealistic long-run acceleration.

Where data from all nations of the UK were incomplete, proportional relationships to England were assumed to remain stable in projection years.

**Limitations**

Historic incidence publications varied in granularity across nations and years, requiring harmonisation and limited imputation.

The COVID-19 period (particularly 2020) disrupted diagnosis volumes. Averaging methods were used cautiously where relevant.

Projected incidence reflects continuation of modelled trends and does not explicitly incorporate future policy interventions, technological advancements, system capacity or shifts in societal risk factor behaviour such as smoking rates.

Direct comparisons of projected incidence growth between nations should be interpreted cautiously, given differences in historic data structure and modelling adjustments.

# 4. Modelling population estimates

**Purpose**

Population modelling allows projected cancer diagnoses and survival to be translated into projected prevalence within a changing demographic context.

[1] Cairns, A. J. G. *et al.* (2009) 'A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States', *North American Actuarial Journal*, 13(1), pp. 1–35. doi: 10.1080/10920277.2009.10597538.

**Data used**

Published mid-year population estimates were used for each UK nation. These provided population counts by single year of age and sex, covering the period 1970 to 2021.

Where detailed age breakdowns were unavailable for earlier historic years, age distributions from the nearest available year were applied. This adjustment affected a small number of early historic years and has negligible impact on projected prevalence for the years of interest (2025 onwards).

For Northern Ireland, where full projected population data were not available separately, population estimates were derived as the difference between UK totals and the sum of England, Scotland and Wales. This approach preserves internal consistency with official UK population totals but may introduce minor rounding effects.

**Projection approach**

Published ONS population projections were used for each nation, providing projected mid-year population counts by single year of age and sex through to 2045. These projections were applied directly without adjustment.

**Key assumptions**

Official ONS population projections provide a reasonable baseline demographic trajectory.

**Limitations**

Population projections are themselves uncertain and subject to revision.

Northern Ireland projections were derived indirectly rather than sourced directly.

No alternative demographic scenarios (e.g. high/low migration variants) were modelled within this analysis. Alternative demographic scenarios could result in different projected prevalence totals. However, while demographic changes are an important driver of the growth in cancer prevalence they are not the primary source of uncertainty within the prevalence projections. The population projections produced by the ONS have well established methodologies and a clear basis for the assumptions implicit in them: the uncertainty due to the modelling of incidence and survival rates is more significant within the cancer prevalence.

# 5. Modelling cancer survival rates

**Purpose**

Survival modelling estimates the probability that a person diagnosed with cancer remains alive at future time points. These survival probabilities are combined with incidence data to determine how many individuals from each diagnosis year contribute to prevalence at a given index date. Accurate modelling of survival is therefore central to estimating both historic and projected prevalence.

**Data used**

Published cancer survival statistics were sourced for England, Scotland and Wales. These typically report 1, 5 and 10-year net survival probabilities by cancer type, age group and sex.

Northern Ireland survival statistics were not available at sufficient granularity for modelling purposes. Accordingly, England survival patterns were applied to Northern Ireland incidence data. England data was chosen for having higher availability of granular data points over a longer period of time.

**Expansion of survival curves**

Published survival statistics are typically reported at discrete durations (e.g. 1, 5 and 10 years post-diagnosis). To integrate survival with annual incidence data, these points were expanded into complete survival curves covering each year following diagnosis up to a maximum of 10 years.

Where survival was only available for specific time points, e.g. 1-year, 5-year and 10-year, intermediate years were estimated using statistical interpolation. Where data were sparse, reasonable and conservative assumptions were applied to ensure internally consistent survival trajectories.

**Projection of survival trends**

Historic survival improvements for England were modelled and projected forward using GLM techniques. Age at diagnosis, year of diagnosis and years since diagnosis plus interaction terms were fitted to historic survival rates and where the year of diagnosis was found to be statistically significant the model was used to predict future survival rates. As with incidence, projected survival improvements were tapered in later projection years to reduce the risk of unrealistic long-run divergence.

No explicit assumptions were made regarding future treatment breakthroughs or structural changes to cancer care. Projected improvements purely reflect continuation of observed trends.

**UK nation adjustments**

Scotland and Wales report significantly fewer cancer cases compared to England. When comparing survival rates, variations by cancer type and age group can occur, often due to statistical fluctuations from small case numbers which may not persist in subsequent periods.

