SAMuEL-2: Stroke Audit with Machine Learning 2

Investigating variation in clinical decision-making with explainable AI

Anna Laws, Kerry Pearn, and Michael Allen

University of Exeter Medical School, PenCHORD

November 2022

Outline

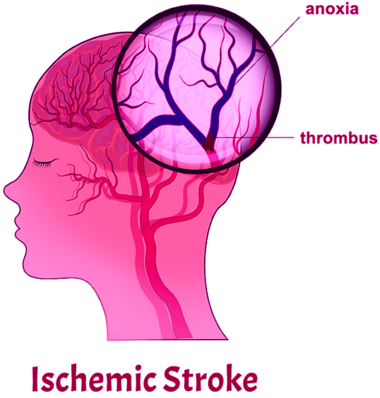
1. Background
    - Stroke and treatment
2. Explainability of machine learning
    - Simplify the model
    - SHAP
3. Outcome modelling
    - Streamlit app

Background

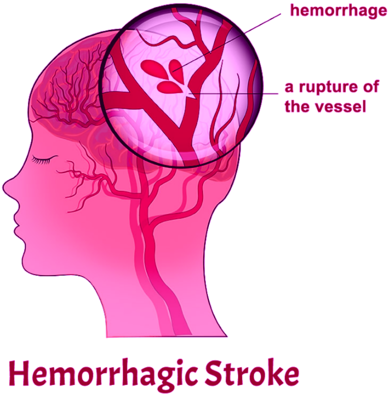
Background:

Stroke and treatment

Two types of stroke



(Infarction)



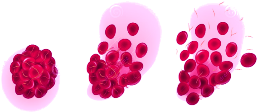
Brain scans reveal which type of stroke a patient has.

Treatments for ischaemic stroke

Thrombolysis aims to break down a clot by activating the body's own clot breakdown mechanisms.

Thrombolysis is given as an injection followed by an infusion (*drip*).

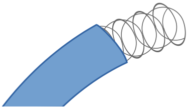
Thrombolysis is available at most hospitals.



Thrombectomy aims to grab and remove the blockage.

Thrombectomy uses a mesh device that enters the blocked blood vessel and physically removes the clot.

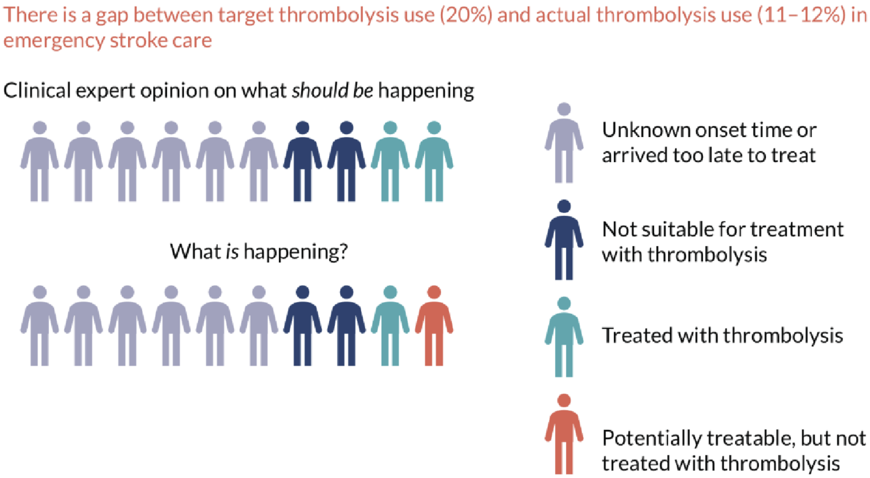
Thrombectomy is available at relatively few hospitals.



Downsides: not everyone is eligible for treatment; the treatments become less effective the later they are given; and there is a small chance of death.

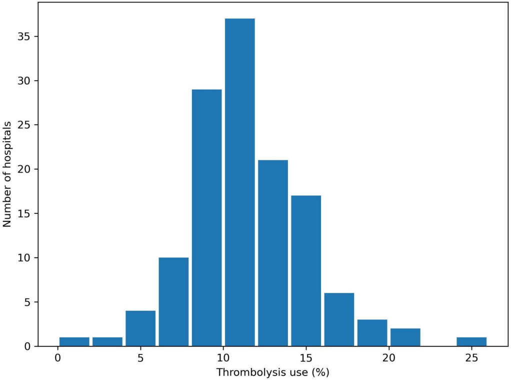
SAMueL-1 focussed on thrombolysis use.

