STROKE-TWIN: Optimising flow and patient outcomes in the emergency stroke pathway using Stroke System Digital Twins (NIHR161438)

Summary of research

> ***Your decision to treat or not treat … That's the difficult part. That's the grey area where everyone does a different thing."*** — Stroke Consultant during interviews for NIHR SAMueL (Stroke Audit Machine Learning) project.

Aim

This research aims to reduce life-altering differences in emergency stroke care by identifying and addressing hospital-level variations in the use and speed of proven clot-removing treatments. These treatments are 1) the use of 'clot-busting' drugs (intravenous thrombolysis, or 'IVT') and, 2) mechanical removal of clots (mechanical thrombectomy, or 'MT'). We will combine multiple modern 'AI' technologies to create what is called a 'digital twin'. A digital twin is a generic model of a real-world system, that may be tailored to a particular instance. In our case a digital twin is a model that can mimic the behaviour of any regional stroke system in England or Wales. Users can see performance of the regional system (including how clinical decision-making compares to other teams), and investigate how changes in the system would likely affect patient outcomes. These digital twins will be made available to the stroke community to use as a web app, prior to which we will prospectively investigate factors that will contribute to its implementation and co-produce theory-informed resources to support implementation. This will pave the way for targeted interventions to improve patient outcomes and streamline healthcare resources.

Background

Thrombolysis and thrombectomy are proven treatments that reduce disability after an acute stroke. Use of these treatments is significantly below NHS ambitions, and varies considerably between hospitals. Our team applies a range of advanced methods to understand variation in emergency stroke care, especially the removal of clots causing the stroke, and the effect that such variation has on patient outcomes. We map access to different types of treatment depending on where people live. We have models that mimic the flow of patients through each hospital (using a modelling approach called 'discrete event simulation'), and can identify for each hospital how improving that flow could improve patient outcomes. Using 'explainable machine learning' we also learn the characteristics of patients each hospital will, or will not, give clot-busting drugs to (to remove the blockage causing a stroke). Differences in these clinical decisions are a major cause of variation in treatment, and understanding these differences will help reduce this variation and improve outcomes. We have recently begun to make these models available to service providers as a web app (https://bit.ly/sam2_app) which is already being used to support an NHS England quality improvement initiative (Thrombolysis in Acute Stroke Collaborative, 'TASC'). Previous work has been highlighted by NIHR (https://bit.ly/samuel-highlight), and has recently won a Health Data Research-UK *Reproducibility Recognition* for the quality of the reproducibility and accessibility of the work.

Planned work

We will build on our previous work in several substantial and important ways by:

1. Extending our work, previously only examining thrombolysis, to cover mechanical removal of clots ('thrombectomy'). We will look at variation in the access to, and speed of, thrombectomy. We will compare decisions on which patients are selected for thrombectomy and we will use explainable machine learning to predict outcomes with and without thrombectomy, depending on patient characteristics.

2. Updating our analysis to include the 2023 National Clinical Guideline for Stroke (strokeguideline.org). This includes extending use of clot-busting drugs to more people who wake up having had a stroke in their sleep.

3. Investigating how variation in treatment is affecting in-patient lengths of stay, 'LoS', (which affects both the patients and the hospitals who provide the inpatient care) and discharge destination.

4. Pulling all the work together in a single model called a 'digital twin'. People may select which hospital or region they wish the digital twin to mimic, and ask 'what if?' questions.

5. Working with stakeholders to understand barriers to implementation, tailoring the output and user-interface of the digital twin to make it more effective in driving and supporting quality improvement, and co-producing resources to support implementation.
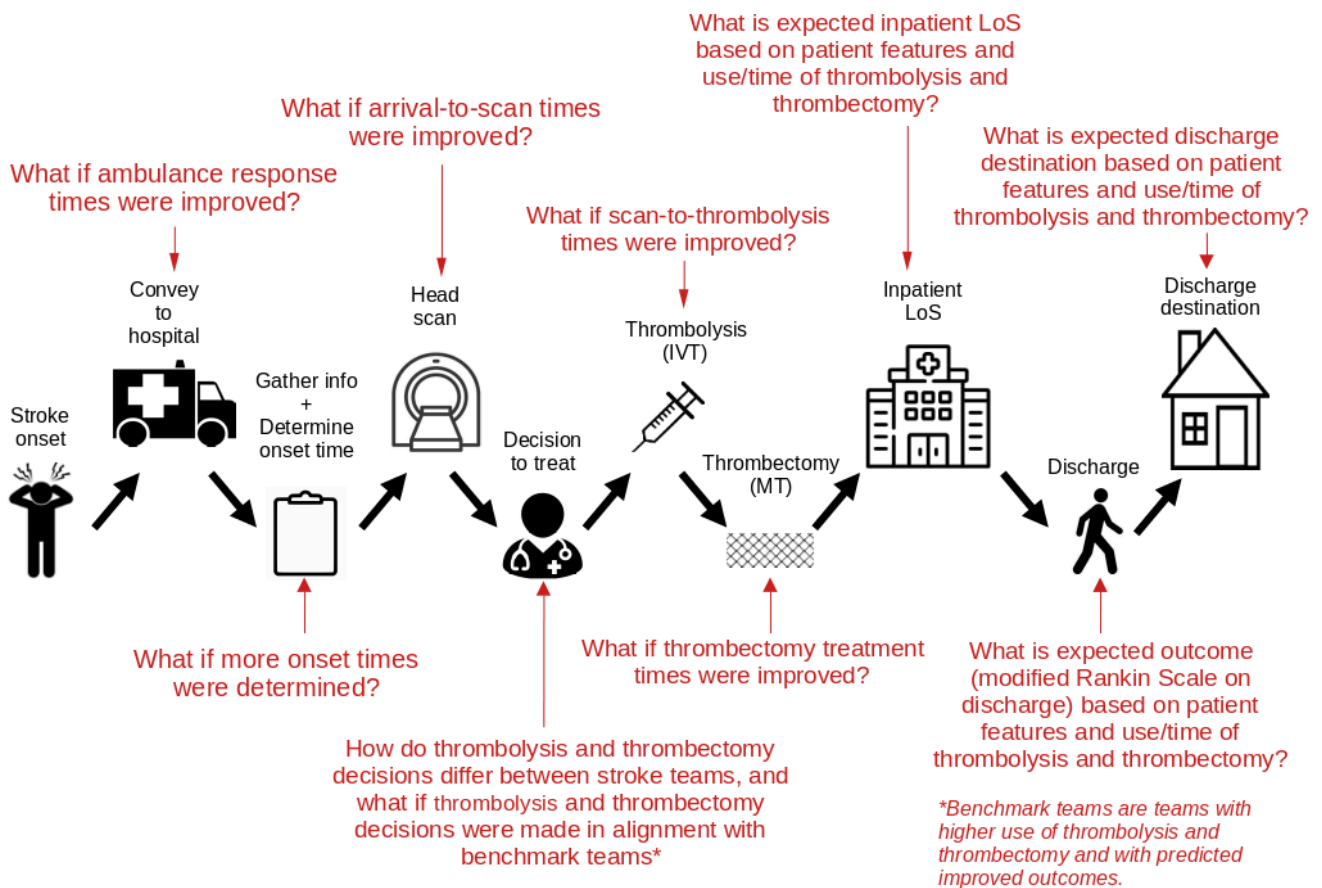
Figure 1: High level view of the digital twin pathway modelling

Public and patient involvement

We will continue to work with our highly engaged Patient and Public Involvement team, which is chaired by a stroke-survivor (Leon Farmer), who is a co-applicant on this application. This team has been key in pushing the team to focus on patient outcomes and not just NHS targets for use of thrombolysis and thrombectomy.

Link to patient benefit

We have strong direct links for how our findings will benefit patients. The national stroke audit 'SSNAP' will incorporate our findings from summer 2024, providing key output from our work to all stroke teams. We are also a key member of a new project, TASC, driven by NHS-England and managed by NHS-Elect, working with stroke teams across England to improve use of thrombolysis. Our current web app (thrombolysis only) will be launched in March 2024 through the TASC project. Our previous stroke modelling work has also been incorporated into policy documents such as the national clinical guidelines for stroke care [36] and the UK

guide to implementation of thrombectomy [37]. We anticipate that the proposed work will have immediate impact through SSNAP and TASC, and we anticipate it to be cited in policy documents on the benefit of thrombolysis and thrombectomy in the real world.

