

Parallel time-series analysis of HbA1c and Systolic BP using a recurrent neural network (RNN) stratifies for 1-year mortality in Type 2 Diabetes independent of age and parameter variability

Background / Aims

Our aim was to investigate clinical associations of HbA1c and systolic blood pressure (SBP) measurements recorded over time. The analysis was performed on a regional subset of the Scottish national SCI-Diabetes database, containing clinical information of all people diagnosed with diabetes within the Greater Glasgow and Clyde health board.

The majority of previous analysis to date has focused on investigating association between summary HbA1c and SBP data (usually comprising measures of average and spread eg mean and standard deviation) and clinical outcomes. We hypothesised that in addition to the numerical values recorded, the sequence of measured values would have an association with clinical outcome, and that interaction between parameters over time may add additional information. Mortality was chosen as it is well recorded within our dataset, and we and many others have previously demonstrated associations between summary measures of the parameters of interest and mortality (1).

Recurrent neural networks (RNNs) are a class of neural network that use their internal memory to process arbitrary sequences of inputs. Long Short Term Memory (LSTM) networks are a variant of the RNN capable of learning long-term dependencies. A multi-dimensional RNN replaces the single recurrent connection found in standard RNNs with as many recurrent connections as there are dimensions in the data, allowing understanding of interaction between dimensions, as well as over time.

Methods

HbA1c, SBP (analysis parameters) and mortality data for all individuals with Type 2 Diabetes in our health board were identified from 2008 – 2013. Inclusion in the analysis required a recorded diagnosis prior to 2008 (ie the start of the run-in period), and survival to 2013 (the end of the run-in period). Sequences of all parameters were generated using 2-month time bins. Missing values were imputed by carrying last value forward. Age and coefficient of variation (CV) of all parameters during the run-in period were calculated.

A recurrent neural network (with 2 LSTM (Long Short-Term Memory) layers) was trained on 80% of parameter time-series data, taking mortality status at the end of a 1-year follow up period as the dependent variable. A survival analysis was performed on individuals within the test-set (20%) - comparing those with a predicted probability of death (within 1 year) above the median probability value vs those below, with age, HbA1c CV, and SBP CV as covariables.

Results

33456 individuals with Type 2 Diabetes diagnosed prior to 2008 were included in the analysis. The RNN training set comprised 26765 individuals, with a test set size of 6691 individuals. Age and summary characteristics of the parameters during the run-in period are shown in Table 1.

279 deaths occurred within 1 year in the test set, with 248 deaths occurring within the group classified as high risk by the RNN.

The area under the curve (AUC) of a receiver operating characteristic (ROC) curve for predicting mortality was 0.82 (Figure 1).

On survival analysis analysed using a cox proportional hazards model, HR for mortality for those with a predicted probability of death above median value was 2.41 (1.77 – 3.28) $p < 0.0001$, when age HbA1c CV and SBP CV included as covariates (Figure 2).

	all subjects	low probability of mortality	high probability of mortality
age (years)	63.7 (54.1 – 72.2)	60.3 (51.3 – 69.5)	66.6 (57.9 – 74.0)
mean HbA1c (mmol/mol)	53.1 (44.8 – 63.8)	59.1 (49.9 – 71.9)	48.8 (41.7 – 56.1)
HbA1c CV	0.28 (0.19 – 0.39)	0.29 (0.20 – 0.40)	0.27 (0.19 – 0.39)
mean SBP (mmHg)	126.6 (116.9- 135.5)	121.0 (112.4 – 129.0)	132.0 (124.0 – 139.8)
SBP CV	0.20 (0.09 – 0.34)	0.21 (0.10 – 0.34)	0.19 (0.09 – 0.30)

Table 1. Parameter characteristics and age of individuals within the test set. Low and high probability of mortality determined by RNN output. Values shown as median (interquartile range). n = 6691

Summary and Conclusions

Multi-parameter time-series analysis allows information to be captured from both the sequence of values and their interaction over time, as well as from the numerical values themselves. In this analysis, the output from the multidimensional RNN is an independent predictor of mortality.

It is interesting to note that the RNN-determined risk of mortality associates with lower mean/CV values of HbA1c and with lower SBP CV. In the presented analysis, a cut-off at the median probability of mortality was used to determine high/low risk. If this threshold is raised, the model strongly favours CV over average values of these parameters (data on file). Overall this analysis suggests that a time-series approach to analysis may provide useful additional information when stratifying for adverse outcome.

