Understanding Progressive CKD from Primary Care Records – a Study Protocol

Norman Poh[1], Simon Bull[1], Santosh Tirunagari[1], John Williams[2], Nicholas Cole[2], Christopher Farmer[3] and Simon de Lusignan[2]

1: Department of Computer Science, University of Surrey

2: Department of Clinical and Experimental Medicine, University of Surrey

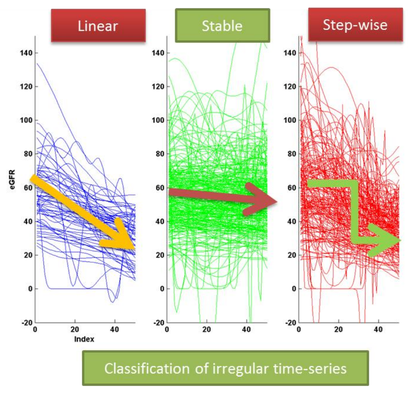
3: East Kent Hospitals University NHS Foundation Trust

Contact: N.Poh@surrey.ac.uk

Introduction

Background: One of the key challenges in managing CKD patients is to identify those who are progressive (worsening eGFR) from those who are non-progressive or even those with reversible CKD. There have been several independent studies attempting at characterising progression:

- Manually classify trends as stable, linear progressive and step-wise progressive (Kilbride, 2015)
- Measure variability in eGFR, i.e., standard deviations in eGFR, adjusted for eGFR mean (Tseng et al, 2015)
- Classify eGFR trajectory types and measure mean eGFR slope (Xie et al, 2016)



Existing findings

- eGFR standard deviation is smaller when eGFR value is smaller
- Increased eGFR variability is associated with increased risk of death in CKD population with diabetes (Tseng et al, 2015)
- Positive eGFR slope is associated with risk of death (Xie et al, 2016)

Technical challenges

1. How to rate CKD patients in the spectrum of non-progressive to progressive given their longitudinal eGFR measurements?
2. What factors might explain the progression?
3. Routinely collected primary care data present additional challenges due to the existence of several coding schemes.

Methodology

1. Use broken-stick model (see our adjacent poster) to estimate the mean eGFR slope. Rank patients by their slope.
2. Identify a list of variables for a range of known factors, which are related to signs and symptoms (*s*), laboratory measurements (*m*) and treatments (t) for CKD as well as its associated co-morbidities or risk factors (*r*) such as diabetes, cardiovascular disease, anaemia and other kidney-related diseases. For each risk factor, *s*, *t*, *m* and *r* are systematically identified (see another adjacent poster), leading to a range of plausible phenotype variables.

Two protocols

2 protocols possible on 10 years' worth of data from Royal College of General Practitioners (RCGP) data set.

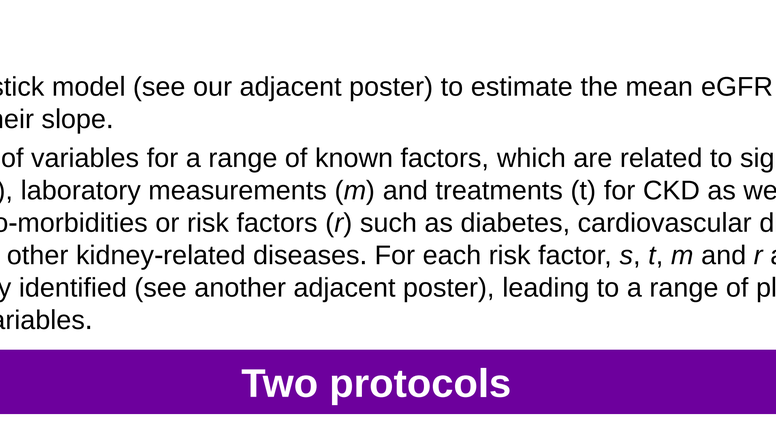
Fixed-length (50-50) protocol

- 1st five years as observation period
- 2nd five years as outcome period
- Outcome is a three-point mace –a composite of cardiovascular death, nonfatal myocardial infarction and nonfatal ischaemic stroke