Overview of the digital twin

The overall pathway under study, and the analysis provided, is shown in figure 1, above. Clinical pathway simulation will be used to model the flow of patients through the system, with timings (including between-patient variation) tailored to individual stroke teams. Explainable machine learning will be used to predict the probability of individual patients receiving thrombolysis and/or thrombectomy depending on patient characteristics and the stroke team(s) attended. Explainable machine learning will also be used to learn patient outcomes (modified Rankin Scale, 'mRS') depending on use and time of thrombolysis and/or thrombectomy. The clinical pathway simulation and explainable machine learning will be combined into a single "digital twin" where "What if?" questions may be asked (see figure 1). This will be made available by a web app (developed from our existing web app which models just use of thrombolysis, and does not include inpatient stays or discharge destinations).

Background - What is the problem?

Stroke remains a leading cause of death and disability, globally and in the UK. Clots, which are the cause of most strokes, may be dissolved or removed with reperfusion treatments if given soon enough after a stroke. These treatments include thrombolysis (a clot-busting medication given by injection/infusion [1]) and, more recently, thrombectomy (the mechanical removal of a clot performed in a specialist centre [2] which can provide both thrombolysis and thrombectomy, hereafter called a thrombectomy centre). Hospitals which only provide thrombolysis will be referred to as thrombolysis centres.

Despite thrombolysis being shown to be beneficial, use varies very significantly between hospitals. In England and Wales the national stroke audit reported that in 2021/22, thrombolysis rates for emergency stroke admissions varied from 1% to 28% between hospitals (median rate of 10.4%) [3]. For contrast, the NHS England Long Term Plan states that 20% of emergency stroke admissions should be receiving thrombolysis. Additionally, in 2022/23, 3% of emergency stroke admissions received thrombectomy, significantly below the expected eligibility of at least 10% of these patients [4].

Our research team uses various modelling techniques, including explainable machine learning and clinical pathway simulation, to investigate the variation across different hospitals during the first few hours of stroke care. We use these models to understand the source and impact of that variation on patients [5;6]. We have identified that the majority of between-hospital variation in thrombolysis use comes from differences in hospital processes and decision-making, rather than from differences in local patient populations [5;6]. Additionally, we perform geographic analysis to study the geographical variation in access to thrombolysis and thrombectomy [7]. We are working with the Sentinel Stroke National Audit Programme (SSNAP) and NHS England, who use our modelling results to better understand the root causes of this between-hospital variation in thrombolysis use, to help them to achieve their aim of improving stroke care.

Why is this work important?

The disparity in treatment of stroke patients can cause a worsened patient outcome, which in turn can also cause an increase in downstream healthcare needs, such as extended inpatient stay. In this proposed project we will build on existing work, by extending the modelling to include a wider representation of the pathway, and bringing the separate work strands together into a single analysis framework. Our modelling will aid understanding of the variation in the delivery of both thrombolysis and thrombectomy between hospitals. We will make our work and analysis accessible for the wider stroke community to interrogate as an interactive web app.

Review of existing evidence

Studies have shown that reasons for low and varying thrombolysis rates are multifactorial, including organisational factors [9-11]. For organisational factors, the centralisation of primary stroke centres has been shown to reduce barriers to thrombolysis [11-14]. However, a balance must be achieved, as we have shown that the centralisation of stroke care can also worsen access to reperfusion therapies in some areas [7].

In addition to organisational factors, clinicians can have varying attitudes as to which patients are suitable candidates for thrombolysis. In a discrete choice experiment [15], 138 clinicians considered hypothetical patient vignettes, and identified the patients they considered eligible for receiving thrombolysis. The authors showed there were considerable differences among respondents in their thrombolysis decision-making.

Based on national audit data from three years of emergency stroke admissions, we have previously built models of the thrombolysis pathway. We use clinical pathway simulation to model the flow of individual patients through the pathway. We use the model to look at the effects of changing pathway processes or the effect of replicating clinical decision-making around thrombolysis from higher thrombolysing hospitals to lower thrombolysing hospitals [5-6]. Using these models we found that it would be credible to target an increase in average thrombolysis in England and Wales, from 11% to 18%, but that each hospital should have its own target, reflecting differences in local populations. We found that only a minority of variation in thrombolysis use arose from differences in population casemix, whereas the largest increase in thrombolysis use would come from replicating thrombolysis decision-making practice from higher to lower thrombolysing hospitals. A summary of the findings is shown in the graphical summary in figure 2.

Though we have performed extensive work on using clinical pathway simulation and explainable machine learning to study variation in thrombolysis, we have not conducted any work on variation in use of thrombectomy. We have also not performed any work predicting and modelling inpatient length of stay, and how that might be affected by thrombolysis or thrombectomy. Also, since our original work, the national clinical guidelines for stroke allow for extended use of thrombolysis - in patients that are late-arriving but where advanced imaging shows salvageable brain tissue. We would like to add these patients into our modelling.

We have recently extended the model to include explainability of our machine learning models [8]. This enables us to understand the patterns that the model has learnt from the data. Example findings are that the odds of receiving thrombolysis were 3-fold greater for patients with a precise stroke onset time (versus an estimated onset time), and that there was a 13-fold difference in odds of receiving thrombolysis between hospitals (after allowing for patient and process speed differences). Our models revealed that hospitals have different levels of risk tolerance for non-ideal patient characteristics. Lower thrombolysing hospitals were particularly less likely to give thrombolysis to patients with 1) mild stroke, 2) imprecisely known onset time, 3) prior disability, 4) any combination of those. We would like to perform similar work for thrombectomy.

We have also recently extended the explainable machine learning to predict outcomes with and without thrombolysis. This has not yet been published but some aspects are included in this proposal. This work found that thrombolysis in practice has the benefit reported in the clinical trials, but has allowed a fuller understanding of patient features that most influence outcome (which are stroke severity, prior disability, age, use/time of thrombolysis, stroke team, and presence of atrial fibrillation). We would like to perform similar work for thrombectomy.

To allow stroke teams to access our modelling, we have just released a pilot web app (bit.ly/sam2_app). The web app allows stroke teams to compare their own thrombolysis use decisions with the decisions made at other hospitals. It also allows teams to investigate the impact of changing the pathway process speeds, the ascertainment of the stroke onset time, or changing decision-making criteria on thrombolysis use and outcomes. Outcome prediction is based on a mathematical model of the clinical trials for thrombolysis. The web app also contains a health economics model to analyse the impact of a patient's characteristics to predict their life expectancy, utility, costs, and quality-adjusted life year (QALY).
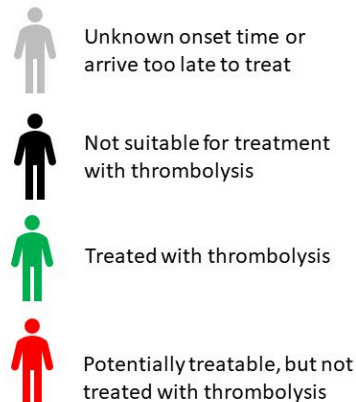
SAMueL – Stroke Audit Machine Learning

What problem are we addressing?

There is a gap between target thrombolysis (**20%**) and actual thrombolysis use (11-12%) in emergency stroke care

Clinical expert opinion on what *should be* happening

What *is* happening?

Unknown onset time or arrive too late to treat

Not suitable for treatment with thrombolysis

Treated with thrombolysis

Potentially treatable, but not treated with thrombolysis

What did we test?

We used clinical pathway simulation and machine learning to analyze a series of 'what if?' questions:

1. What if arrival-to-treatment speed was 30 minutes?
2. What if all hospitals determined stroke onset time as frequently as the 'upper quartile' hospital (a hospital ranked 25 out of 100 hospitals)?
3. What if decisions were made according to a majority vote of 30 benchmark hospitals?

What did we find?

We found that making all these changes would increase thrombolysis use in England and Wales to 18-19%. Out of every 10 patients who were potentially treatable, but did not receive treatment, we found the cause to be:

Hospital processes were **too slow**

Stroke onset time was not determined when it potentially could have been

Doctors chose not to use thrombolysis when other higher-thrombolysing hospitals would have done

Figure 2: Graphic summary of findings of previous work on use of thrombolysis.

The web app is currently in use as part of the TASC programme ('Thrombolysis in Acute Stroke Collaborative') - a collaboration of NHS-Elect, NHS-England, our stroke modelling team, and the national stroke audit SSNAP to help improve use of thrombolysis. This collaborative will be working with six low-thrombolysing centres in 2024, before extending work to all emergency stroke centres.

Current NIHR-funded work (https://fundingawards.nihr.ac.uk/award/NIHR134326) is focusing on building explainable machine learning models to predict death and disability-level outcomes based on patient characteristics, with and without thrombolysis (and with differing time to thrombolysis). Our team is also modelling the effect of pre-hospital selection of patients for thrombectomy on the time to thrombolysis and thrombectomy, and their outcomes (based on mathematical models of the clinical trials for thrombolysis and thrombectomy). As yet, these models are not part of the web app (the models are ready for a web app, but we wish to consult more on how to present them).