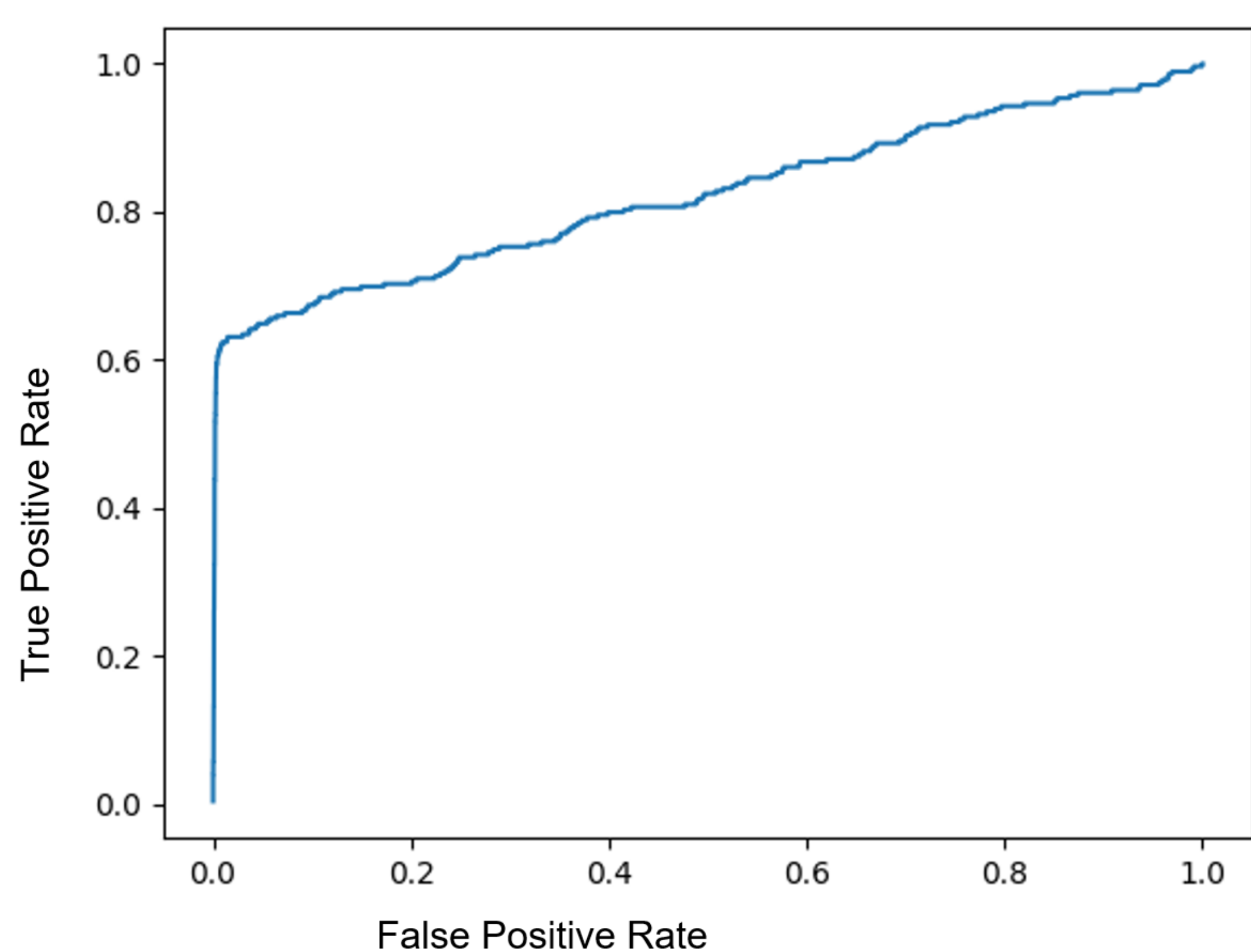


Figure 1. Receiver Operator Characteristic Curve illustrates the diagnostic ability of the multidimensional RNN to predict mortality at 1 year post run-in period, as the discrimination threshold is varied

