Predicting Cardiovascular Disease Across Populations Using Deep Learning

Bruno Batinica[1], Sebastiano Barbieri[2], Jingyuan Liang[1], Katrina Poppe[1], Suneela Mehta[3], Rod Jackson[1]

1. Section of Epidemiology and Biostatistics, University of Auckland, Auckland, New Zealand, 2. Centre for Big Data Research in Health, University of New South Wales, Sydney, NSW, Australia, 3. Planning, Funding and Outcomes Team, Waitematā and Auckland District Health Boards, Auckland, New Zealand

Background

- Cardiovascular disease (CVD) is the leading single cause of **premature mortality** worldwide [1].
- **Risk prediction models** trained on **longitudinal research cohorts** are a well-established tool to aid clinical decision-making regarding CVD prevention [2].
- Models trained on **administrative datasets** are a low-cost alternative that can predict risk **across entire populations**, including in **resource-deprived areas** [3].
- The predictive power of **traditional models** trained on **administrative datasets** is restricted by the limited number of **standard predictors** available in these datasets.
- Flexible **deep learning** models may be able to offset this limitation by utilising **additional health information** found in large administrative datasets [4].

Aims

To **develop** novel deep learning-based CVD risk prediction models from administrative datasets and **compare** them to traditional statistical models.

Methods

Data sources: Ministry of Health† and CareConnect‡ datasets containing:

- **Demographic information†**
- **Hospital diagnoses & procedures†** (ICD-10)
- **Pharmaceutical dispensing†**
- **Community and hospital laboratory test results‡** performed in Auckland and Northland

Figure 1. Study catchment area

Cohort: 768,629 Individuals aged 30-74 without prior CVD from Auckland & Northland followed for 5 years **[Fig 1.]**.

