Time-series analysis of HbA1c using Recurrent Neural Networks stratifies for 1-year mortality in Type 1 and Type 2 Diabetes, independent of age and variability

CAR Sainsbury, J Osmanska, GC Jones

glucose.ai Research Group, Department of Diabetes, Gartnavel General Hospital, Glasgow, Scotland

http://glucose.ai
@csainsbury

Background / Aims

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

The majority of analysis to date has focused on investigating association between summary HbA1c data (usually comprising measures of average and spread) 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. Mortality is well recorded within our dataset, and we and many others have previously demonstrated an association between summary measures of HbA1c and mortality (1).

Recurrent neural networks (RNNs) are a class of neural network that use their internal memory to process arbitrary sequences of inputs (figure 1). Long Short Term Memory (LSTM) networks are a variant of the RNN capable of learning long-term dependencies.

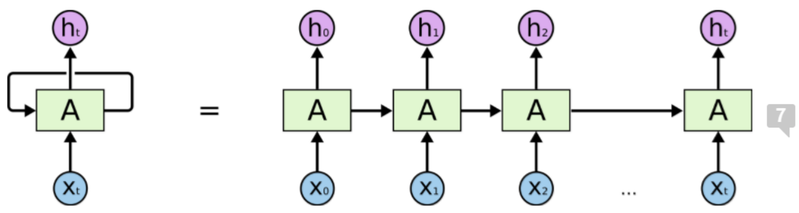

Figure 1. An unrolled recurrent neural network (2)

Methods

HbA1c and mortality data for individuals with T1DM or T2DM in our health board were identified from 2008-2013. T1DM and T2DM were analysed separately. Inclusion in the analysis required a recorded diagnosis prior to 2008 (ie the start of the run-in period), and survival to the 2013 (the end of the run-in period). An HbA1c sequence was generated using 2-month time bins. Missing values were imputed using a last value carried forward technique. Age and a measure of HbA1c variability - coefficient of variation (CV) - were calculated.

A recurrent neural network (with an embedding layer, and a 128-node LSTM (Long Short-Term Memory) layer) was trained on 80% of HbA1c 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 generated by the RNN (within 1 year) above the median probability value vs those with a predicted probability below the median value, with age and HbA1c CV as covariables.

Results

Type 1 Diabetes

6620 individuals were included in the analysis. The RNN training set comprised 5296 individuals meeting the criteria for inclusion, with 1324 individuals comprising the test set. Age and glycaemic (HbA1c) characteristics of the test set shown in table 1. There were 18 deaths within 1 year, with 15 deaths occurring within the group classified as high probability by the RNN.

On survival analysis HR for mortality for those with a predicted probability of death above median was 4.01 (95% CI 1.12-14.33) p=0.03, when age and HbA1c CV included as covariables (Figure 2).

	all subjects	low probability of mortality	high probability of mortality
age (years)	38.6 (25.0 – 86.7)	39.8 (25.9 – 51.3)	36.9 (24.5 – 49.7)
median HbA1c (mmol/mol)	71.0 (63.0 – 83.5)	70.0 (63.0 – 81.4)	72.5 (63.0 – 85.0)
HbA1c CV	0.26 (0.15 - 0.40)	0.23 (0.13 – 0.34)	0.28 (0.18 – 0.46)

Table 1. Glycaemic characteristics and age of individuals within the Type 1 Diabetes test set. Values shown as median (interquartile range). n = 1324

Type 2 Diabetes

34060 individuals were included in the analysis. The RNN training set comprised 27244 individuals meeting the criteria for inclusion, with a test set size of 6816 individuals. Age and glycaemic (HbA1c) characteristics of the test set shown in table 2. There were 318 deaths within 1 year, with 222 deaths occurring within the group classified as high risk by the RNN.