Research questions

We now propose to extend our modelling and web app especially to include variation in use of thrombectomy, to model in-patient length of stay (and how they would change with improved thrombolysis and thrombectomy

use), and discharge destination. Informed by discussions with clinicians and patients, we have created focused research questions, all of which are novel to this proposal.

1. How do thrombectomy centres vary in their selection of patients for thrombectomy?
2. How does variation in use and speed of thrombectomy affect patient outcomes?
3. How does variation in use and speed of thrombolysis and thrombectomy affect in-patient length of stay?
4. How does variation in use and speed of thrombolysis and thrombectomy affect discharge destination?
5. What is the effect of the new National Clinical Guideline for Stroke of an extended treatment window for late-presenting patients or patients who wake up with stroke?
6. How can the web app (user interface for the digital twin model) be implemented, embedded and normalised in practice to inform improvements in service delivery?
   a. What are barriers and facilitators to implementation?
   b. What resources should be developed to facilitate implementation?

Aims and objectives: Comparison of previous and planned work

The table below summaries what work has previously been completed, and what will be added in this project.

Previous/current work	Planned work

Explainable machine learning
● Prediction of use of thrombolysis for patients arriving within 4 hours of known stroke onset (depending on individual patient characteristics).
● Prediction of outcomes (disability: mRS) with and without thrombolysis for individual patients (depending on individual patient characteristics).
● Model of how organisational factors (from SSNAP organisational audit) affect thrombolysis use and outcomes.
 	
Pathway simulation
● Pathway from stroke onset to use of thrombolysis for patients arriving within 4 hours of known stroke onset.
 	
Health economics
● Adaptation of health economics models of McMeekin et al. applied to thrombolysis (conversion to Python, and adaptation for web app; see https://bit.ly/stroke-he). with and without
 	
Web app
● Web app launched: https://bit.ly/sam2_app. Web app includes:
  ○ Descriptive statistics for stroke teams.
  ○ Effect of thrombolysis pathway improvement for each team.
  ○ Thrombolysis choice at each team: how likely a team is to give thrombolysis to a defined patient, compared with other teams.
  ○ Health economics based on disability at discharge.
 	
Qualitative/implementation research
● The focus so far has been on what we should include in the modelling, by:
	Explainable machine learning
	● Prediction of use of thrombectomy for patients arriving within 6 hours of known stroke onset (depending on individual patient characteristics).
	● Prediction of outcomes (disability: mRS) with and without thrombectomy for individual patients (depending on individual patient characteristics).
	● Extension of thrombolysis explainable machine learning (use and outcome) to late-arriving patients, and patients with unknown stroke onset times.
	● Prediction of inpatient lengths of stay with/without thrombolysis/thrombectomy.
	● Prediction of discharge destination with/without thrombolysis/thrombectomy.
	
	Pathway simulation
	● Pathway extended to include thrombectomy, and thrombolysis in late-arriving patients, and patients with unknown stroke onset times.
	
	Health economics
	● Adaptation of health economics models of McMeekin et al. applied to thrombolysis and thrombectomy. Explainable machine learning provides mRS distribution with and without thrombolysis and thrombectomy as a key input into the health economics model.
	
	Web app
	● Extend web app to include thrombectomy, including user consultation and usability testing.
	● Develop a version of the choice/outcome explainable machine learning as an educational tool to be used on best use of thrombolysis and thrombectomy.
	
	Qualitative/implementation research
	● Nested qualitative study informed by Normalisation Process Theory to prospectively examine barriers and facilitators to implementation.

○ Ethnographic studies on thrombolysis pathway. ○ Interviews on use of thrombolysis.	● Co-production workshops with relevant stakeholders to develop theory-informed resources to support implementation. ● Focus on implementation of project outputs: ○ Co-production of project outputs and web app with TASC (an NHS-England quality improvement project).

Research plan / methods

Design and theoretical/conceptual framework

The overall pathway under study is shown in figure 1 above. The digital twin of this pathway will have these main elements:

- Explainable machine learning will be used to learn which patients are currently likely to receive thrombolysis and thrombectomy at each stroke team (allowing an understanding of variation in selection of patients).
- Clinical pathway simulation will be used to predict changes in thrombolysis/thrombectomy use and outcomes when combining possible changes to the pathway (e.g. changing pathway speed, selection of patients for direct conveyance to a thrombectomy centre, changing selection criteria for treatment with thrombolysis/thrombectomy).
- Health economics models to convert pathway model outputs (which will predict disability on discharge with alternative scenarios) to quality of life and life-expectancy measurements.
- Qualitative/implementation research informed by Normalisation Process Theory to prospectively examine barriers and facilitators to implementation, usability testing of the web app, and co-production workshops to develop theory-informed resources to support implementation.

We have previously used these explainable machine learning and clinical pathway simulation methods extensively for thrombolysis. Figure 3 shows the accuracy these models have. We can predict thrombolysis use at each hospital with very high accuracy ($r^2 > 0.96$ for prediction of percentage of patients who will receive thrombolysis at each stroke team). We can predict which patients, of those who arrive within 4 hours of stroke onset, will receive thrombolysis, with an area under the curve of the receiver operating characteristic (ROC AUC) of 0.916. We can predict outcomes (disability on discharge) with an ROC AUC of 0.811 for a multi-class model predicting all mRS levels (or a mean ROC AUC of 0.873 for dichotomised models with a separate model for predicting being less than each mRS threshold). For this project we plan to extend these methods to thrombectomy, and for extending outcomes to inpatient length of stay and discharge destination.

Accuracy of full thrombolysis pathway simulation model for prediction of thrombolysis in each stroke team for all out-of-hospital emergency stroke admissions.

Accuracy of machine learning model for prediction of thrombolysis use in each stroke team for patients arriving within 4 hours of known stroke onset (ROC AUC for individual patient predictions=0.916).

Accuracy of individual patient stroke outcome machine learning model, shown as a 'confusion matrix' (ROC AUC=0.811). Stroke outcome is shown as modified Rankin Scale (mRS).
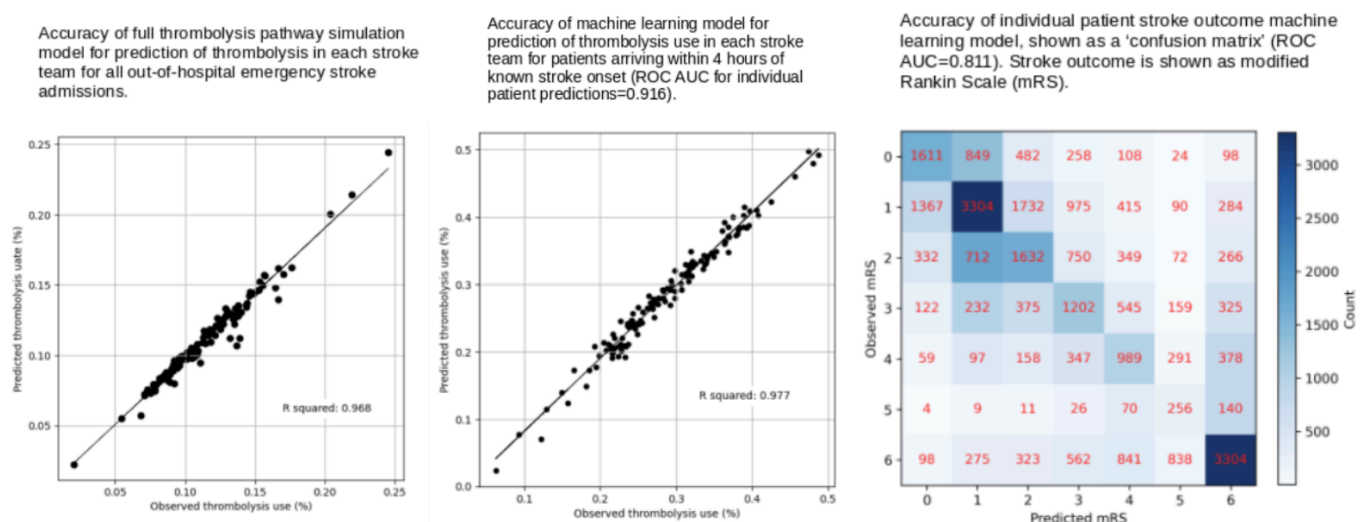
Figure 3: Accuracy of explainable machine learning and clinical pathway simulation in predicting variation in use of thrombolysis, and outcomes of patients (disability on discharge).

Quantitative Modelling Data

Quantitative data for the project will come mostly from SSNAP. We are well-rehearsed in obtaining national data from this source, ensuring data remains secure and anonymised.

Explainable machine learning for decisions to use thrombolysis

In this section we will describe work we have performed with thrombolysis. Our plan for the proposed project involves extending these methods to use of thrombectomy.

