**Extended explanation of models used in Henson et al 2019**

22 May 2024

> Henson K, Brock R, Charnock K, Wickramasinghe B, Will O, Pitman A (2019) Risk of suicide after a cancer diagnosis in England: a population-based study. *JAMA Psychiatry*; 76(1):51-60 doi: 10.1001/jamapsychiatry.2018.3181
> https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2714596

*Summary prepared by Dr Alexandra Pitman, UCL Division of Psychiatry, Dr Justin Yang, UCL Division of Psychiatry, and Dr Katherine Thackray (née Henson), NHS England.*

The study used cancer registration data for the whole of England linked to Office for National Statistics (ONS) mortality data to identify all individuals diagnosed with malignant cancer (excluding non-melanoma skin cancer) in England during 1995-2015. We included all individuals who were aged 18-99 years at diagnosis. ONS mortality data provided the cause of death information to identify suicides.

Our **outcome** was all open verdicts and verdicts of suicide. An open verdict is a legal decision that records a death but does not state its cause. There is substantial evidence to show that suicides often get coded as open verdicts, and so these were included as per routine practice in official statistics. Suicide may be under-reported as a cause of death because legal requirements for identifying suicide are less stringent than health professionals' criteria, and coroners may avoid recording a suicide verdict in case it adds to a family's distress. This under-reporting may impact cancer patients more than the general population, which suggests that some suicide deaths in patients with cancer may be classified as an open verdict and some may be classified as a death by natural causes. It is therefore important to recognise that findings from any study of suicide risk in patients with cancer may under-estimate the risk of suicide.

To understand the risk of suicide among the cancer population, we performed a comparison of the number of deaths among patients with cancer to the number of deaths expected from the general population of the same age, sex and during the same year.

We reported this comparison as two measures:

the **Standardised Mortality Ratio** (SMR) as a relative measure of suicide risk between individuals with cancer and the general population, calculated as the observed number divided by the expected

the **Absolute Excess Risk** (AER) as an absolute measure of the number of additional cancer patients who died from suicide (or who did not die from suicide) as compared to the general population by person-time at risk, calculated as the observed minus the expected divided by the person-years at risk.

We also calculated **relative risks** (ratio of SMRs) and **relative excess risks** (ratio of AERs) using multivariate models.

**Cohort definition**

As we were looking at the outcome following a diagnosis date, we used **time to event data** and cohort techniques.

Patients entered the cohort on the date of cancer diagnosis and were followed up until date of death, date when lost to follow-up, or August 31, 2017.

**Models**

Poisson modelling allows you to model count data taking into account the underlying population to produce a rate. This allows for comparisons across subgroups.

Multivariable Poisson regression models were used to evaluate the simultaneous effect of the factors of interest. All factors were fitted in the regression model with no interaction terms. These factors were sex, cancer type, deprivation, race/ethnicity, age at cancer diagnosis, year of diagnosis, attained age, and follow-up period (or attained age).

We defined the covariates of deprivation and race/ethnicity as follows:

- **Deprivation** was captured using the income deprivation domain of the Index of Multiple Deprivation (IMD). The IMD is a relative measure of deprivation used in England (see https://assets.publishing.service.gov.uk/media/5a7f0e5ded915d74e33f410b/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf ). The income deprivation domain of the IMD measures the proportion of the population in an area experiencing deprivation relating to low income. The definition of low income used includes both those people that are out-of-work, and those that are in work but who have low earnings (based on an established earnings threshold).
- **Ethnicity** was self-reported, as recorded in hospital patient administration systems, and submitted to the cancer registry. If a patient's ethnicity was unknown, it was supplemented by information on self-reported ethnicity as recorded in the Hospital Episode Statistics dataset. As ethnicity may be recorded differently during each hospital admission, the modal value was used in cases of discrepancy, preferring the most recent ethnicity documented in the event of a tied modal value.

The SMR was defined as the number of deaths by suicide observed in the cancer population / number of expected deaths by suicide.

The AER was defined as the ([number observed − number expected] / [person-years at risk]) × 10 000. These measures estimate the proportional increase in the suicide rate compared with rates in the general population and the absolute excess suicide rate compared with the general population, respectively.

Population-based expected deaths were derived from age (5-year groups), sex, and calendar-year (1-year groups) specific death rates for England.

Poisson regression models were used to test for heterogeneity and to evaluate the effect of all confounders simultaneously. Tests for heterogeneity were performed using likelihood ratio tests based on Poisson regression models. This test compared the deviance of a model including the factor of interest to the deviance of a model without the factor of interest.

Statistical significance was defined as 2-sided with significance level set at $P < .05$.

**Sub-group analyses**

During the exploration of the cohort characteristics, it is important to consider the sub-groups for follow-up period. You should balance the clinical context and evidence base against your cohort size and the need for large enough groups to provide the statistical power to produce reliable estimates.

We explored the effect of**:** (although only the last of these was described explicitly as a sub-group analysis)

- sex
- years since cancer diagnosis

- attained age
- stage at diagnosis

**Sensitivity analyses**

During the development of the cohort, we investigated the impact of using all parts of the death certificate to identify relevant deaths, and also the inclusion of open verdicts. For the study, we followed ONS recommended coding, but it is important to understand the impact of coding decisions and the resulting ascertainment of deaths. It may be appropriate to use a tighter definition in the sensitivity analysis.

If there were any changes in registration or ascertainment at any point over follow-up (for example a change in diagnostic systems or coding systems), subset analyses could be performed on a subset of the sample covering diagnoses made over a specific period to assess if the change has an impact on the overall estimates.

In our analysis, we tested the effect of excluding open verdicts to see if any issues with coding these had introduced ascertainment bias. Other sensitivity analyses that analysts might consider in order to address any local sources of bias include:

- using a stricter/broader definition of suicide if there are concerns about death certificate coding of suicide;
- combining tumour types where there is uncertainty over precise tumour site.