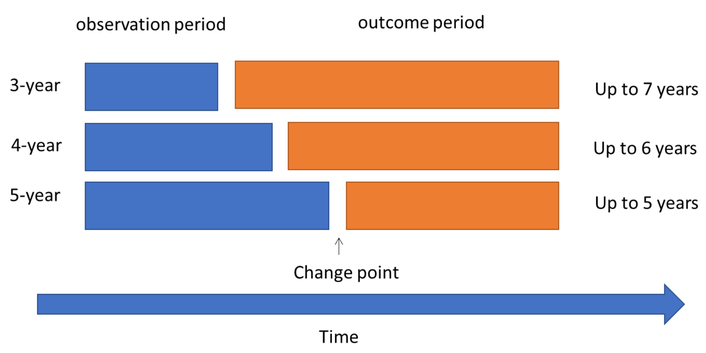
Variable-length protocol

- A special case being the 50-50 protocol
- Requires at least 3-years of observation period; and the rest as outcome period with variable length
- Advantage: More efficient use of data since it does not impose any cut-off of before/after time period

Fixed-length protocol



Variable-length protocol



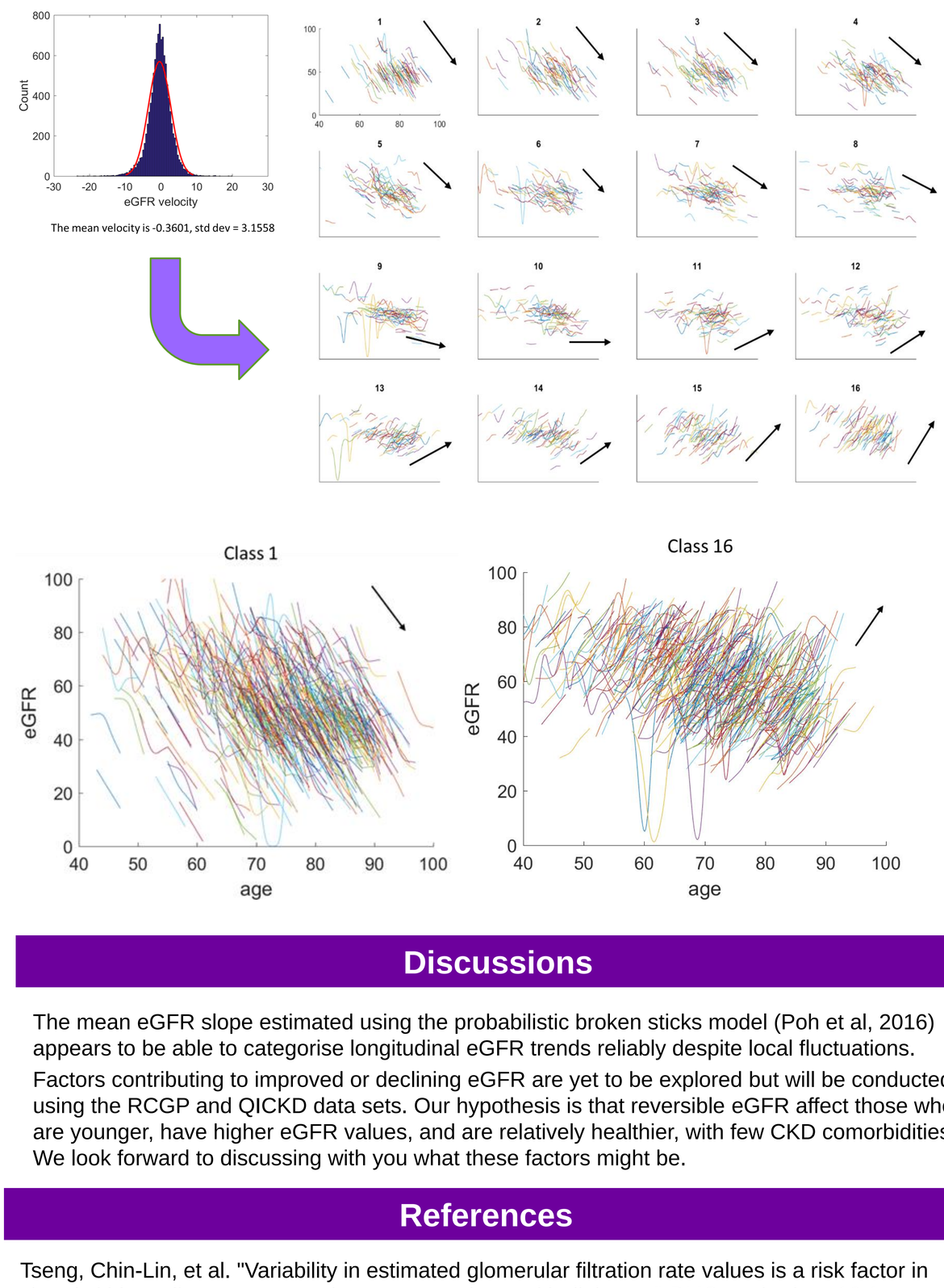
Methods

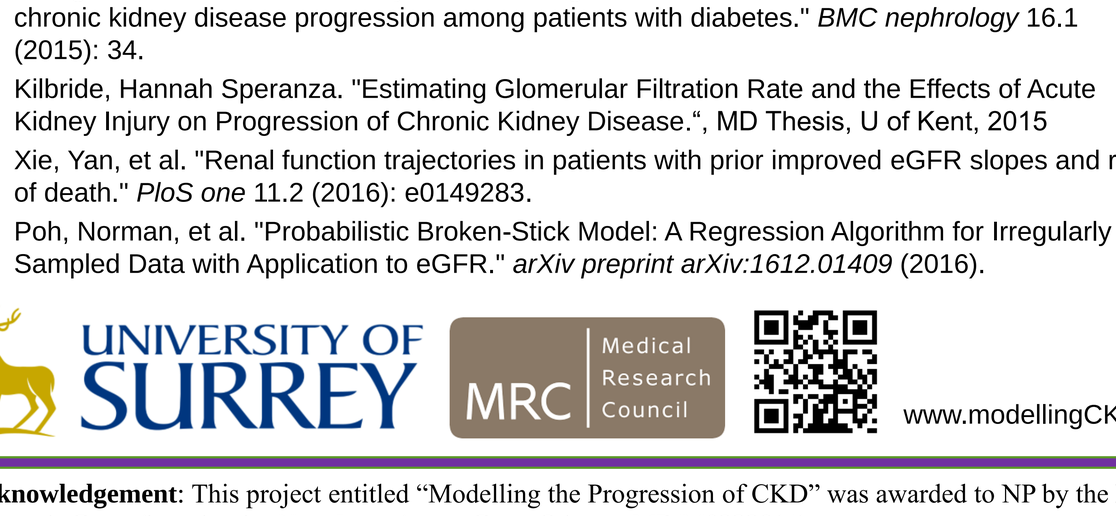
We have developed our own regression algorithm – called the broken-stick model (Poh et al, 2016) – capable of estimating the rate change of eGFR by using a 'Bayesian' sliding window of three years in order to provide a stable estimate of the annual rate change of eGFR, while still being sensitive to underlying genuine patterns. The global eGFR slope (annual rate change) of a patient is defined as the average eGFR slope over the period of analysis.

Patient inclusion/exclusion criteria: All patients with eGFR measurements were included. Patients with one or more acute kidney injuries or hereditary kidney diseases were excluded. Additionally, patients without *consistent* eGFR trends, defined as having a standard deviation of the eGFR slope of 2 units per year, were excluded.

Observation

We conducted a pilot study based on Quality Improvement CKD (QIKCD) data set. The longitudinal observational data was divided into equal groups of approximately 600 patients. In 9 out of 16 groups we found that there was a deterioration in eGFR, one group was equivocal and six groups showed improvement.





