International Journal of Population Data Science

Journal Website: www.ijpds.org

Survival to Diabetes by age, gender and BMI category

Nock, Kelly^{1*}, Bailey, Kerry¹, Barker, Craig², and Thomas, Helen²

¹We Predict

²Abertawe Bro Morgannwg University Health Board

Aim

To carry out a 10 year survival analysis to provide insight to NHS partners regarding the time to Type-2 Diabetes across the population of ABM HB and by a variety of key criteria including BMI, gender, age and deprivation.

Approach

Survival time was taken from the window start date (current data date – 10 years) to patient exit from the study (Death, Type-2 Diabetes diagnosis or window end date). Survival output was generated using the Kaplan-Meier estimate as part of the survival package (Therneau T. M., 2015) in R, with Type-2 Diabetes diagnosis as the event variable with the following control and stratification variables: Gender, Age Group (at window start), Average BMI Category and Deprivation Quintile.

To identify Diabetic and I particular type-2 patients we developed dictionaries using diagnosis codes from both ICD-10 and READ. These dictionaries were then mapped against the GP, Inpatient and Outpatient datasets to capture a diagnosis of Diabetes as early as possible. The logic to determine whether patients are Diabetic involved the use of 277 READ and 56 ICD-10 codes, incorporating and expanding upon the QOF standard rule-set.

To identify as many BMIs as possible three methods were used firstly using absolute BMI measurements. Secondly, using READ codes for pre-binned BMI categories. Thirdly, by combining separate height and weight measurements. The latter method in particular introduces data quality problems, both with erroneous values and in cases where different GPs use different units of measurement (e.g. metres, centimetre, feet, stones). To use these measurements in a reliable manner, we developed complex conditional statements to extract valid measurement records and store them uniformly, such that we can use them easily in a variety of data queries to classify patients by weight category.

Results

Cumulative probability (calculated from inverse 'survival time') to diabetes is greater in obese and more elderly. There was not the expected impact of deprivation. Representation of survival curves were difficult for lay people and some other stakeholders to interpret so stacked bar charts of cumulative probabilities were visualised. Stakeholders valued being able to interact with the different aggregated visualisations to display different aspects of the results. The results provided valuable insights for NHS partners and informed predictive models.

*Corresponding Author:

Email Address: james@fn.com.au (J. Farrow)