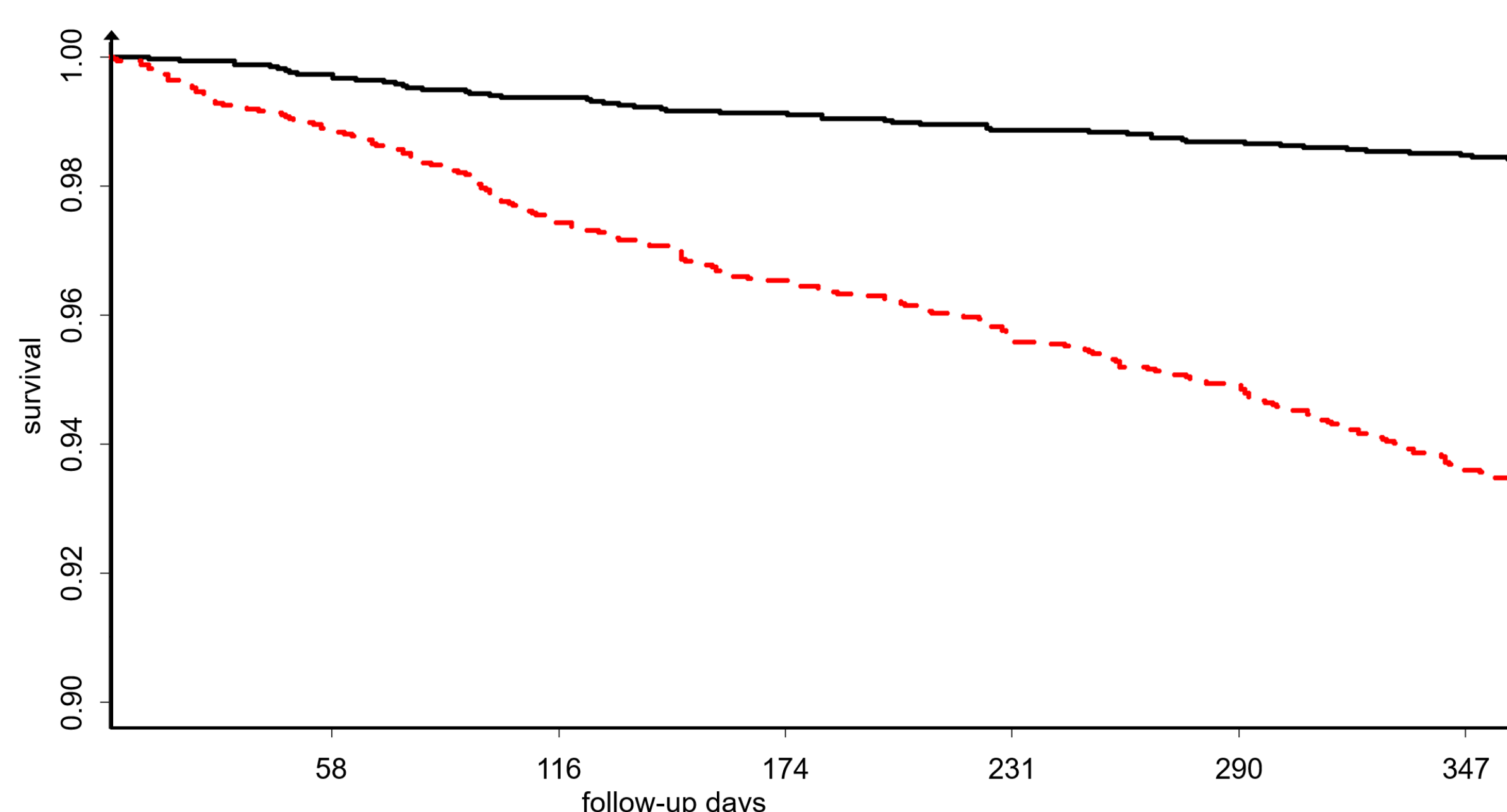


Figure 2. Survival analysis. 1 year follow-up. n = 6691
HR probability of death > median probability 2.41 (1.77 – 3.28) $p < 0.0001$
HR age: 1.09 (1.08 – 1.10), $p < 0.0001$
HR HbA1c CV: 0.85 (0.74 – 0.98), $p = 0.02$
HR SBP CV: 3.79 (3.33 – 4.3), $p < 0.0001$

References

- (1) Wightman SS, Sainsbury CAR, Jones GC. Visit-to-visit HbA1c variability and systolic blood pressure (SBP) variability are significantly and additively associated with mortality in individuals with type 1 diabetes: An observational study. *Diabetes Obes Metab.* 2018;1–4. <https://doi.org/10.1111/dom.13193>