Discussions

The mean eGFR slope estimated using the probabilistic broken sticks model (Poh et al, 2016) appears to be able to categorise longitudinal eGFR trends reliably despite local fluctuations.

Factors contributing to improved or declining eGFR are yet to be explored but will be conducted using the RCGP and QICKD data sets. Our hypothesis is that reversible eGFR affect those who are younger, have higher eGFR values, and are relatively healthier, with few CKD comorbidities. We look forward to discussing with you what these factors might be.

References

Tseng, Chin-Lin, et al. "Variability in estimated glomerular filtration rate values is a risk factor in chronic kidney disease progression among patients with diabetes." *BMC nephrology* 16.1 (2015): 34.

Kilbride, Hannah Speranza. "Estimating Glomerular Filtration Rate and the Effects of Acute Kidney Injury on Progression of Chronic Kidney Disease.", MD Thesis, U of Kent, 2015

Xie, Yan, et al. "Renal function trajectories in patients with prior improved eGFR slopes and risk of death." *PloS one* 11.2 (2016): e0149283.

Poh, Norman, et al. "Probabilistic Broken-Stick Model: A Regression Algorithm for Irregularly Sampled Data with Application to eGFR." *arXiv preprint arXiv:1612.01409* (2016).

www.modellingCKD.org

Acknowledgement: This project entitled "Modelling the Progression of CKD" was awarded to NP by the Medical Research Council, under the New Investigator Grant Scheme MR/M023281/1.