However, at a higher level of aggregation, genuine differences in survival rates between nations become evident, and these should be accounted for in the modelling. To address this, while minimizing over-reliance on less reliable data, survival rate estimates for Scotland and Wales are calculated as a weighted average of the relevant English survival rate and the same from the

respective home nation. The weighting reflects the number of observations available for each home nation's rate.

As a result, differences between national estimates are smoothed and proportionally applied, with greater weight given to cancer types and age groups with more robust data. These adjustments are then used to modify projected survival rates for England, producing projections for Scotland and Wales. This method assumes relative survival differences between nations remain generally consistent throughout the projection period.

For Northern Ireland, due to limited data, English survival curves were used directly.

**Background mortality interaction**

Cancer-specific survival was combined with background general population mortality to ensure that survival beyond the observed cancer-specific follow-up period was not overstated. This is particularly important for longer durations post-diagnosis and older age groups.

**Key assumptions**

Published 1, 5 and 10-year net survival points provide sufficient basis to reconstruct full survival curves.

Relative differences in survival between England and Scotland/Wales remain broadly stable over time.

Survival improvements continue in line with recent observed trends, subject to tapering.

Childhood cancer survival was not assumed to improve beyond the most recent observed data, reflecting limited trend information and small absolute case numbers.

England survival is a reasonable proxy for Northern Ireland in the absence of sufficiently granular data.

**Limitations**

Survival data availability varies by nation and cancer type, requiring interpolation and harmonisation.

For some cancers and recent years, minor adjustments were applied where validation indicated inconsistencies between published survival and prevalence outputs.

Survival modelling does not explicitly incorporate changes in stage at diagnosis or treatment effectiveness.

Applying England survival to Northern Ireland may mask genuine national differences.

Long-term survival beyond published follow-up periods relies partly on background mortality assumptions.

Survival assumptions are a material driver of projected prevalence, particularly for cancers with improving long-term outcomes.

# 6. Modelling background mortality

**Purpose**

Background mortality modelling accounts for the risk that individuals diagnosed with cancer may die from causes unrelated to their cancer. Incorporating general population mortality ensures that long-term survival, and therefore prevalence, is not overstated, particularly for older age groups and for diagnoses many years in the past. Additionally, since background mortality includes mortality caused by cancer, the two sets of mortality/survival probabilities need to be carefully integrated and not simply summed.

**Data used**

General population mortality data were sourced from published ONS life tables.

For future years, projected mortality improvements were modelled using a standard actuarial projection framework.

A consistent background mortality basis was applied across all four UK nations.

**Modelling approach**

Cancer survival statistics are typically reported net of background mortality and only up to a fixed follow-up period (commonly 10 years post-diagnosis).

To estimate survival over longer durations:

- Cancer-specific mortality rates for the first 10 years were added to general population mortality for each age and year to give overall one-year mortality & survival rates.
- 10-year survival was calculated as the product of these one-year survival rates
- Beyond the 10-year follow-up period, adjusted background mortality rates were applied to avoid overstating long-term survival.
- Survival probabilities were aligned to mid-year index dates to ensure consistency between incidence timing and prevalence estimation.

This approach ensures that individuals diagnosed many years previously do not unrealistically accumulate in prevalence totals.

Without incorporating background mortality beyond the observed cancer follow-up window, prevalence estimates would be systematically inflated, particularly for cancers with good long-term survival.

**Key assumptions**

General population mortality projections provide a reasonable estimate of non-cancer mortality risk.

The same mortality projection framework can be applied consistently across UK nations.

Adjustments applied beyond the observed cancer follow-up period prevent overstating survival without materially distorting shorter-term outcomes.

Future mortality improvements follow the core projected trajectory and do not incorporate alternative high/low scenarios.

**Limitations**

A common background mortality basis was used across all UK nations, whereas in reality, mortality rates differ slightly between countries.

Long-term prevalence estimates (for diagnoses more than 10 years prior) are sensitive to assumptions about cancer specific or related mortality beyond the 10-year published statistics.

Alternative mortality improvement scenarios could produce modestly different projected totals.

However, background mortality is generally a second-order driver relative to incidence and cancer-specific survival trends.

# 7. Double-counting adjustments

**Purpose**

Cancer incidence statistics count tumour diagnoses. However, prevalence estimates aim to count people living with cancer.

Some individuals are diagnosed with more than one independent primary cancer over their lifetime. Without adjustment, simply summing prevalence across cancer types would overstate the total number of people living with cancer.