On survival analysis, HR for mortality for those with a predicted probability of death above median was 1.53 (95% CI 1.19-1.96) p=<0.001, when age and HbA1c CV included as covariables (Figure 3).

	all subjects	low probability of mortality	high probability of mortality
age (years)	63.4 (54.0 – 71.9)	61.2 (52.6 – 70.3)	65.0 (56.2 – 73.1)
median HbA1c (mmol/mol)	56.0 (49.0 – 67.5)	62.0 (54.5 – 75.0)	51.0 (49.0 – 58.0)
HbA1c CV	0.28 (0.20 – 0.40)	0.28 (0.19 – 0.36)	0.29 (0.20 – 0.42)

Table 2. Glycaemic characteristics and age of individuals within the Type 2 Diabetes test set. Values shown as median (interquartile range). n = 6816

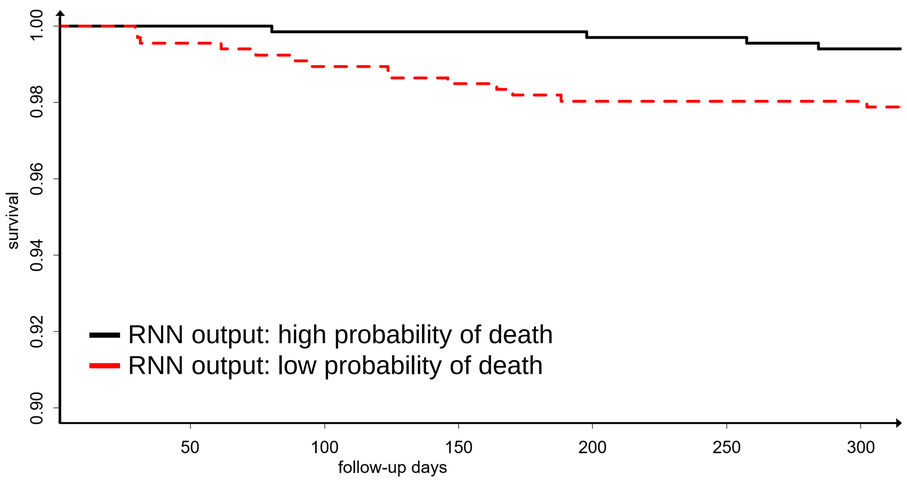

Figure 2. Survival analysis T1DM. 1 year follow-up. n = 1324
HR probability of death > median probability 4.01 (1.12 – 14.33), p=0.03
HR age: 1.08 (1.05 – 1.11), p<0.001
HR HbA1c CV: 1.88 (1.43 – 2.47), p<0.001

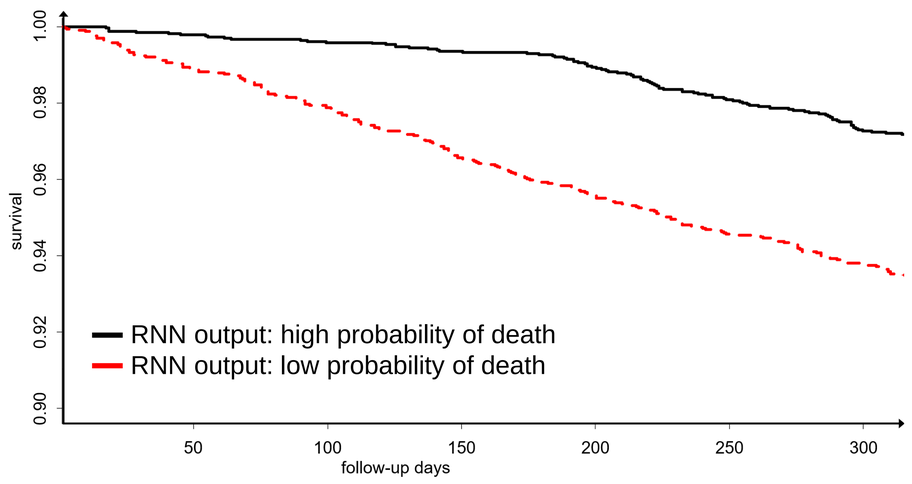

Figure 3. Survival analysis T2DM. 1 year follow-up. n = 6816
HR probability of death > median probability 1.53 (1.19 – 1.96), p<0.001
HR age: 1.09 (1.07 – 1.10), p<0.001
HR HbA1c CV: 2.55 (2.31 – 2.80), p<0.001

Summary and Conclusions

Time-series analysis allows information to be captured from the sequence of numerical values, as well as from the values themselves. This analysis demonstrates that this may be able to add additional information over summary metrics such as CV when stratifying for outcomes of interest.

References

(1) HbA1c variability is associated with increased mortality and earlier hospital admission in people with Type 1 diabetes. G Walker, S Cunningham, C Sainsbury, G Jones. Diabet. Med. 34, 1541–1545 (2017)

(2) https://colah.github.io/posts/2015-08-Understanding-LSTMs/