**Background/Methods**

Traditionally, CVD risk equations have been developed using Cox proportional-hazards models. The Cox model is highly interpretable and can be used to investigate how the hazard changes in response to explanatory covariates. However, the Cox model is limited in its ability to account for large numbers of explanatory covariates (such as individual diagnostic codes or medications), complex event sequences and interactions. Recent work showed that the linear predictor of the Cox proportional-hazards model can be replaced by a deep neural network, leading to improved predictive accuracy in terms of concordance index [1]. A popular component of many contemporary neural network architectures are attention mechanisms, which allow the network to focus on specific subsets of relevant inputs and can be used to formulate interpretable machine learning models [2]. By applying a Cox model combined with an attention-based neural network to very large-linked administrative datasets we are going to compute hazard ratios associated with individual codes (e.g. diagnoses, medications), thereby identifying novel risk predictors that will inform future risk prediction equations. Models will be trained using a recently proposed cost function which compares individuals experiencing an event to randomly sampled individuals at risk, an approach which has been shown to scale to large datasets and can readily be applied to non-proportional extensions of the Cox model [3]. The feasibility of this novel approach has already been demonstrated by preliminary work using the New Zealand VARIANZ data [4]; in the Australian context these methods will be applied to the National Integrated Health Services Information dataset, which captures whole-of-population longitudinal person-level information on hospitalisations, outpatient visits, Medicare claims, pharmaceutical dispensing and mortality.

**Feasibility**

We applied a Cox model combined with an attention-based neural network to estimate the risk of CVD death or hospitalization within 5 years for 1.7 million New Zealanders, aged 30-74 years and without previous hospitalizations for CVD or heart failure [4]. The employed VARIANZ dataset comprises demographic variables, hospitalizations, pharmaceutical dispensings and mortality. Table 1 shows the hazard ratios between an individual with a specific primary diagnostic code and a (perfectly healthy) individual with the same demographic variables but no recorded diagnostic codes or pharmaceutical dispensings in the 5 years prior to risk assessment. Confidence intervals for the estimated hazard ratios were computed using an ensemble of 100 deep neural networks [5].

Table 1. Largest hazard ratios for primary diagnostic codes (mean, [95% confidence interval]).

|  |  |  |
| --- | --- | --- |
| ICD-10 | DESCRIPTION | Hazard Ratio |
| I10 | Essential (primary) hypertension | 22.1 [16.5, 29.5] |
| Z490 | Preparatory care for dialysis | 14.1 [8.9, 22.3] |
| M8697 | Unspecified osteomyelitis, ankle and foot | 10.8 [8.0, 14.5] |
| M329 | Systemic lupus erythematosus, unspecified | 10.7 [7.8, 14.8] |
| E1173 | Type 2 diabetes mellitus with foot ulcer due to multiple causes | 9.3 [5.4, 15.7] |
| J440 | Chronic obstructive pulmonary disease with acute lower respiratory infection | 8.6 [6.3, 11.6] |
| O100 | Pre-existing essential hypertension complicating pregnancy, childbirth and the puerperium | 8.5 [7.0, 10.3] |
| L97 | Ulcer of lower limb, not elsewhere classified | 7.9 [6.2, 10.0] |
| C343 | Malignant neoplasm of lower lobe, bronchus or lung | 7.8 [6.1, 9.8] |
| I671 | Cerebral aneurysm, nonruptured | 6.9 [5.3, 8.8] |

**References**

1. Katzman, J.L., et al., *DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network.* 2018. **18**(1): p. 24.

2. Vaswani, A., et al. *Attention is all you need*. in *Advances in neural information processing systems*. 2017.

3. Kvamme, H., Ø. Borgan, and I.J.J.o.M.L.R. Scheel, *Time-to-event prediction with neural networks and Cox regression.* 2019. **20**(129): p. 1-30.

4. Mehta, S., et al., *Development and validation of alternative cardiovascular risk prediction equations for population health planning: a routine health data linkage study of 1.7 million New Zealanders.* 2018. **47**(5): p. 1571-1584.

5. Lakshminarayanan, B., A. Pritzel, and C. Blundell. *Simple and scalable predictive uncertainty estimation using deep ensembles*. in *Advances in neural information processing systems*. 2017.