SAMueL Project Meeting

Please feel free to use, or not use, your cameras as much as you like.

May 2022

Agenda

- Introductions & apologies
- Why SAMueL?
- SAMueL-1 report and papers
- Qualitative research
  - Aims
  - Recruitment
  - Ethics
- PPI
- Quantitative research
  - Workstreams
  - Recruitment
  - HQIP data access
  - Machine learning
  - Outcome modelling
- AOB

Introductions & apologies

Why SAMueL?

SAMueL-1 Summary videos

- Short (10 min): `https://bit.ly/samuel_vid`
- Long (45 mins): `https://bit.ly/samuel_vid_long`

Summary of findings

SAMueL-1 report and papers

SAMueL-1 report and papers

Report (*NIHR Journals*, in press):

- Allen, M., James, C., Frost, J. Liabo, K., Pearn, K., Monks, T., Zhelev, Z., Logan, S., Everson, R., James, M., Stein, K. (2022). Stroke Audit Machine Learning (SAMueL): Use of simulation and machine learning to identify key levers for maximising the benefit of intravenous thrombolysis in acute stroke. Health and Social Care Delivery Research. (in press). Accompanying online book: Available online at: https://samuel-book.github.io/samuel-1/. DOI: 10.5281/zenodo.5078131

Summary paper (*Stroke*, in press):

- Allen, M., James, C., Frost, J. Liabo, K., Pearn, K., Monks, T., Everson, R., Stein, K., James, M., (2022). Use of simulation and machine learning to identify key levers for maximising the benefit of thrombolysis in acute stroke. Stroke (in press).

Paper on random forests (currently under revision for *Intelligence in Medicine*):

- James, C., Allen, M., James, M., Everson, R. Using machine learning and clinical registry data to uncover variation in clinical decision making.

Qualitative research

PPI

Quantitative research

Workstreams

- Explainable AI
- Further work on summarising differences in clinical decision-making
- Machine learning outcome prediction
- Inclusion of organisational audit factors in use of thrombolysis
- Granular outcome (mRS) prediction in pathway simulation model
- Health economics model
- Synthetic data
- Production code and prototype web application

Recruitment

Anna Laws joined us at the beginning of May. Welcome Anna!

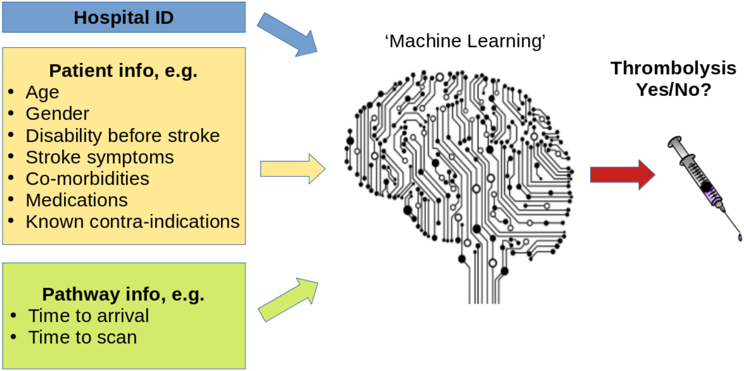
HQIP Data Access

- Some extra complications since SAMueL-1
- Draft HQIP request with SSNAP and an advisor from HQIP

Machine Learning (Kerry Pearn)

Learning the thrombolysis decisions made at each hospital

We restrict the machine learning model to those 40% patients who have a chance of thrombolysis by arriving at hospital within 4 hours of known stroke onset.



Models are about 85% accurate.

Machine Learning - general work covered

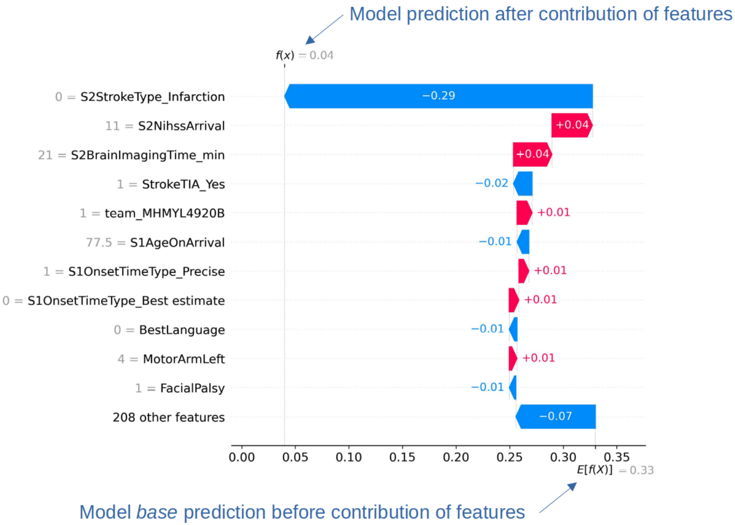
Current work is using the SAMueL-1 data set.

