Get Data Out cancer survival analysis SOP.

This analysis follows the UK and Ireland Association of Cancer Registries ratified Standard Operating Procedure [Guidelines on Population Based Cancer Survival Analysis](http://www.ukiacr.org/publication/guidelines-population-based-cancer-survival-analysis); these are the same guidelines that are used to produce the [National Statistics on cancer survival](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinengland/nationalestimatesforpatientsfollowedupto2017) in England. The official 2016 registration extract is the basis for the data analysed here and is supplemented by vital status updates obtained from NHS Digital.

The survival estimates have been calculated using the [Pohar-Perme net survival estimator](http://onlinelibrary.wiley.com/doi/10.1111/j.1541-0420.2011.01640.x/abstract), as implemented by [Isabelle Clerc-Urmès, Michel Grzebyk and Guy Hédelin’s stns programme in Stata](http://www.stata-journal.com/article.html?article=st0326). Net survival is estimated with an excess hazard model in which the all-cause mortality is modelled as the sum of the excess (cancer-related) mortality hazard and the expected (or background) mortality.

Background mortality is derived from population life tables. The life tables used are supplied by the [Office for National Statistics](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalsmoothedlifetables). English life tables for 2014-2016 are used for 2015 to 2017. In 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.

In the overall adult survival analyses, for age groups where the estimates do not meet the following quality criteria, the result is suppressed for that particular age group of the specific cancer site:

* A minimum of ten patients should be alive at the duration(s) of survival being estimated (no cohorts failed this criterion)
* At least two patients should die in the years before or after the duration(s) being estimated (cohorts failing this criterion are denoted “.f” in the results)
* The standard error of the survival estimates should be lower than 20% (cohorts failing this criterion are denoted “.g” in the results).
* The level of the survival estimates should not increase with duration. E.g. the survival estimated at 12-months following diagnosis should be lower than the survival estimated at 3-months following diagnosis (cohorts failing this criterion are denoted “.h” in the results)

We implement survival analysis data quality [criteria](http://www.sciencedirect.com/science/article/pii/S1877782114000356)  to identify the patients that are eligible to be included in the analysis. Other decisions applied include:

* where a patient dies on the date of diagnosis but is not a DCO registration, then these patients should be included in the survival analyses but should have 1 day added to the recorded date of death to accommodate limits in [Stata’s stset command](http://www.stata.com/manuals13/ststset.pdf)
* when two or more tumours of the same type are diagnosed on the same day for a patient, the one with the worst prognosis is chosen for inclusion