**Modelling approach**

To estimate prevalence for “all cancers combined”, an adjustment was applied to account for individuals with multiple primary diagnoses. The approach assumes that, within each age group and year, new cancer incidence occurs independently of prior diagnoses. Under this assumption:

- Age-specific prevalence was first calculated as a proportion of the underlying population.
- New incidence was then reduced proportionally to reflect the share of the population already living with cancer.
- The calculation was iterated to ensure internal consistency between incidence and prevalence totals.

This produces an adjusted estimate of the number of individuals living with at least one cancer diagnosis.

**Key assumptions**

The probability of a new cancer diagnosis is independent of having had a prior diagnosis.

The proportional adjustment adequately captures the aggregate effect of multiple primaries at population level.

Iterative convergence provides a stable population-level estimate.

**Limitations**

The independence assumption is a simplification. In reality, individuals with a prior cancer diagnosis tend to have an elevated risk of subsequent cancers due to shared risk factors, genetic predisposition or treatment-related effects.

This adjustment primarily affects the “all cancers combined” total and does not materially alter site-specific prevalence estimates, as the risk of multiple of the same type of cancer is less than the risk of having any second cancer diagnosis.

The estimate for double counting in 2021 is close to the difference between all-cancer prevalence and the sum of site-specific prevalence in the published NDRS Cancer Prevalence statistics, giving confidence that the approach is not unreasonable.

**Treatment of 5-year prevalence**

For 5-year prevalence estimates, no double-counting adjustment was applied. In this context, the number of diagnoses occurring within the previous five years was assumed to approximate the number of individuals diagnosed within that period.

This simplification is considered reasonable because:

- The probability of multiple independent primary cancers within a five-year window is relatively low at population level
- Any overstatement arising from multiple primaries within five years is expected to be modest relative to total prevalence

However, 5-year prevalence totals may marginally overstate the number of individuals if multiple primary diagnoses occur within the five-year window.

# 8. Attribution of underlying drivers of change

**Purpose**

Projected changes in cancer prevalence arise from three interacting components:

- Changes in cancer incidence rates

- Changes in cancer survival
- Changes in population size, age structure and background survival

To illustrate the relative contribution of incidence, survival and demographic change to projected prevalence growth, we decomposed the projected change between the baseline year and 2045.

**Modelling approach**

To understand what was driving projected growth we ran a series of alternative scenarios.

Starting from the validated baseline year:

- Population structure was projected forward alongside background mortality projections while holding incidence and cancer survival rates constant
- Incidence trends were then introduced, holding cancer survival constant
- Survival trends were then introduced

By comparing prevalence under each scenario, the incremental contribution of each component can be estimated. This approach allows projected growth in prevalence to be expressed in terms of:

- Population size and structure, incorporating general life expectancy
- Incidence trend effects
- Cancer survival

Because incidence, survival and demographic effects interact multiplicatively within the model, the attribution reflects the order and structure of the decomposition rather than an exact causal partition.

# 9. Limitations and Interpretation

Cancer prevalence projections necessarily involve modelling assumptions and harmonisation of multiple data sources. The following considerations are important when interpreting the estimates presented in this report.

**Data harmonisation across nations**

Cancer incidence and survival publications differ in structure, granularity and historical coverage across England, Scotland, Wales and Northern Ireland.

To produce consistent UK-wide estimates:

- Cancer types were mapped to a common ICD-10 framework.
- Missing age breakdowns were proportionally allocated using adjacent years or comparable nations.
- Northern Ireland survival was proxied using England survival patterns.

These adjustments were made carefully to produce robust estimates for each nation. However, since they each rely on slightly different sets of model assumptions, cross-nation comparisons should be avoided or at the very least interpreted with appropriate caution.

**Projection uncertainty**

The projections assume that recent trends in age-specific incidence rates and cancer survival continue broadly in line with modelled trajectories.

The model does not explicitly simulate any future changes to underlying drivers that could influence incidence and survival rates, including:

- Changes in screening uptake and cancer prevention initiatives
- Shifts in stage at diagnosis
- Innovations in diagnostics and treatments
- Changes in risk factor prevalence
- Changes in health system capacity
- Policy reform
- Increased public awareness, such as through media coverage

If future developments materially alter incidence or survival trends, actual prevalence will differ from these projections. These figures should therefore be interpreted as conditional projections, not forecasts of what will necessarily occur.

The population projections used in this analysis are based on the ONS principal projection, which is conditional on assumptions about future fertility, mortality and net international migration. Alternative variant projections are published by ONS to illustrate the uncertainty around these demographic drivers.