- *Model optimisation using Optuna*: A reviewer criticised lack of ML optimisation in SAMueL-1, so we are covering this more in SAMueL-2. Optimisation so far shows model defaults, and parameters used in SAMueL-1, are close to optimial.
- *New model types*: Two new model types have been evaluated: XG-Boost and LightGBM (both derivatives of random forests), no increase in accuracy but these models are significantly faster than random forests to run, so may be of benefit when running code on SSNAP servers.
- *Feature re-engineering*: Simplified features help improve explainability of the model.
- *Explainable AI*: 'Shap' values (more detail in following slides).

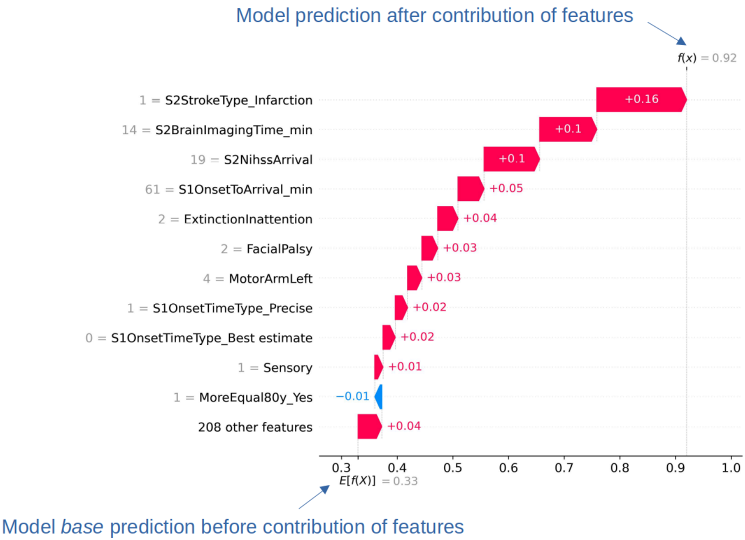
Shapley values

- Named after the Game Theorist Lloyd Shapley, who won the Nobel Prize in Economic Sciences for his work.
- Shapley values provide the marginal contribution of each player in a game - that is the value they add when present on a team compared with if they were not present on a team.
- Imagine 4 people are needed for a pub quiz team, and there is a group of 6 people who could be selected from the team:
  - There 15 possible combinations of people.
  - If you could test all 15 combinations the Shapley value of each team member would be the difference the presence of any player makes on average compared to the average quiz score across all combinations of players.
- Shapley values describe the contribution of features in a ML model in a way that may be applied to any type of ML model (*'model agnostic'*).

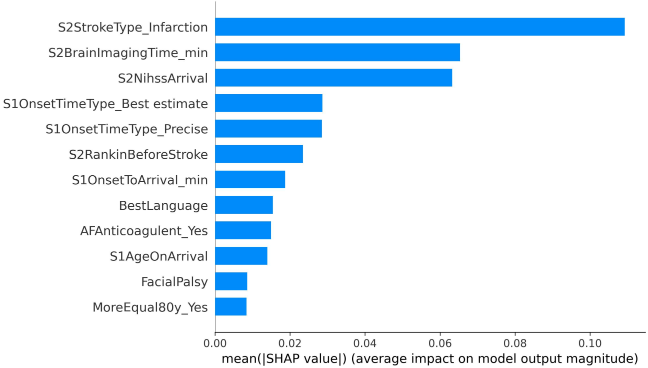
Shap example for a patient who is predicted to not receive thrombolysis



Shap example for a patient who is predicted to receive thrombolysis



Overall effect of features across test data set



(This is where we discovered we needed to refine feature engineering as ischaemic and haemorrhagic stroke types were present as separate features).

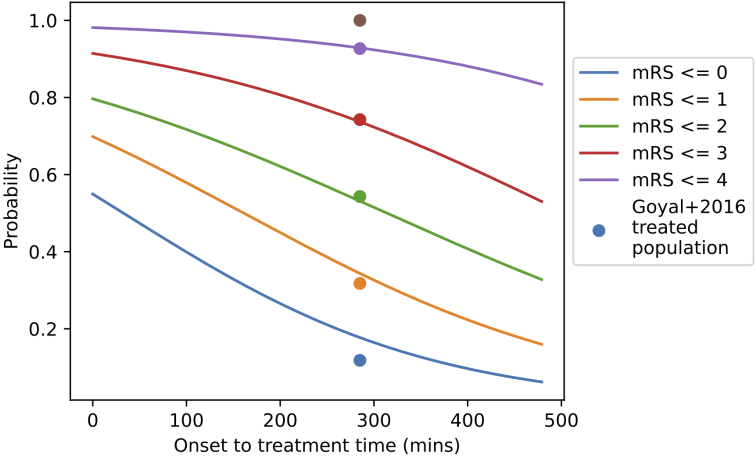
We also found that looking at global influential factors misses features that are important locally (e.g. to a particular one stroke team).