Key to our explainable machine learning models is providing explanations about how a model makes its predictions. The model type used (XGBoost) is considered a 'black box' model, but by building an additional explainability model on top of the main model we can show how individual features affected the model's final prediction. This is a SHAP (or SHapley Additive exPlanations) model. Figure 4 shows the model prediction, for use of thrombolysis, from an individual patient, and how each of the patient features contributed to the final model prediction.
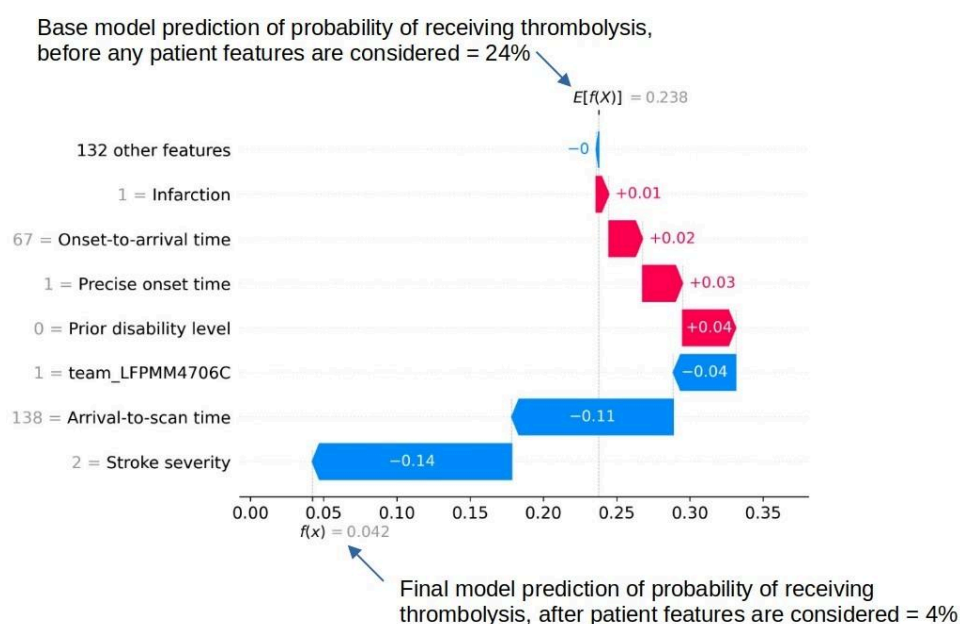

.

Figure 4: A SHAP waterfall plot showing the most influential features for the model's prediction of a patient's probability of receiving thrombolysis (in this case a patient with a very low probability of receiving thrombolysis). The feature values for this patient are shown in grey next to each feature name on the y-axis. In this example the three most influential features, reducing the chance of receiving thrombolysis, were 1) low stroke severity (NIHSS 2), 2) slow arrival-to-scan time (138 mins), and 3) the hospital attended (anonymised stroke team 'LFPMM4706C').

Individual patient explanations (using SHAP) may be combined to investigate general patterns learned by the explainable machine learning model. Figure 5 shows the general patterns identified in use of thrombolysis. For example, the chances of receiving thrombolysis fall with increasing arrival-to-scan times, and are also highly dependent on stroke severity (with the chances of receiving thrombolysis being highest with moderate and moderate-severe stroke). We can also see that the hospital attended significantly affects the chances of receiving thrombolysis (independent of other characteristics of the patient).
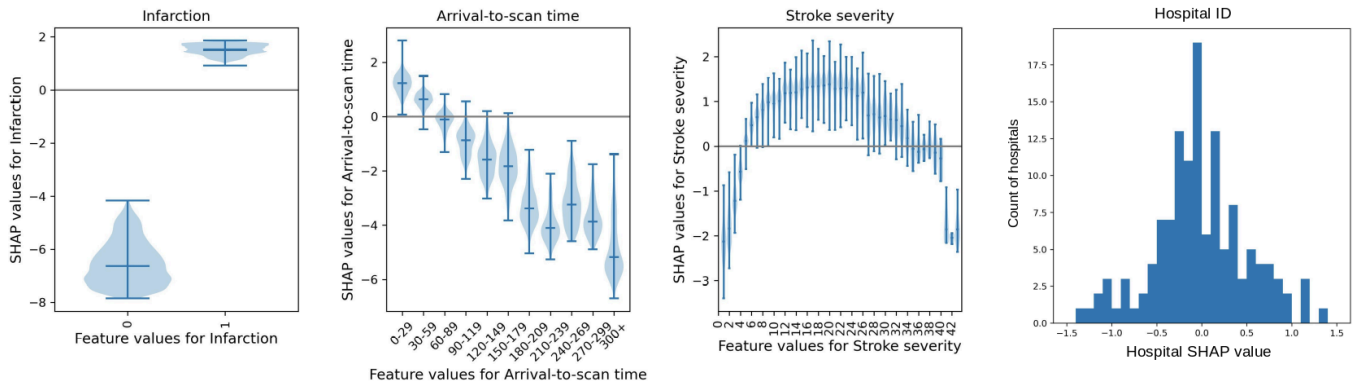
Figure 5: Violin plots showing the SHAP values (for decision to use thrombolysis) for three features, and histogram of hospital SHAP values. The shape of the violin shows the spread of the SHAP values for each feature value. The histogram shows the number of hospitals with different SHAP values. SHAP values show how feature values affect the log-odds of receiving thrombolysis. A positive SHAP value pushes the model towards predicting the patient would receive thrombolysis.

Based on patterns we observed when using SHAP, we can identify patient prototypes and see how likely different teams are to give these different patients thrombolysis. Figure 6 shows an example of a hospital which is very unlikely to use thrombolysis in mild stroke (stroke severity scale [NIHSS] 3/42) compared with benchmark hospitals. This output may be used to investigate clinical variation in attitudes to use of thrombolysis.
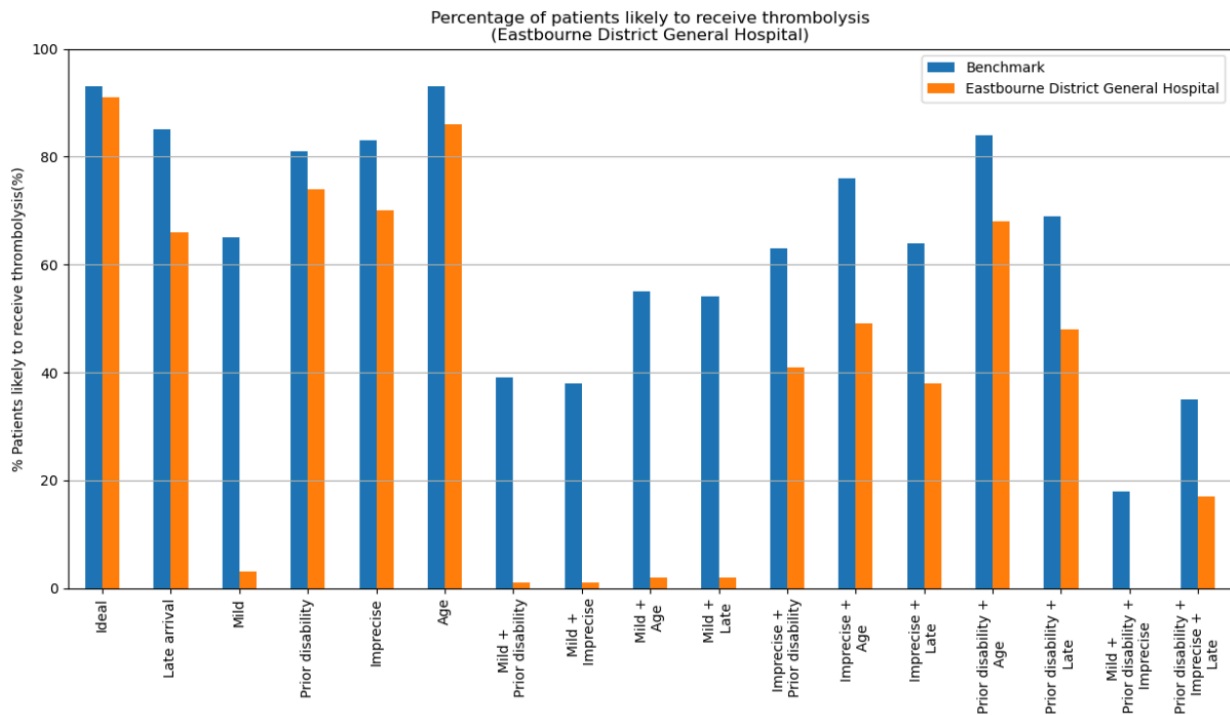
Figure 6: Predicted use of thrombolysis in different patient prototypes at an example hospital, compared with predicted use of thrombolysis in 25 'benchmark' hospitals (the 25 hospitals most likely to give thrombolysis to patients if all hospitals saw the same patients).