What is the problem?



Thrombolysis rates in England and Wales have been stable at 11-12% for 10 years.

Use of thrombolysis varies considerably between hospitals



https://samuel-book.github.io/samuel-1/pathway_sim/explained_variance_in_thrombolysis.html

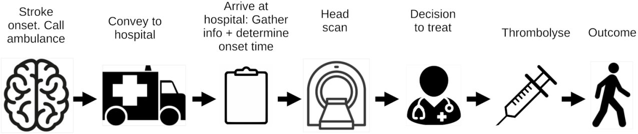
https://samuel-book.github.io/samuel-1/introduction/scientific_summary.html

A high level view of learning clinical decision-making

We accessed 240,000 emergency stroke admissions in England and Wales over three years.

Patient pathway:

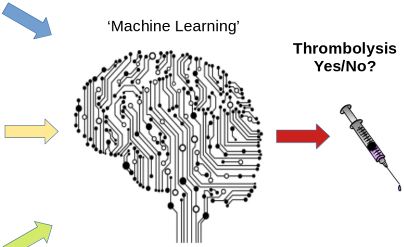
The onset time might not be known precisely.



Hospital ID

Patient info, e.g.
- Age
- Gender
- Disability before stroke
- Stroke symptoms
- Co-morbidities
- Medications
- Known contra-indications

Pathway info, e.g.
- Time to arrival
- Time to scan



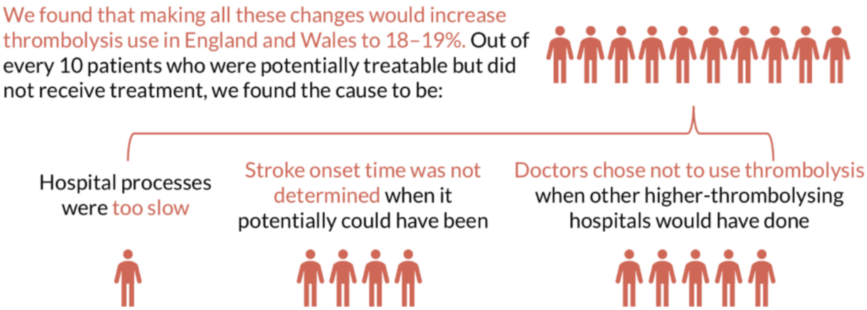
Machine learning models include logistic regression, random forest, XGBoost, and neural networks.

Here we are using XGBoost models.

"What if?" questions from SAMueL-1

- What if arrival-to-treatment time was 30 minutes?
- What if all hospitals determined stroke onset time as frequently as an *upper quartile* hospital (ranked 25 out of 100 for determining stroke onset time).
- What if decisions to thrombolyse were made according to a majority vote of 30 benchmark hospitals?

For each hospital we use their own patients to ask these questions, to allow for differences in local patient populations.



What questions are we asking in SAMueL-2?

What causes this variation in thrombolysis rates, and what could reasonably be achieved at each hospital (allowing for each hospital's own patient population)?

- Which patients do clinicians agree and disagree on, when considering when they should receive thrombolysis?
- How are the thrombolysis pathway and decision-making affected by *organisational factors* (such as use of specialist stroke nurses)?
- How best can we engage clinicians in our work, and prompt them to reconsider their emergency stroke pathway and/or decision-making?
  - Communication of general findings.
  - Web application for individual hospitals.
  - A 'hospital profile' for each hospital.