**Survival modelling constraints**

Published survival data are typically available at discrete durations (e.g. 1, 5 and 10 years) and for limited follow-up periods. Reconstruction of full survival curves required interpolation and modelling assumptions.

In addition:

- Relative survival differences between England and other nations were assumed to remain broadly stable over time.
- Long-term survival beyond observed follow-up periods depends partly on general population mortality assumptions.
- Survival assumptions are a material driver of long-term prevalence projections, particularly for cancers with improving long-term outcomes.

**Double counting and multiple primaries**

Incidence statistics count tumour diagnoses, whereas prevalence aims to count individual people. To avoid overstating the number of people living with cancer, an adjustment was applied when estimating “all cancers combined” prevalence.

This adjustment assumes independence between prior and subsequent diagnoses at population level. In reality, individuals with one cancer tend to have increased risk of another. The assumption is considered appropriate for population-level modelling, but is acknowledged as an oversimplification and introduces modest structural uncertainty into total prevalence estimates. Site-specific prevalence figures are only minimally affected by this adjustment.

**Mortality assumptions**

Background mortality projections were applied consistently across nations. In practice, mortality patterns differ slightly between countries.

Long-term prevalence (particularly for diagnoses more than 10 years prior) is sensitive to assumptions about mortality improvement. Alternative mortality scenarios would result in modestly different projected totals.

**Interpretation of attribution analysis**

The decomposition of projected growth into demographic, incidence and survival components is model-based. Because these components interact multiplicatively, attribution figures reflect the structure of the model rather than a precise causal partition of real-world drivers. They should be interpreted as illustrative of relative contributions within the modelling framework.

**Overall interpretation**

These projections provide a structured and internally consistent estimate of how cancer prevalence could evolve under continuation of recent incidence, survival and demographic trends.

They are intended to inform strategic discussion, planning and public understanding. They are not precise predictions and should be interpreted alongside other evidence and emerging data.

# 10. Sex and gender in cancer statistics

All estimates in this report are derived from routinely collected NHS patient records. Within these systems, the variable typically labelled “gender” or “sex” is not recorded consistently according to a single definition. In some cases it reflects biological sex assigned at birth and in others it reflects gender as recorded or presented at the point of care. We therefore cannot assume conceptual consistency across providers, time periods, or clinical contexts. Nor can we reliably identify whether an individual recorded as male or female is cisgender or transgender, nor whether their recorded sex corresponds to sex assigned at birth.

Furthermore, the recorded category is binary, allowing only for values ‘male’ and ‘female’. There are no published statistics for non-binary or other gender identities that could have been used for this analysis. There is also no consistent way to reliably distinguish sex assigned at birth from gender identity within the data. Any growing population of transgender, non-binary and other gender-diverse people living with cancer cannot be quantified using these data.

These limitations directly affect interpretation of any statistics presented by “male” and “female”. As such, the categories “men”/”male and “women”/”female” should be interpreted as shorthand for “People recorded as male in NHS data systems” and “People recorded as female in NHS data systems”.

To improve clarity, where appropriate we use phrasing such as “Men and people assigned male at birth” and “Women and people assigned female at birth”. This language reflects clinical realities (e.g. organ-specific cancers) while acknowledging the recording limitations of the underlying data.

Prostate cancer statistics will therefore be reported for:

- Those assigned male at birth, present as male to clinicians and were recorded as male in patient records and
- Those assigned male at birth, present as female to clinicians and were recorded as male in patient records

However, it may be excluding those who were assigned male at birth, present as female to clinicians and were recorded as female in patient records. Prostate cancer statistics are only reported for men and diagnoses in female records will in some cases be excluded from analysis. We cannot determine the extent to which such exclusions occur.

Statistics for other cancer types are also at risk of exclusions and inconsistencies.

As the number of people openly identifying with diverse gender identities increases, this limitation becomes increasingly material for population statistics and service planning. Accurate and inclusive population statistics require clear and standardised recording of sex assigned at birth, gender identity, structured options beyond a binary classification and transparent guidance on clinical coding rules for sex-specific cancers.

Until such practices are embedded consistently across NHS systems, it will remain challenging to produce robust, disaggregated cancer statistics for transgender, non-binary, or other gender-diverse populations. This report therefore reflects the current structural limitations of routinely collected administrative data, and findings should be interpreted within that context.