Explainable machine learning for outcome after stroke

Similar to the model predicting whether a patient will receive thrombolysis at any given hospital, explainable machine learning (XGBoost with SHAP) may be used to predict the patients disability outcome at discharge, and the SHAP values can be used to understand the influence of patient features on this outcome. Patient disability is measured using mRS, which is a 7 level scale from 0 (no disability), to 6 (death). These models may either be multi-class classification models (where the probability of each of the seven mRS levels is

predicted for each patient), or may be dichotomous models to predict whether a patient will be below a given disability threshold at time of discharge. Figure 7 shows an example of a dichotomous model for likelihood for being discharged with no or little disability (mRS 0-2). We see that the probability of having no or little disability at discharge reduces with increasing pre-stroke disability, reduces with age, increases with use of thrombolysis (with a decline in benefit over time), and reduces with increasing stroke severity.
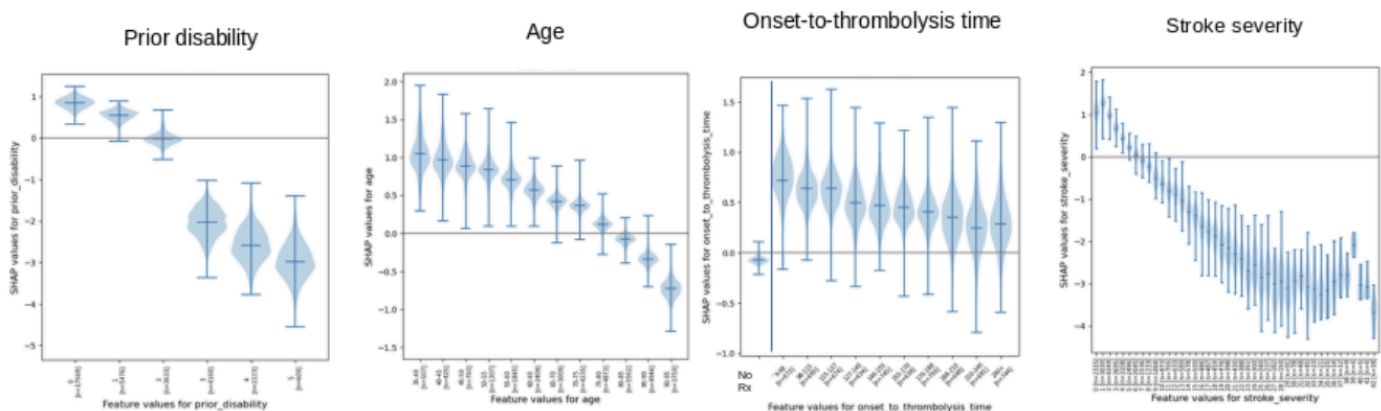

Figure 7: Violin plots showing the SHAP values for four of the model features for predicting patient outcomes. For this model, the SHAP values show how feature values affect the log-odds of being discharged with minimal disability (mRS 0-2). A positive SHAP value pushes the model towards predicting that a patient will be discharged with minimal disability.

Individual patient outcome modelling

When differences in choices around use of thrombolysis emerge, clinicians ask 'how do we know what is best?' Our explainable machine learning models predict the likely probability of outcomes (disability level at discharge) for any patient with and without thrombolysis. Figure 8 shows an example patient with significantly lower chance of receiving thrombolysis at a particular hospital, but where the chances of a good outcome would be expected to be improved by use of thrombolysis.
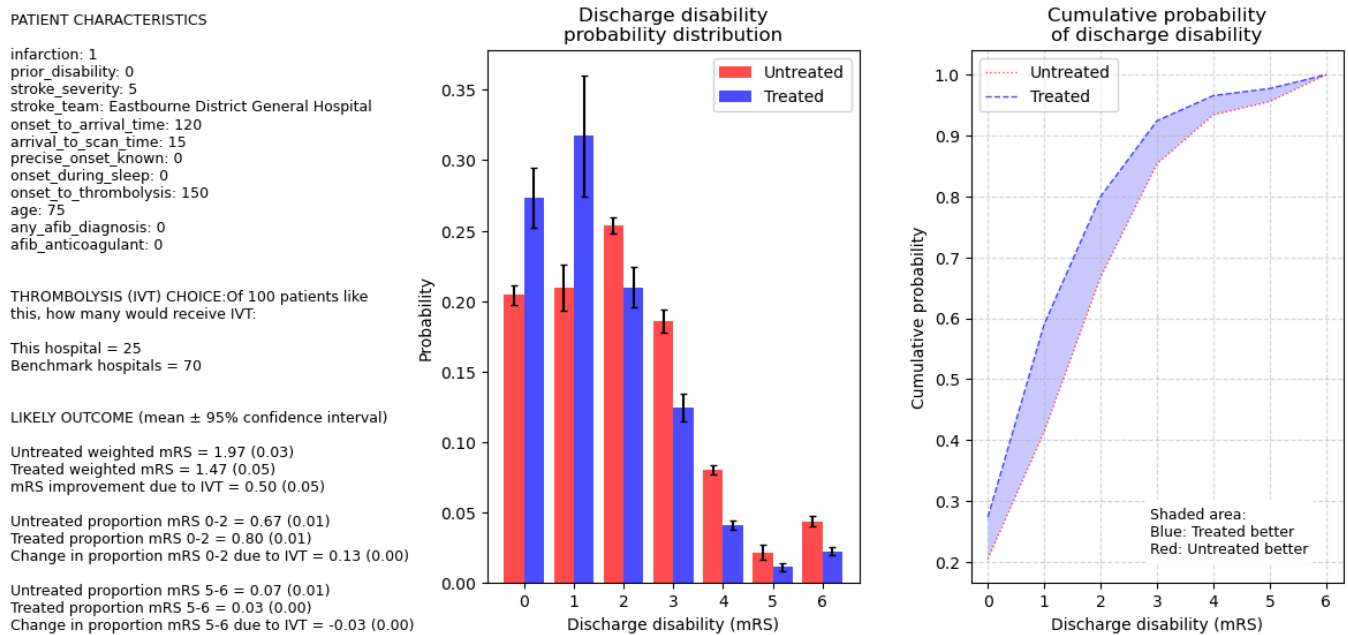
Figure 8: Thrombolysis use choice and outcome predictions for a single patient. This patient, with mild stroke, has a 25% chance of receiving thrombolysis at the example hospital, compared to 70% chance across 25 benchmark hospitals. The predicted outcome modelling shows that thrombolysis is likely to increase the probability of the patient being discharged with low disability (mRS 0-2) and reduce probability of being discharged with severe disability or death (mRS 4-6).

Inpatient length of stay and discharge destination

We have not built any models for inpatient lengths of stay, but we will use the same methodology described above to predict inpatient length of stay and discharge destination, and use SHAP values to examine the effect of patient characteristics, and the use of thrombolysis and thrombectomy on length of stay and discharge destination.

Clinical pathway simulation

Clinical pathway simulation is used to examine the effect of multiple changes to the pathway. This is used to identify which changes would be the best focus for improvement. The three changes examined for thrombolysis are:

1. **Pathway speed**: What if all patients received thrombolysis within 30 minutes from arrival?
2. **Onset determination**: What if the hospital was at least as good as the upper quartile hospital in the proportion of patients where time of stroke onset was determined?
3. **Benchmark decisions**: What if decisions to thrombolyse were made by the majority-vote decision of the 25 most enthusiastic users of thrombolysis ('benchmark hospitals')?

(See the *Web application* section below for an example of applying this clinical pathway simulation to thrombolysis only).

Application of explainable machine learning and clinical pathway simulation to thrombectomy and extended use of thrombolysis

The modelling methods to be used in this project have been described in relation to thrombolysis. We plan to take these same methods and apply them to thrombectomy and in extended use of thrombolysis (thrombolysis in patients arriving later than 4 hours after onset, and patients arriving with no known onset time). Specifically:

- Thrombolysis:
  - Explainable machine learning work will be extended to predicting, and comparing (between hospitals, and depending on treatment), inpatient lengths of stay, and discharge destination.
  - Repeat work on thrombolysis, but for patients where perfusion-imaging is used to select late-arriving patients, or patients with unknown stroke onset, by estimating the area of the brain which may still be salvaged by reperfusion treatment.
- Thrombectomy:
  - Use explainable machine learning to learn which patients are selected for thrombectomy, depending on the admitting unit.
  - Use explainable machine learning to learn outcomes, inpatient length of stay, and discharge destination with and without thrombectomy, taking into account patient characteristics.
  - Use clinical pathway simulation to combine multiple effects, such as speed of process, selection of patients for thrombolysis (which may frequently be a gateway step to receiving thrombectomy) and thrombectomy, and the effect of bypassing a local thrombolysis centre to directly attend a more distant thrombectomy centre.

