

SAMueL-2 project meeting (July 2022)

Attendees:

- Leon Farmer [LF] – PCI lead
- Mike Allen [MA] – PenCHORD modeller
- Anna Laws [AL] – PenCHORD modeller
- Martin James [MJ] – Stroke consultant
- Peter McMeekin [PM] – Health economist
- Tom Monks [TM] – Deputy director of PenCHORD, modeller
- Catherine Pope [CP] – Medical sociologist. Qualitative
- Julia Frost – Qualitative researcher
- Kerry Pearn [KP] – PenCHORD modeller (note taker)

Apologies:

- Kristin Liabo [KL] – PPI researcher
- Richard Everson [RE] – Professor of Machine Learning
- Chrissie Walker [CW] – PenCHORD Research and Operations Manager
- Ken Stein [KS] – Professor of Public Health
- Keira Pratt-Boyden [KPB]

Patient and Carer Involvement (PCI) Group

LF presented a summary of the first PCI meeting:

- First PCI meeting held 27th June
- PCI group has 7 people from across the UK
- First meeting was dedicated to introductions and a very brief review of the project
- Plans for the next three sessions (approx dates):
 - October: Project overview
 - December: Explainable machine learning
 - February: qualitative research plans

Machine Learning - SHAP

Latest work on explainable machine learning may be found in our online Jupyter book (to accompany a paper): <https://bit.ly/explainable-ml>

Summary:

- To aid explainability we have reduced the machine learning model to 8 features (ROC AUC of 0.915 c.f. 0.922 for all features). These features have minimal covariance (maximum R-squared = 0.045).
- During testing models (random forest + XG-Boost) we found models built for each hospital independently would occasionally predict thrombolysis was given to a patient with a haemorrhagic stroke. While this was rare,

this could undermine confidence in the model. Models fitted to all data with hospital ID as a feature did not ever predict thrombolysis was given to a patient with a haemorrhagic stroke. We will therefore use models fitted to all data.

- The model predicts that thrombolysis can be given to a patient taking AF anticoagulation. MJ queried whether this occurs in reality. Post-meeting note: Of patients arriving within 4 hours of known stroke onset, 3.7% receive thrombolysis if taking AF anticoagulation, compared with 38% not taking AF anticoagulation.
- ‘Waterfall plots’ of SHAP values for individual predictions may be better if reversed (features become increasingly important as you look down the plot).
- When passing the 10K cohort of patients through all hospitals, hospital SHAP accounts for 86% of the observed variance in thrombolysis use.
- We trained an XG-Boost model to predict different choices in thrombolysis between hospitals with a high or low propensity to use thrombolysis (ROC AUC 0.73). Using this model we found that lower thrombolysing hospitals were less likely to give thrombolysis:
 1. In milder, or very severe, strokes.
 2. With increasing disability before stroke.
 3. When stroke onset time had been estimated (rather than known precisely).
 4. With longer onset-to-arrival times.
 5. With longer arrival-to-scan times.
 6. When patient is on anticoagulants for atrial fibrillation.

Actions

- Start SHAP presentations with a waterfall plot.
- Can waterfall plot be reversed in order?
- Find better name for ‘low’ and ‘high’ thrombolysing hospitals
- Put frequency (total # points) on SHAP violin plots

Qualitative research

- MJ informed the team that NHS England is putting together a proposal for a community practice which pairs up a low with a high IVT rate stroke team. This should be useful for qual work (post-meeting note: We have been contacted by Emma Jupp from NHS Sussex Integrated Care Board to be involved in these type of discussions, especially when we have results with identified hospitals).
- MJ: Aim for end of September for first meeting with RD&E stroke physicians (new intake was start of August).
- JF: The RDE&E workshop/chat with RD&E staff will not require extra paperwork. But any interaction with others (such as Keira hanging out on the ward) will require an NHS passport. Need a ‘carrot’ for best chance of

engaging sites - this could include giving them some detailed feedback on their site (needs new data).

- Qual work will be discussed in more detail in September project meeting.

Disability-level outcome model

- Current on the disability-level outcome model (based on time to thrombolysis and/or thrombectomy) may be found at: <https://bit.ly/stroke-outcome-modelg>.

Papers

The summary paper from SAMueL-1 has been published, with an editorial by *Stroke*:

Michael Allen, Charlotte James, Julia Frost, Kristin Liabo, Kerry Pearn, Thomas Monks, Richard Everson, Ken Stein and Martin James. Use of Clinical Pathway Simulation and Machine Learning to Identify Key Levers for Maximizing the Benefit of Intravenous Thrombolysis in Acute Stroke. *Stroke*. 2022;53:2758–2767 DOI 10.1161/STROKEAHA.121.038454 (<https://www.ahajournals.org/doi/10.1161/STROKEAHA.121.038454>). Editorial at: <https://www.ahajournals.org/doi/10.1161/STROKEAHA.122.039954>

Next meeting

Next project meeting is planned for 21 September 2022.