# 11. Table of ICD10 codes and cancer types

The following lookup table was used to translate ICD10 codes into a set of cancer types for reporting on.

| ICD-10 | NCRAS description | Cancer type |
|---|---|---|
| C00 | Lip cancer | Head and neck |
| C01 | Tongue cancer (base) | Head and neck |
| C02 | Tongue cancer (other) | Head and neck |
| C03 | Mouth cancer (gum) | Head and neck |
| C04 | Mouth cancer (floor) | Head and neck |
| C05 | Mouth cancer (palate) | Head and neck |
| C06 | Mouth cancer (other) | Head and neck |
| C07 | Salivary gland cancer (parotid) | Head and neck |
| C08 | Salivary gland cancer (other) | Head and neck |
| C09 | Tonsil cancer | Head and neck |
| C10 | Pharyngeal cancer (oropharynx) | Head and neck |
| C11 | Pharyngeal cancer (nasopharynx) | Head and neck |
| C12 | Pharyngeal cancer (piriform sinus) | Head and neck |
| C13 | Pharyngeal cancer (hypopharynx) | Head and neck |
| C14 | Head and neck cancer (other) | Head and neck |
| C15 | Oesophageal (gullet) cancer | Upper GI |
| C16 | Stomach cancer | Upper GI |
| C17 | Small intestine cancer | Upper GI |
| C18 | Colon cancer | Colorectal |
| C19 | Rectosigmoid junction cancer | Colorectal |
| C20 | Rectal cancer | Colorectal |
| C21 | Anal cancer | Other cancer types |
| C22 | Liver cancer | Upper GI |
| C23 | Gall bladder cancer | Upper GI |
| C24 | Biliary tract cancer | Upper GI |
| C25 | Pancreatic cancer | Upper GI |
| C26 | Digestive organ cancer (other) | Other cancer types |
| C30 | Nasal cavity and middle ear cancer | Head and neck |
| C31 | Accessory sinus cancer | Head and neck |
| C32 | Larynx cancer | Head and neck |
| C33 | Trachea (windpipe) cancer | Lung |
| C34 | Lung cancer | Lung |
| C37 | Thymus cancer | Other cancer types |
| C38 | Heart, mediastinum and pleura cancer | Other cancer types |
| C39 | Respiratory and thoracic cancer (other) | Lung |
| C40 | Bone and cartilage cancer (limbs) | Sarcoma |
| C41 | Bone and cartilage cancer (other) | Sarcoma |

| | | |
|---|---|---|
| C43 | Melanoma | Melanoma |
| C45 | Mesothelioma | Lung |
| C46 | Kaposi's sarcoma | Other cancer types |
| C47 | Peripheral nerves/autonomic nervous system cancer | Sarcoma |
| C48 | Peritoneal cancer | Sarcoma |
| C49 | Soft tissue sarcomas | Sarcoma |
| C50 | Breast cancer | Breast |
| C51 | Vulva cancer | Gynaecology |
| C52 | Vagina cancer | Gynaecology |
| C53 | Cervical cancer | Gynaecology |
| C54 | Uterine cancer (corpus uteri) | Gynaecology |
| C55 | Uterine cancer (other) | Gynaecology |
| C56 | Ovarian cancer | Gynaecology |
| C57 | Female genital organ cancer (other/unspecified) | Gynaecology |
| C58 | Placenta | Gynaecology |
| C60 | Penis cancer | Urology, excl prostate |
| C61 | Prostate cancer | Prostate |
| C62 | Testicular cancer | Urology, excl prostate |
| C63 | Male genital organ cancer (other/unspecified) | Urology, excl prostate |
| C64 | Kidney cancer (except renal pelvis) | Urology, excl prostate |
| C65 | Kidney cancer (renal pelvis) | Urology, excl prostate |
| C66 | Kidney cancer (ureter) | Urology, excl prostate |
| C67 | Bladder cancer | Urology, excl prostate |
| C68 | Kidney cancer (other) | Urology, excl prostate |
| C69 | Eye cancer (ocular melanoma) | Brain/CNS |
| C70 | Meningeal cancer | Brain/CNS |
| C71 | Brain tumours | Brain/CNS |
| C72 | Spinal cord tumours | Brain/CNS |
| C73 | Thyroid cancer | Head and neck |
| C74 | Adrenal gland cancer | Other cancer types |
| C75 | Endocrine tumours | Other cancer types |
| C76 | Other and ill-defined sites | Other cancer types |
| C77 | CUP (lymph nodes) | Other cancer types |
| C78 | CUP (respiratory and digestive) | Other cancer types |
| C79 | CUP (other specified sites) | Other cancer types |
| C80 | CUP (unspecified site) | Other cancer types |
| C81 | Lymphoma (Hodgkin) | Haematology |
| C82 | Lymphoma (follicular) | Haematology |
| C83 | Lymphoma (non-follicular) | Haematology |
| C84 | Lymphoma (mature T/NK cell) | Haematology |
| C85 | Lymphoma (non-Hodgkin, other) | Haematology |
| C86 | Lymphoma (other T/NK cell) | Haematology |
| C88 | Lymphoma (Immunoproliferative diseases, B cell) | Haematology |
| C90 | Myeloma and malignant plasma cell | Haematology |
| C91 | Leukaemia (lymphoid, other) | Haematology |
| C92 | Leukaemia (myeloid, other) | Haematology |