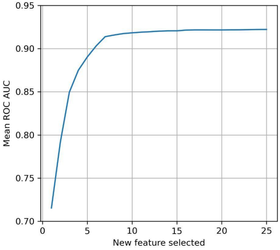
Explainability of machine learning

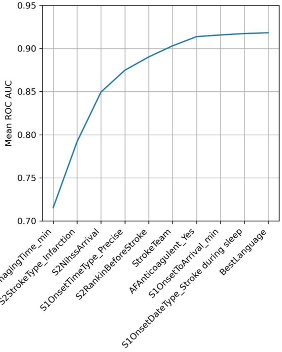
Explainability of machine learning:

Simplify the model

Simplifying the model with feature selection

- Our full data set has 85 feature - but which are most important?
- A model with fewer features is easier to understand and explain
- We build features by selecting them one at a time according to which new feature adds most accuracy
- We measure accuracy with ROC-AUC (Receiver Operating Characteristic Curve - Area Under the Curve). But don't worry about that - it is just a robust accuracy measure.
- We find 8 features give us nearly as much accuracy as all features



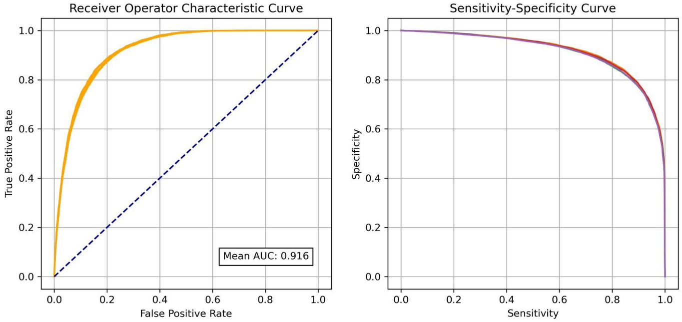


Features selected:

- Arrival-to-scan time
- Infarction
- Stroke severity
- Precise onset time
- Prior disability level
- Stroke team
- Use of AF anticoagulants
- Onset-to-arrival time

https://samuel-book.github.io/samuel_shap_paper_1/xgb_with_feature_selection/01_xgb_combined_fit_feature_selection.html

Model accuracy after selection of 8 features



- Accuracy = 85%
- ROC AUC = 0.916
- The model can achieve 84% sensitivity and specificity simultaneously
  - Sensitivity: The proportion of patients who receive thrombolysis who are correctly classified
  - Specificity: The proportion of patients who do not receive thrombolysis who are correctly classified

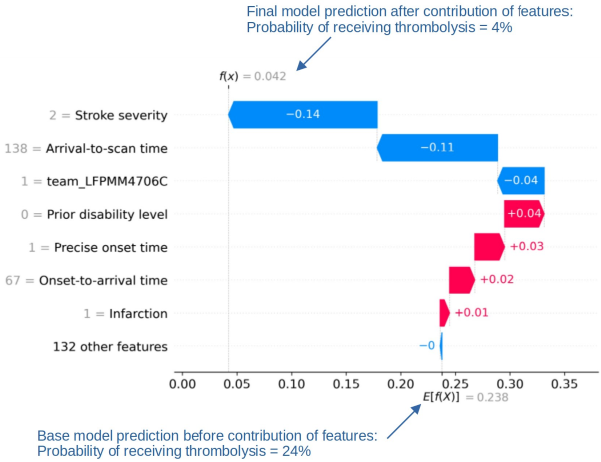
https://samuel-book.github.io/samuel_shap_paper_1/xgb_with_feature_selection/02_xgb_combined_fit_accuracy_key_features.html

Explainability of machine learning:

SHAP

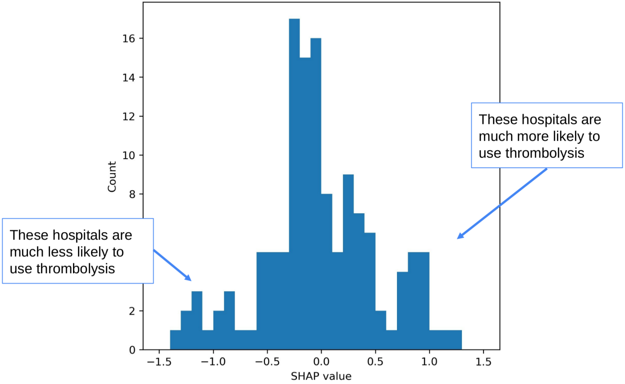
SHAP output for an individual patient

SHAP values show the influence of features (even for *'black box'* models).