Health economics data

The team have broad experience of health economics of stroke care (e.g. [20-22]). The health economics aspects of this project are very largely based on importing health economics models from other projects that the team is connected to. Specifically, we are incorporating a lifetime economic model of mortality and secondary care use for patients discharged from hospital following acute stroke that has been developed by Peter McMeekin's group based on primary collection of data for 1,509 patients for 6 years post-stroke, but has not yet been published. All patients received brain imaging to confirm a diagnosis of ischaemic stroke if required.

The long-term model consists of a set of risk equations which estimate life expectancy (two equations) and lifetime secondary care usage (three equations). Life expectancy modelling mimicked the approach of the UK Diabetes Outcome model [24]. Each equation takes the mRS score at discharge, age at stroke onset and gender as inputs to predict survival or resource use in order to calculate the risk of an event at a single point in time, the cumulative risk up to a point in time, or the number of events up to a point in time. Simulation using life tables (a table showing the probability that a person will die before their next birthday based on their age) was used to estimate the relative mortality risk of patients in the cohort compared to the general population. Lifetime secondary care use was estimated from the patient cohort adjusted for projected survival. Conversion of mRS to utility uses published conversion factors specific to stroke [23].

This work, though not yet published academically, has been coded into Python and made available by the Exeter modelling team (https://github.com/samuel-book/stroke_outcome_app), which is also available in a web app (https://stroke-predictions.streamlit.app/Lifetime_outcomes).

We will apply this work to the proposed project, with the new proposed project providing modelling of mRS at discharge after thrombolysis and/or thrombectomy, based on observed disability benefit learned through explainable machine learning (see the *Individual patient outcome modelling* section above for example of mRS outcome explainable machine learning for thrombolysis; this will be extended to thrombectomy, and late-presenting thrombolysis). By applying this long-term model we will extend patient outcomes to life expectancy, utility, QALYs, and long term healthcare costs.

Web application

The web app is developed in streamlit (streamlit.io) which allows data scientists to create web applications without the need of IT specialists. Figure 9, on the next page, shows example output from the clinical pathway simulation, made available through the current web app (https://bit.ly/sam2_app). For each team, the effect of three key changes, alone or in combination, on thrombolysis use and benefit is shown below. This helps teams identify the part of the pathway that would give maximum return for quality improvement (QI) investment. This will be extended to use of thrombectomy, and the effect of changes on more outcome measures (full mRS profiles, inpatient lengths of stay, and discharge destinations).

Additionally, as part of the project we will develop a prototype web app aimed at training new and existing stroke clinicians in choice of use of, and outcome from, thrombolysis and thrombectomy. This will allow clinicians to investigate interactively the likely outcome of thrombolysis and thrombectomy as patient features are changed.

Web app target audience

NOTE: The project will produce substantial academic output in addition to the very applied web app (see the *Dissemination* section for proposed paper titles).

There are three primary target audiences for the web app. The qualitative implementation work (described below, specifically the usability testing) will help inform whether one combined web app is produced, or whether it would have alternative front-ends for different audiences. The three target audiences are:

1. **Stroke team leads and/or stroke team data analytics experts**. The objective is to provide stroke team leads and analysts with tools to understand how their reperfusion treatment (thrombolysis and thrombectomy) compares with others, and the effect on outcomes and inpatient lengths of stay.
2. **Integrated Stroke Delivery Networks (ISDNs)**. The objective is to provide ISDNs with an overview of their system, allowing them to target quality improvement initiatives where most benefit is likely to be achieved.
3. **Quality improvement (QI) teams**. The objective is to provide QI teams, such as TASC or individual hospital QI teams, with the analysis they need to identify which parts of the emergency stroke pathway should be prioritised for improvement.
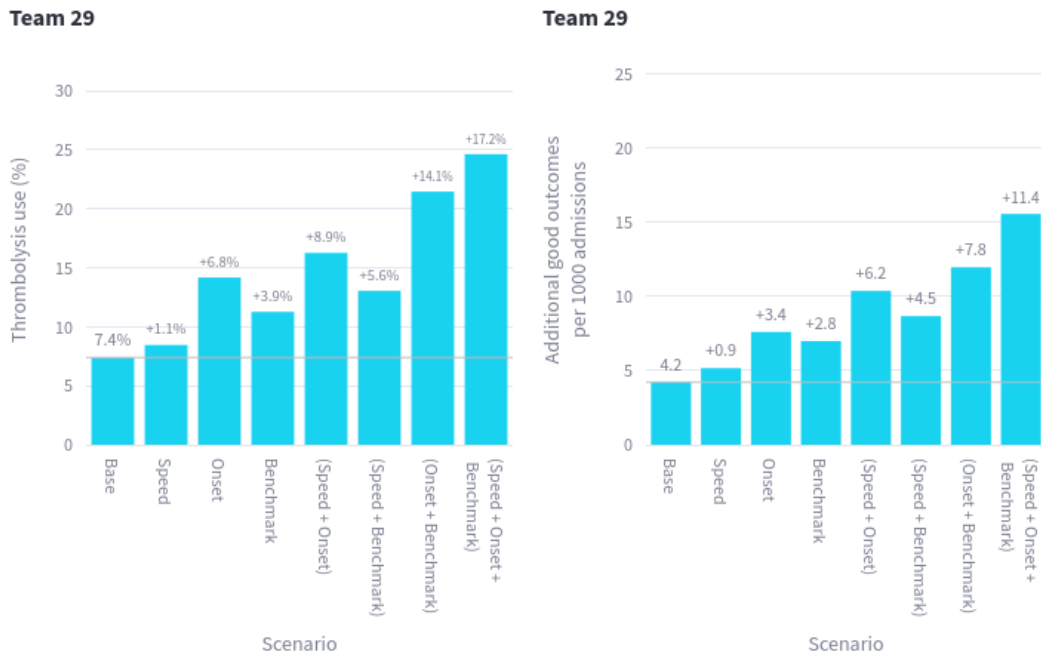
Figure 9: Web app screenshot on use of clinical pathway simulation and explainable machine learning to model the effect of changing different aspects of the pathway: speed, determination of stroke onset time, and applying benchmark thrombolysis decisions. These changes are applied to predict changes in use of thrombolysis (left) and outcome (right).

Web app sustainability

The web app is built using Streamlit (https://streamlit.io/) which is a web app platform designed for data scientists to produce web apps, without the need for specialist IT resources. This makes the app easier to build and support long term. Our aim is to pass core code to the national stroke audit to deploy, with NIHR PenARC (who have a demonstrable long term commitment to applied research in stroke) also agreeing to underwrite support for the web app. We reduce our models to key patient and hospital factors; this aids explainability of the model but also helps to protect the model against changes in detailed data availability that may occur over time.

Qualitative research

Design

Qualitative data will focus on implementation of the web app, collected over two sequential stages:

1. Examining implementation using 'real-life' modelled vignettes.
2. Usability testing of the web app and co-production of theory-informed implementation resources.

Qualitative data collection will have a start date offset by six months from the wider study to allow for the development of initial models, obtaining Health Research Authority (HRA) and local governance approvals, and construction of the vignette templates.

In stage 1, 'real-life' vignettes will be constructed based on the developed models, followed by semi-structured Normalisation Process Theory (NPT) [25,26] interviews to investigate how the developed web app can be

13

implemented, embedded and normalised in practice to inform improvements in service delivery. 'Real-life' vignettes, i.e. those based on real rather than hypothetical data or situations, have been suggested to establish credibility and to reduce idealised answers amongst participants [27]. By combining the vignettes with NPT-based interviews, we will prospectively examine factors that contribute to the potential implementation of the web app. The interviews will be informed by the four NPT constructs of coherence, cognitive participation, collective action and reflexive monitoring that together will provide the explanatory power for understanding how the web app can become integrated into existing practice through individual and collective action, along with a broader examination of the extent to which modelled comparisons between sites are considered able to inform quality improvement activities. These constructs are an important focus for pre-implementation work as alignment of sociotechnical factors may be a more significant challenge for implementation than challenges associated with hardware or software [28].

In stage 2, we will first conduct qualitative usability testing of the web app using Think Aloud interviews [29]. Data generated will inform refinements to the project-specific web app, including whether one combined web app is produced, or whether it would have alternative front-ends for different audiences, as well as our broader understanding of how to develop user-friendly web apps based on explainable machine learning. Following usability testing, we will then host two co-production workshops which will likely develop theory-informed implementation resources, informed by data from the 'real-life' vignettes and usability testing, to support implementation of the web app. The specific outcome of the co-production process will be determined in collaboration with participants and will be responsive to the findings from stage 1 to ensure suitability for practice, but may include modifications to the web app, guidance, recommendations, training materials and/or report structures.