Next steps

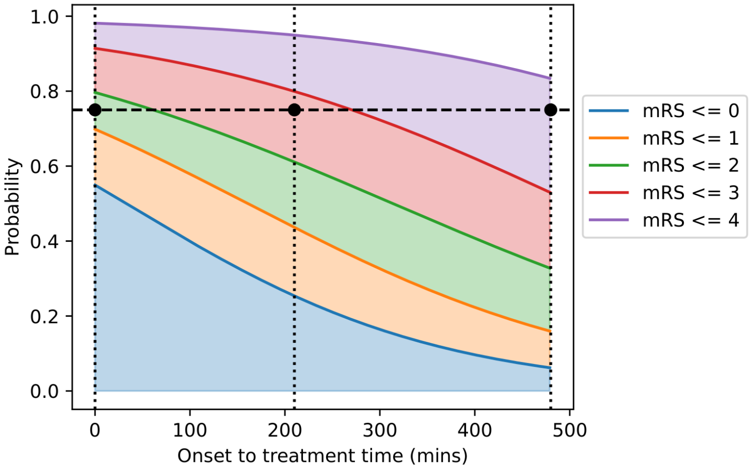
- Digging deeper into Shapley values and what they are telling us about the models.
- Simplified models (feature selection)?
  - But may lose features important only in a small number of units?

Granular stroke outcome prediction (Anna Laws, Charlotte James & Peter McMeekin)

Predicting outcome (modified Rankin Scale) based on time to treatment (thrombectomy)

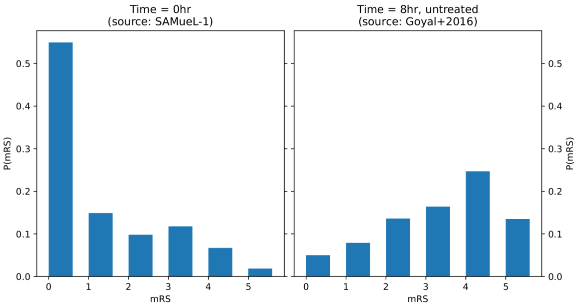


Predicting outcome (modified Rankin Scale) based on time to treatment (thrombectomy)



Assumed Modified Rankin Scale scores for treated at t=0 or at time of no effect (8hrs)

- If thrombectomy performed at *t=0*, then there is no additional disability due to stroke (use baseline mRS distribution).
- If thrombectomy performed at 8hrs (time to no net benefit in Goyal's analysis), then mRS distribution is the same as Goyal's control group.
- These distributions, and the time to no effect may be changed (e.g. perform sensitivity analysis around alternative possible values).



Next steps

- Code up for use in models.
- Explore different assumptions
- Investigate different model structures (longer term)

AOB

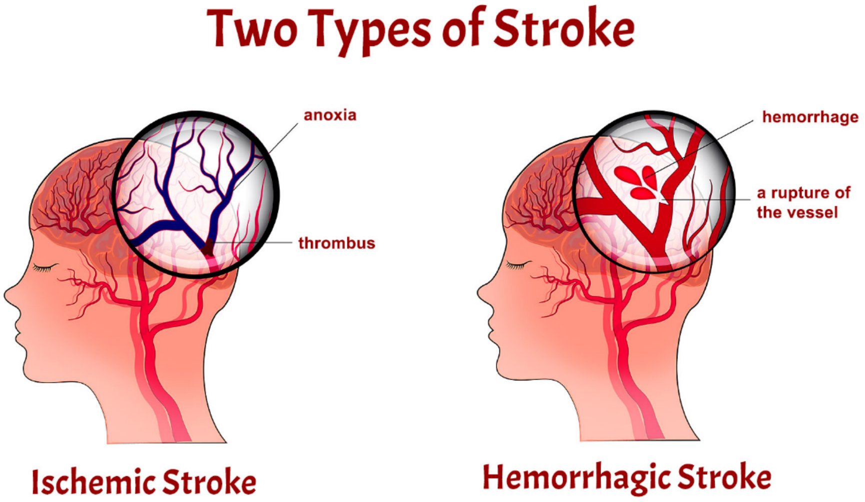
Online project books

https://samuel-book.github.io/samuel-1/

https://samuel-book.github.io/samuel-2/

Additional slides: SAMueL-1 Summary

Two types of stroke



Thrombolysis rates in UK are stable at 11-12% (compared to target of 20%)