For more on odds, probabilities and SHAP see: `https://samuel-book.github.io/samuel_shap_paper_1/introduction/odds_prob.html`

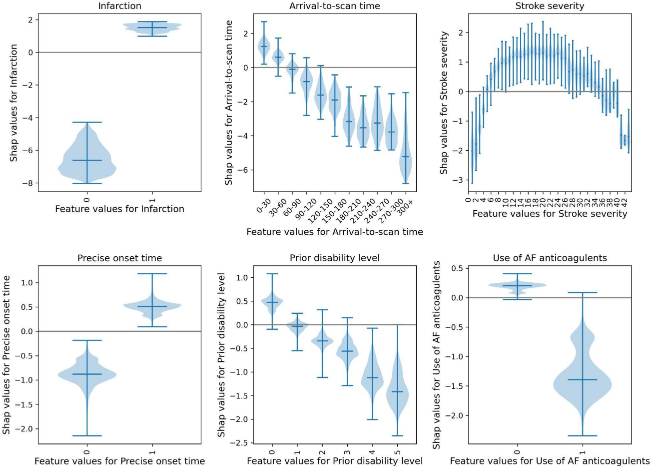
Average SHAP values for each hospital



Average SHAP values for each hospital range from -1.3 to +1.3.

https://samuel-book.github.io/samuel_shap_paper_1/xgb_with_feature_selection/03_xgb_combined_shap_key_features.html

SHAP values for predicting use of thrombolysis across all hospitals



Note: SHAP values here are *log odds*. Each step-change in value of ± 1 changes the chances of receiving thrombolysis about 3-fold. (Plots are in order of feature importance.)

https://samuel-book.github.io/samuel_shap_paper_1/xgb_with_feature_selection/03_xgb_combined_shap_key_features.html

Outcome modelling

How do we define outcomes?

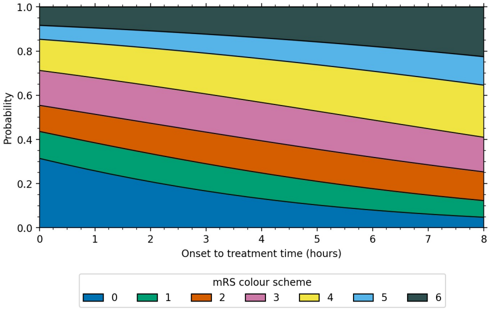
Modified Rankin Scale (mRS) is a measure of disability or dependence for people with stroke. Because mRS bands are not linear, we also use utility as a measure of quality of life.

Outcome	mRS	Utility	Description
"Good" outcome	0	0.97	*No symptoms.*
"Good" outcome	1	0.88	*No significant disability.* Able to carry out all usual activities, despite some symptoms.
"Bad" outcome	2	0.74	*Slight disability*. Able to look after own affairs without assistance, but unable to carry out all previous activities.
"Bad" outcome	3	0.55	*Moderate disability.* Requires some help, but able to walk unassisted.
"Bad" outcome	4	0.20	*Moderately severe disability.* Unable to attend to own bodily needs without assistance, and unable to walk unassisted.
"Bad" outcome	5	-0.19	*Severe disability.* Requires constant nursing care and attention, bedridden, incontinent.
"Bad" outcome	6	0.00	*Dead.*

One of our aims is to calculate changes using all of the mRS bands, so the results are more finely-grained than the previous "good" or "bad" outcome options.

Outcome variation with time

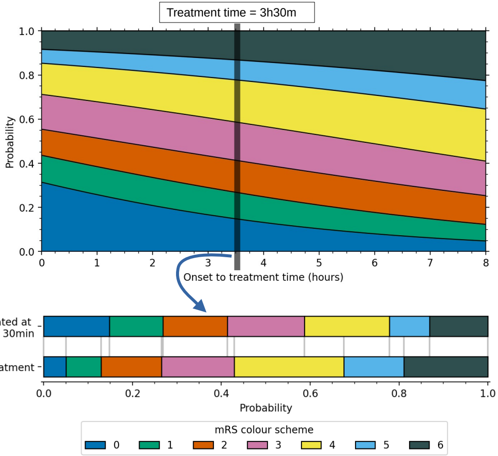
Expected outcomes of patients with large-vessel occlusions treated with thrombectomy:



The variation with time is a logistic function.

https://github.com/samuel-book/stroke_outcome/blob/main/mRS_outcomes_maths.ipynb

Expected outcomes of patients with large-vessel occlusions treated with thrombectomy:



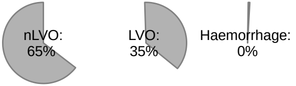
Treated		Not treated	Proportion
(0	–	1)	$\times 0.079 = -0.08$
(0	–	2)	$\times 0.019 = -0.04$
(1	–	2)	$\times 0.117 = -0.12$
(1	–	3)	$\times 0.003 = -0.01$
(2	–	3)	$\times 0.145 = -0.15$
(3	–	4)	$\times 0.158 = -0.16$
(4	–	5)	$\times 0.102 = -0.10$
(5	–	6)	$\times 0.058 = -0.06$
			Total : -0.72

https://github.com/samuel-book/stroke_outcome/blob/main/mRS_outcomes_maths.ipynb

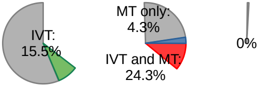
Patient population

Patient population

1: Decide what proportion of the population has each stroke type.



2: Decide what proportion of each stroke type will receive each treatment.



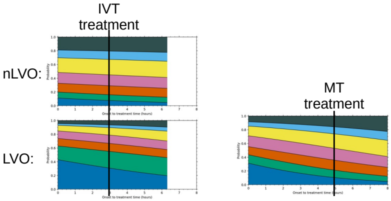
Treatment time

1: Pick a treatment time for IVT and for MT.

IVT: 3 hours

MT: 5 hours

2: Find the mRS probability distributions at the chosen times.



3: Find the mean changes in mRS and utility at these times compared with the non-treated population.

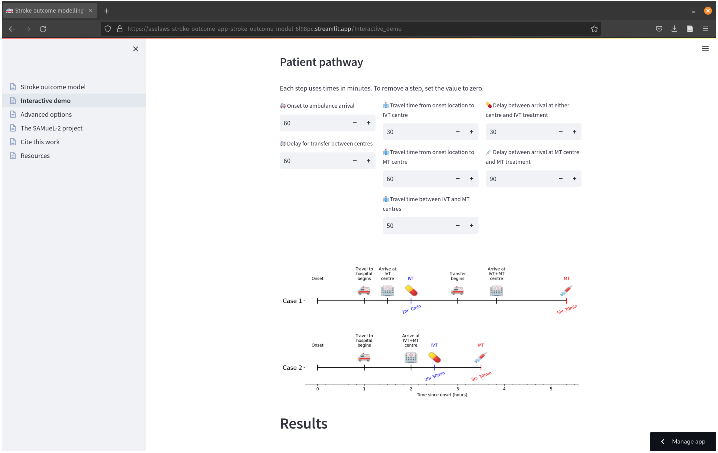
Result: Multiply these proportions by these changes to find the mean population change in mRS and in utility.

Outcome modelling:

Streamlit app

Streamlit app

Available here: https://samuel2-stroke-outcome.streamlit.app/



The app is linked to this (messy!) GitHub repository: https://github.com/aselaws/stroke_outcome_app

Summary

SHAP:

- An XGBoost model with 8 features has 85% accuracy at predicting use of thrombolysis
- Shapley (or SHAP) values show how much each feature influences the model prediction
- The probability of receiving thrombolysis is increased with:
    - Ischaemic stroke
    - Short arrival to scan time
    - Stroke severity 6-35
    - Precisely known onset time
    - Low prior disability
    - No use of anticoagulants for AF
    - Short onset-to-arrival times
    - Attending hospitals with high propensity to use thrombolysis
- We find that low-thrombolysing hospitals are more sensitive to these features.

Stroke outcome modelling:

- Many data sources have been combined to create modified Rankin scale (mRS) distributions as a function of time
- These can be used to find the expected changes in mRS and utility of a patient population
- An application of this is comparing the population outcomes depending on choice of hospital

Thank you!!

Thank you for your time and attention!