| | | |
|---|---|---|
| C93 | Myeloid leukaemia, unspecified | Haematology |
| C94 | Leukaemia (other, specified) | Haematology |
| C95 | Leukaemia (other, unspecified) | Haematology |
| C96 | Lymphoid (hematopoietic tissue, other) | Haematology |
| C97 | Independent (primary) multiple sites | Other cancer types |

# 12. List of data sources used

*Cancer survival for England and Wales articles (1986-2001; British Journal of Cancer, 2008)*

| Data Item | Years | Source | Date Accessed |
|---|---|---|---|
| England Incidence | 2021 | National Disease Registration Service (NDRS)/NHS Digital | 19/09/2024 |
| England Incidence | 2020 | NHS Digital | 19/09/2024 |
| England Incidence | 2016-2019 | Public Health England (PHE) | 15/11/2022 |
| England Incidence | 1999-2015 | Office for National Statistics (ONS) | 02/12/2022 |
| England Incidence | 1998,2003,2007 | ONS | 24/02/2023 |
| England Incidence | 1995-1997 | ONS | 03/01/2025 |
| England Incidence | 1971-1994 | ONS | 02/12/2022 |
| Scotland Incidence | 1998-2022 | Public Health Scotland (PHS) | 23/07/2025 |
| Wales Incidence | 2002-2022 | Public Health Wales (PHW) | 23/07/2025 |
| Wales Incidence | 1992-2001 | NHS in Wales/ Welsh Cancer Intelligence and Surveillance Unit | 02/12/2022 |
| Northern Irish Incidence | 1998-2022 | Queens University Belfast (QUB) | 23/07/2025 |
| England and Wales Survival | 1986-2001 | Survival Analysis articles, British Journal of Cancer, 2008 | 18/11/2022 |
| England Survival | 2001-2014 | ONS | 28/01/2025 |
| England Survival | 2001-2016 | PHE | 02/12/2022 |
| England Survival | 2002-2017 | PHE | 27/03/2023 |
| England Survival | 2003-2007 | | 29/11/2022 |

| | | | |
|---|---|---|---|
| England Survival | 2004-2008 | ONS/London School of Hygiene & Tropical Medicine (LSHTM) | 29/11/2022 |
| England Survival | 2005-2009 | ONS | 29/11/2022 |
| England Survival | 2006-2010 | ONS | 24/01/2025 |
| England Survival | 2009-2013 | ONS | 29/11/2022 |
| England Survival | 2011-2015 | ONS | 27/01/2025 |
| England Survival | 2012-2016 | ONS | 27/01/2025 |
| England Survival | 2013-2017 | ONS | 29/11/2022 |
| England Survival | 2016-2020 | NDRS | 18/05/2023 |
| England Survival (Children) | 2001-2017 | ONS | 29/11/2022 |
| Scottish Survival | 1993-2017 | PHS | 29/11/2022 |
| Wales Survival | 2002-2019 | PHW | 24/02/2023 |
| England Cancer Prevalence | 2021 | NDRS | 03/12/2024 |
| Scottish Cancer Prevalence | 2019 | PHS | 25/07/2024 |
| Population Data | 1970-2024 | ONS | 20/09/2025 |
| Home Nation Population Data | 1970-2024 | ONS | 23/01/2026 |
| Population Projections | 2025-2045 | ONS | 21/09/2025 |
| General Population Mortality | 1980-2020 | ONS | 25/08/2025 |
| ICD9 to ICD10 Map | n/a | cancer.gov | 06/01/2025 |