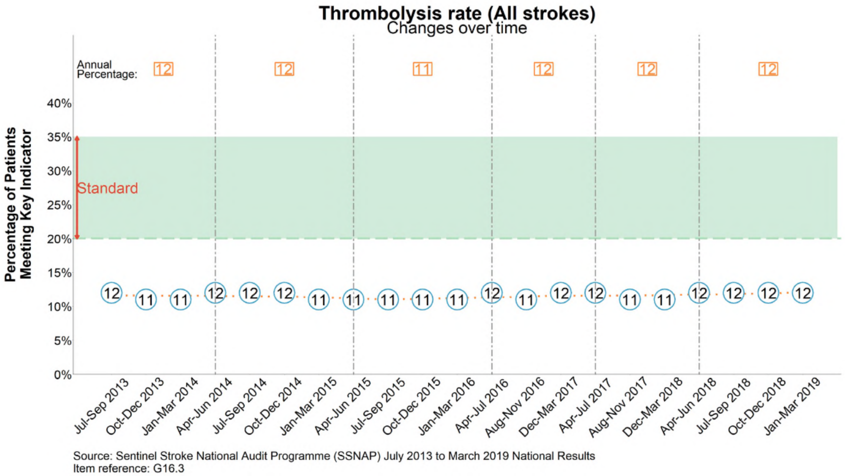
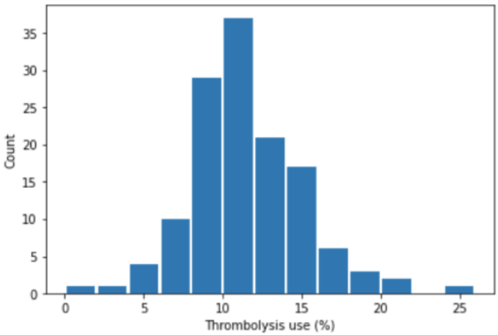


Figure 8. Percentage of patients who received Thrombolysis between 2013 and 2019.

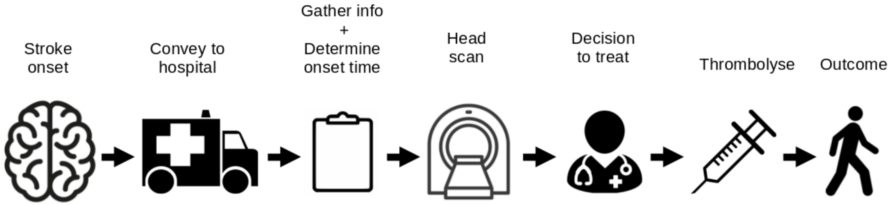
Thrombolysis rates vary considerably between hospitals - why?



What causes this variation?

- Patients?
- Processes?
- Decision making?
- A mix of the above?

Our model pathway

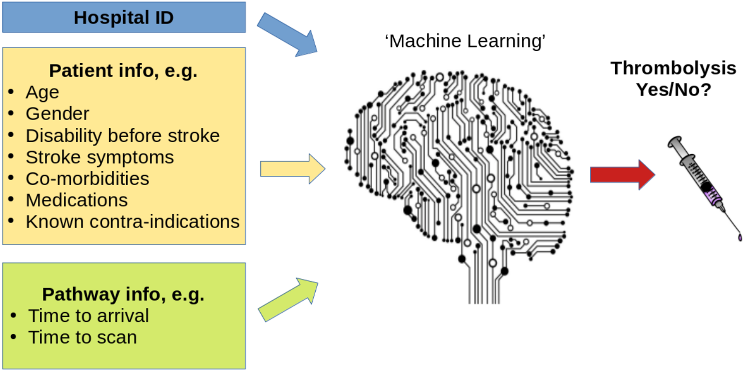


We model three key changes (alone or in combination):

1. Pathway speed
2. Proportion of patients with determined stroke onset time
3. Making decision-to-thrombolyse based on decisions learned at a set of *benchmark* hospitals

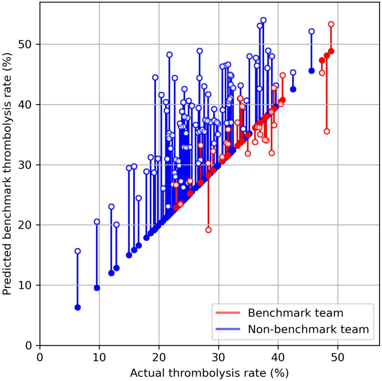
Learning the thrombolysis decisions made at each hospital

We restrict the machine learning model to those 40% patients who have a chance of thrombolysis by arriving at hospital within 4 hours of known stroke onset.



Models are about 85% accurate.

What if all decisions were made according to the majority vote of 30 benchmark hospitals?



If all decisions were made according to the majority vote of the 30 benchmark hospitals:

- Thrombolysis would increase from 29.5% to 36.9% (for patients arriving within 4 hours of known stroke onset) - a 25% increase in use of thrombolysis.

Summary of findings