Participants, sampling and recruitment

Up to 40 participants will be recruited to take part in the 'real-life' vignette interviews. Participants will be purposively sampled based on the type of stroke care delivered in their organisation (those with and without specialist stroke care including thrombolysis and thrombectomy). All participants will have significant responsibility for the organisation of stroke care within their organisations, primarily stroke leads, though additional snowball sampling may be employed should participants consider others to be relevant. Participants will also be sampled based on their organisation's current thrombolysis rates, including whether they are 'over-performing', 'under-performing', or 'performing as standard'. Recruitment of participants will be targeted to obtain representatives of the three groups that we anticipate to be the target audience: (1) stroke team leads and/or stroke team data analytics experts, (2) ISDNs, and (3) QI teams. To recruit participants we will use our team's well-established networks with TASC and ISDNs, as well as direct approaches to stroke and QI teams in individual organisations. It is anticipated that 40 participants, who will likely together represent at least 12 NHS Trusts, will provide sufficient information power to inform our understanding of implementation challenges [30].

For stage 2 (usability testing and co-production) up to 20 participants who took part in the 'real-life' vignettes will be invited to contribute, and we will additionally recruit up to four patient and public involvement (PPI) representatives with experience of stroke as either survivors or carers, who will ensure that the patient voice is represented in the produced implementation resources. PPI representatives will be identified with the support of the Stroke Association and/or from those who have been supporting our other work on co-producing acute stroke pathways (https://fundingawards.nihr.ac.uk/award/NIHR153982). PPI representatives will be paid for their time following INVOLVE guidelines (https://www.nihr.ac.uk/documents/payment-guidance-for-researchers-and-professionals/27392), and will also be offered additional support prior to each workshop so that any technical terms can be explained using lay language, with a glossary of terms we have developed for previous co-production adapted and updated as the project progresses so that they can be referred to at any time. Our PPI co-applicant and the stroke representatives will support us in ensuring any materials are aphasia-friendly, following the Stroke Association Accessible Information Guidelines [31].

Data collection

Prior to data collection in stage 1 we will construct a series of standardised 'real-life' vignettes. These vignettes will then be adapted based on the participants' own organisation and provide visual (graphical, see earlier sections) and narrative comparisons with other organisations matched based on the type of care delivered (e.g. thrombectomy centre compared with another thrombectomy centre). Where possible, and depending on the participants' own organisation's performance, this will include comparisons with organisations over-performing, under-performing and/or performing as standard. Including participants from over-performing organisations will allow for examination of how the web app may also inform understanding of best practice, rather than just examining opportunities for improvement. Vignettes will be shared with participants prior to the interview. During the interview, which will be in-person or online, participants will first be asked to reflect on their overall perception of what the vignettes represent, then will be asked questions based on Normalisation Process Theory to ensure we develop a theory-informed understanding of implementation. Questions will use a format that we have developed and used previously [32] and will have a focus on how the use of modelled data could inform quality improvement practice. NPT has previously been used to prospectively examine the implementation of a risk tool in primary care with a combination of vignettes and NPT-based interviews [33].

Think Aloud interviews (stage 2), conducted in-person or online, will consist of participants being given full access to the web app and asked to constantly narrate their actions. Where interviews are hosted online, participants will be asked to share their screen so that the researcher can see how participants are interacting with the web app. Participants will be questioned on why they are performing specific actions, and whether they encounter any frustrations or reservations related to the underlying data and/or the user-interface.

To develop implementation resources, we will host two co-production workshops lasting up to two hours, with supplementary semi-structured interviews offered where participants wish to contribute but cannot attend the workshop. Our experience of conducting co-production workshops in this field indicates that due to diary constraints, particularly of clinicians, around half of participants (n=12) would attend the workshops, and the other half would participate in interviews. Co-production, whether in workshops or interviews, will begin with a presentation of the study aims and a summary of the findings of the qualitative work to date. Participants will be asked to identify resources required to address the identified implementation challenges. Once identified, we will then co-produce a framework for the resources, including components such as structure, mode of published delivery and responsibility. An example is that if participants identify the need for training materials on the use of the web app, the framework could include the structure of the materials (length, format [slides, video]), mode of delivery (online, face-to-face) and responsibility (model of training, i.e. centrally delivered by ISDNs, stroke leads or generic online training). These components would be designed to be sufficiently flexible, allowing us to be responsive to data across the whole project. Following the first workshop, the research team will develop an example of the resources which will be presented back to participants in the second workshop for refinement and modification. All qualitative data will be audio-recorded and transcribed verbatim.

Data analysis

Data from the 'real-life' vignette interviews will be analysed using framework analysis [34], with the four NPT mechanisms (coherence, cognitive participation, collective action and reflexive monitoring) and their sub-constructs forming the framework. To improve trustworthiness and rigour, analysis will be led by the qualitative researcher supported by Jason Scott who has extensive expertise in qualitative data analysis. A minimum of 10% of transcripts will be dual-coded and then discussed, with further presentation and discussion of analyses in dedicated data analysis sessions to which the whole research team will be invited. Think Aloud interviews will initially be analysed informally, with the research team documenting required changes, to inform speedy changes to the web app [35]. This will be followed by framework analysis, consisting of justifications for the required changes (which form the initial framework), including additional mapping of data to the stage 1 NPT framework should participants discuss broader implementation challenges associated with usability. Co-production data will be analysed concurrently with participants during data collection, with a more in-depth inductive thematic analysis conducted following each workshop round, incorporating all the data to provide an audit trail of the co-design process to ensure trustworthiness of the findings.

Sampling and selection of research sites

The explainable machine learning and modelling work is performed using data from all acutely-admitting stroke teams in England and Wales. The qualitative/implementation work will use six stroke teams involved with the TASC project, but will be supplemented with stroke teams with higher use of thrombolysis (as TASC is focussed initially on lower thrombolysing sites) to ensure we use a mix of teams. We will also ensure that we have a mix of smaller and larger units, and rural and urban teams, using 12 teams in total.

Project timetable

Period	Quantitative	Qualitative
Pre-work	- HRA (proportional ethics) approval for anonymised SSNAP data - Current web app (thrombolysis only) launched for teams involved in TASC	
Months 0-6	- New SSNAP data obtained. - Descriptive statistics of data - Preliminary thrombolysis choice and outcome models developed (using existing data prior to provision of new data) - Experimental work published to GitHub and as an online Jupyter book	Set up: - Researcher recruitment - HRA approval - Development of standardised 'real-life' vignettes.
Months 7-12	- Finalise thrombolysis choice and outcome models - Extend thrombolysis model to extended use (post 4.5 hours) - Build length of stay and discharge destination models - Extend pathway simulation to use of thrombolysis and inpatient lengths of stay - Have all models available in a prototype web application (with initial training materials) - Write academic papers 1-2 from list in *Dissemination* section - Experimental work published to GitHub and as an online Jupyter book	- Participant recruitment - Stage 1 'real-life' vignette data collection - Data analysis (feeding into web app design) - Preparation of web app for usability testing. - Write academic paper 4 from list in *Dissemination* section
Months 13-18	- Write academic paper 3 - Model refinement - Web app refinement - Publish new models - Experimental work published to GitHub and as an online Jupyter book.	- Stage 2 usability testing - Data collection (feeding into web app design) - Write academic paper 5
Months 19-24	- Web app refinement - Co-production of final training materials for web app - Co-write academic paper 6 - Experimental work published to GitHub and as an online Jupyter book	- Stage 2 co-production and analysis (feeding into web app design and training materials) - Co-production of final training materials for web app - Co-write academic paper 6

Possible barriers for further research, development, adoption and implementation?

We are confident that there are no technical or data-related barriers to the successful completion of the project, as this project will use methods that the team have a high level of experience and expertise in. The main barrier to further research, adoption and implementation is in acceptance and use of the project outputs. The team, however, are experienced in co-production of project outputs to overcome these type of barriers, and the project includes substantial qualitative/implementation work to fully support development of the

technology and especially the presentation of the technology (for example, feedback for our current SAMueL project is that we need to provide some simpler output and interfaces). By working closely with a dedicated QI team from NHS-Elect we also reduce barriers to use and implementation.

Impact

The potential impact of the project is to improve outcomes in stroke by better use of thrombolysis and thrombectomy, by eliminating (or reducing) unwanted between-hospital variation in use and speed of these life-saving and disability-saving treatments. The project will provide stroke specialists with a richer understanding of where thrombolysis and thrombectomy are of most benefit, and what pathway improvements will lead to most patient benefit. The web app will also provide regional planners (such as ISDNs) with an overview of their region, and where QI work will lead to most improvement in patient outcomes. The thrombolysis-only web app is being launched through TASC in March 2024, so we have an immediate route to use and impact, and we have the opportunity to utilise and fine-tune our models through collaborative working with the NHS-England TASC project.