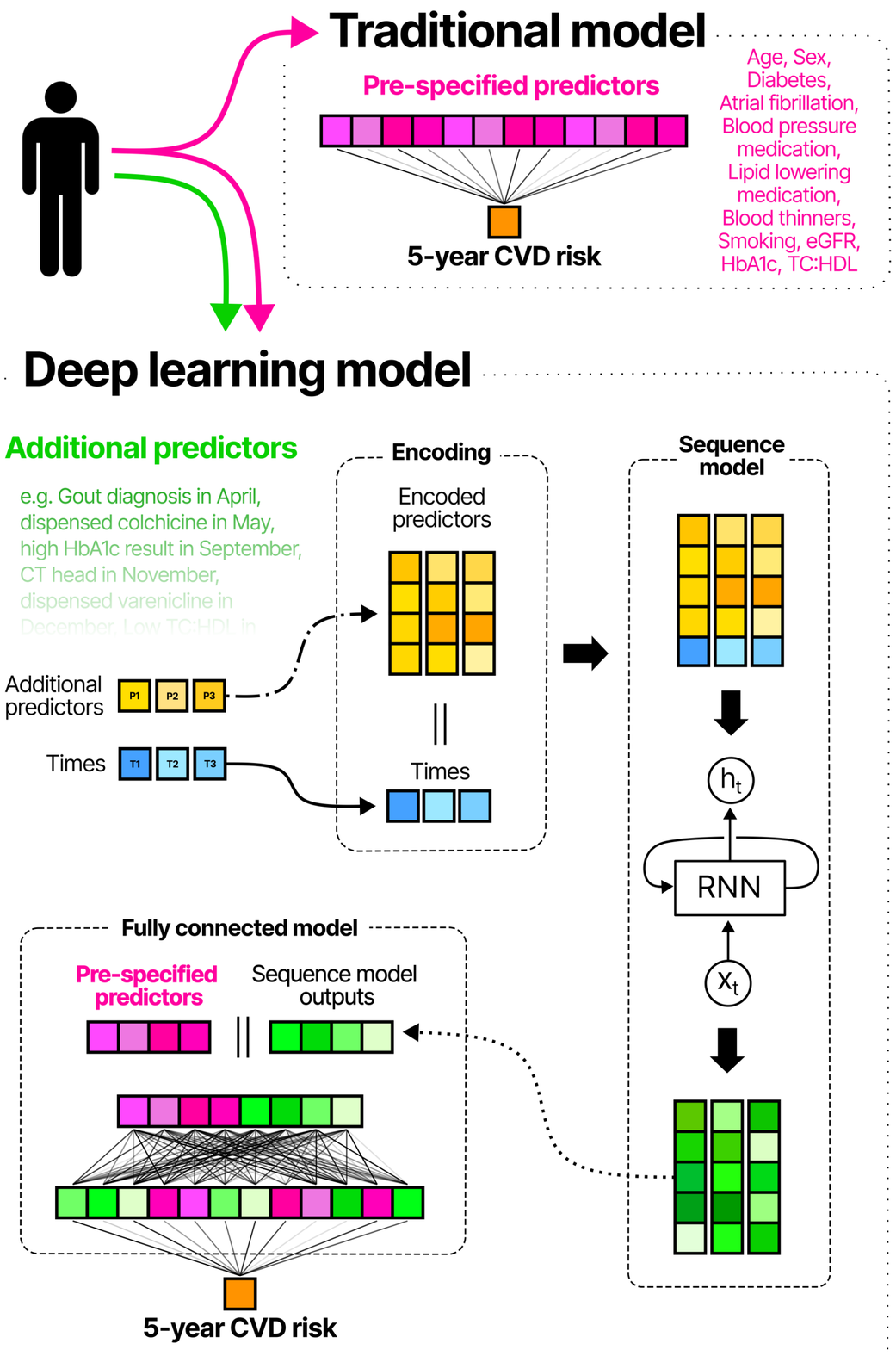

Figure 2. Model architectures

Models: Sex-specific deep learning models were trained and compared against traditional Cox-proportional hazard models **[Fig 2.]**.

Results

Interpretability

- The deep learning models identified novel **causal** and **proxy** variables and quantified their effect on CVD risk **[Fig 3.]**.

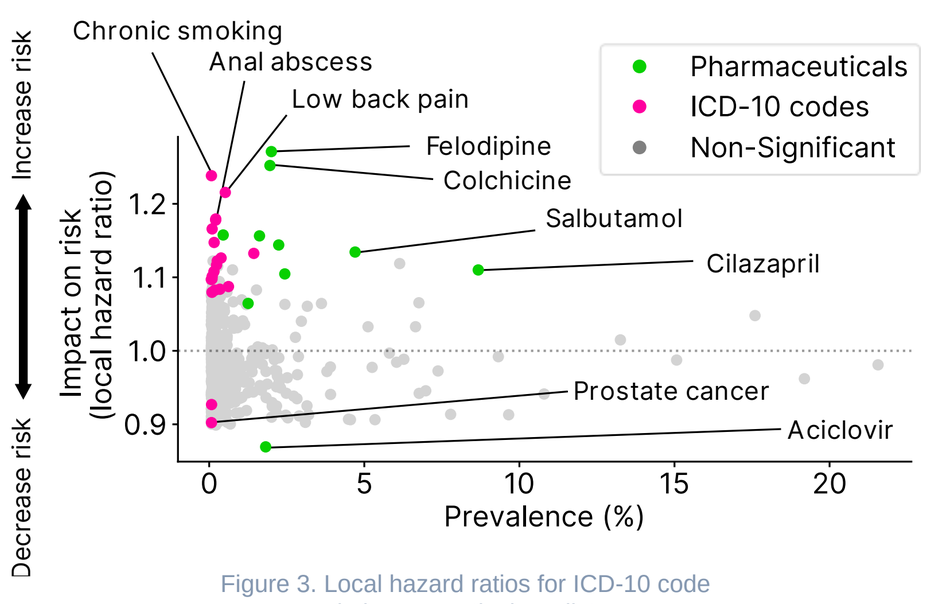

Figure 3. Local hazard ratios for ICD-10 code and pharmaceutical predictors

Performance

- Deep learning models demonstrated better **calibration** and **discrimination** in the whole cohort and all assessed subpopulations **[Fig 4.]**.

Figure 4. Sex-specific calibration plots

- Deep learning models had improved **R^2 values** ($P < 0.01$), **C statistics** ($p < 0.001$) and **Integrated Brier Scores** ($P < 0.01$).

Conclusion

When predicting CVD risk using administrative data, deep learning models **outperform** traditional statistical models. The models presented here are readily **available for use** by policymakers and researchers in the Auckland/Northland region. Similar models could be developed in many countries if simple **data linkage** efforts were undertaken.

1. Labarthe, Darwin R. "Epidemiology and Prevention of Cardiovascular Diseases: A Global Challenge: A Global Challenge." (2010). 2. Damen, Johanna AAG, et al. "Prediction models for cardiovascular disease risk in the general population: systematic review." bmj 353 (2016). 3. Mehta, Suneela, 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." International Journal of Epidemiology 47.5 (2018): 1571-1584. 4. Barbieri, Sebastiano, et al. "Predicting cardiovascular risk from national administrative databases using a combined survival analysis and deep learning approach." International Journal of Epidemiology 51.3 (2022): 931-944.