We anticipate impact will also be aided by this being a fully Open Science project, with open code published on GitHub, and through online books of all work (e.g. see https://bit.ly/samuel_book for a related project).

Through already working with NHS Elect, our current modelling is already proving engaging to providers:

"Dear Martin (James).

I just wanted to drop you a Thank You email for visiting our Trust today [to present the stroke audit machine learning and modelling work] with your colleagues; your presentation was superb, so inspirational in terms of the difference we can make for the patients who are not currently receiving thrombolysis; a great illustration of how we compare in our decision making to the 'premier' hospitals.....I think this is amazing work with huge clinical benefit. If it helps, I'd be very happy to provide a letter of support on behalf of the Trust/ISDN.

Ailsa Brotherton,
Executive Director of Improvement, Research and Innovation
Improvement Director, National Improvement Board"

Our previous stroke modelling work has also been incorporated into policy documents such as the national clinical guidelines for stroke care [36] and the UK guide to implementation of thrombectomy [37]. We anticipate that the proposed work will have immediate impact through SSNAP and TASC, and we anticipate it to be cited in policy documents on the benefit of thrombolysis and thrombectomy in the real world. Additionally our work has been highlighted as being ground-breaking in an editorial on our work in *Stroke* [38], and we would expect the proposed work to have the same profile.

Additionally, the models developed in this proposed project will be of use to other projects including concurrent NIHR projects. For example, we provide geographic modelling of stroke organisation to a range of NHS organisations and for the NIHR OptImIST project (NIHR202361). We use geographic modelling extensively [7, 17-19], modelling differences in likely outcome with different models of service provision. Up to now, however, these models have been limited to predicting outcomes at population-level based on clinical trial data, and do not include any pathway analysis/modelling of variation in provision of thrombolysis and thrombectomy beyond estimated travel times to units.

Ethics

For the modelling and explainable machine learning work we are using anonymised data only. But as we may wish to make generalised conclusions, we will apply for proportional ethics for the modelling work. For the

qualitative / implementation work, we will obtain HRA approval and, where applicable, local NHS Trust governance approvals.

Project management

The project will have project meetings every 2 months. There will be an external advisory group that meets three times during the project. Prof Martin James will take responsibility for clinical leadership of the project, and Dr Michael Allen will take responsibility for technical leadership. Dr Jason Scott will lead the qualitative / implementation work. Lauren Asare will lead PPI (though the meetings will be chaired by Leon Farmer, a stroke survivor).

Project / research expertise

Prof Martin James is an experienced stroke consultant, and is clinical lead of the national stroke audit. Martin has broad research experience, and is currently the clinical lead for the NIHR SAMueL-2 project (NIHR134326) applying explainable machine learning and simulation to thrombolysis alone. Prof. James will be the clinical lead, and co-PI, for the project, ensuring that the project maintains the correct clinical focus.
Dr Michael Allen is an experienced researcher in clinical pathway simulation, explainable machine learning, and geographic analysis in stroke. Dr Allen is the technical lead and co-PI, for the NIHR SAMueL-2 project. Dr Allen leads the modelling team that has worked on multiple NIHR stroke projects, and also commissioned stroke service analysis work for NHS England, NHS Scotland, NHS Wales, and Health and Social Care in Northern Ireland.
Kerry Pearn works with Michael Allen, and has been the driving force in developing the explainable machine learning models to predict thrombolysis use and outcome models developed in the SAMueL-2 project.
Dr Anna Laws will support Michael Allen and Kerry Pearn, and has developed the mathematical models of stroke outcome after thrombolysis and thrombectomy (based on clinical meta-analysis), and has developed the current web app for the SAMueL project.
Prof Richard Everson is professor of machine learning at the University of Exeter, and will continue to provide expert advice to the project team, and critical analysis of methods and results. In particular he will help the team strengthen uncertainty analysis of the models for Stroke-Twin.
Prof Peter McMeekin is an experienced health economist, with particular experience of working in stroke, and the health economics of thrombolysis and thrombectomy.
Dr Jason Scott is Associate Professor of Health and Social Care Quality and a Chartered Psychologist. He has completed ,or is involved in, a range of research projects to develop and evaluate innovative healthcare, social care and community interventions, specifically using co-production informed by implementation science. Jason will lead the qualitative/implementation aspects of the project.
Leon Farmer is a stroke survivor and will chair the PPI group, as he has been doing for the NIHR SAMueL project. This PPI group has been important in keeping research centred first and foremost on patient outcomes.
Lauren Asare of the Exeter PPI team will provide support for Leon, with pre-meets and debriefs, and support for administrative tasks such as scheduling and booking meetings.

Success criteria and barriers to proposed work

The ultimate success of the project is improved outcomes through better use of thrombolysis and thrombectomy. In practice, measuring thrombolysis use (potentially broken down by patient subgroups) would be a measure of impact, along with qualitative research supporting that our work influences these outcomes. Additionally adoption of the modelling work by the National Stroke Audit and by TASC would provide evidence of adoption by key QI bodies.

Barriers to adoption are discussed above. Additionally there are risks of failure to recruit research assistants (anticipated to be small, and mitigated by having a pool of researchers from which to draw on and an ability to attract and retain talent within the university). There is also the risk of failure to recruit participants, though this is also anticipated to be small as both modelling and qualitative teams have performed similar work in stroke

before and have met recruitment targets. We have also built in lead-in time for recruitment for qualitative interviews and workshops.

What further funding or support will be required if this research is successful (e.g. From NIHR, other Government departments, charity or industry)?

The project plan is to hand over all key work to the national stroke audit to continue running. NIHR PenARC has provided assurance of support for as long as necessary for hand-over. The project code is designed with hand-over in mind (simplify production code as much as reasonably possible, clear code structure, good documentation).

Dissemination

In addition to working with the national stroke audit and stroke teams directly (see above), we will publish our work in leading stroke journals, and present it at stroke and health services research conferences. All of our detailed work is published online (e.g. see https://bit.ly/explainable-ml).

The project is well connected to key avenues of impact, both 'on the ground' and through policy:

- The project is working in collaboration with SSNAP. Key outputs of the project will be incorporated into routine reports generated by SSNAP.
- The project is working in collaboration with TASC (a NHS England sponsored project, being run by NHS-Elect, for improving the use of thrombolysis). The policy lead of TASC David Hargroves (who is also the GIRFT National clinical lead for stroke) has agreed to be a member of the External Advisory Group for the project. The Exeter modelling team are providing analysis for stroke teams involved in the TASC project.
- The project is supported by the national clinical director of stroke (see attached letter), who will be kept abreast of findings.
- Our team has also worked with national and regional bodies on organisation and improvement of stroke services (including national bodies, NHS England, NHS Scotland, NHS Wales, Northern Ireland HSC, and regional bodies of East of England, Sussex, South West Peninsula, London). We expect to continue to work with national and regional bodies, and the output of this project will be of key value to this continued work).
- The web app will undergo changes as a result of usability testing, with supporting implementation resources co-produced, helping to ensure its readiness for broader theory-informed implementation.

Key expected project academic outputs - planned paper titles

1. How does use of thrombolysis and variation in thrombolysis practice affect inpatient length of stay? An explainable machine learning study.
2. How does use of thrombectomy and variation in thrombectomy practice affect inpatient lengths of stay and patient outcomes? An explainable machine learning study.
3. How would optimising the combined thrombolysis and thrombectomy pathway affect patient outcomes and inpatient length of stay? An explainable machine learning and clinical pathway simulation study.
4. Prospective understanding of implementing an explainable machine learning web application to drive reduce variation in stroke care.
5. Usability testing of an explainable machine learning web app for understanding variation in thrombolysis and thrombectomy.
6. Project overview paper

Letter of support from NHS England National Clinical Director for Stroke

See also collaborator letter of support from the Managing Director of NHS Elect, uploaded separately.

Dr Deborah Lowe
National Clinical Director for Stroke, NHS England
Deborah.lowe4@nhs.net

Professor Martin James
SW Peninsula Applied Research Collaboration
Academic Dept. of Healthcare for Older People
Royal Devon & Exeter Hospital
Barrack Road
EXETER EX2 5DW
By email to martinjames@nhs.net

14 January 2024

Dear Martin

NIHR161438 - STROKE-TWIN: Optimising flow and patient outcomes in the emergency stroke pathway using Stroke System Digital Twins

Many thanks for sharing with me your latest proposal for the NIHR Health Services and Delivery Research programme 2023 researcher-led project round. I have been very encouraged by our current collaboration using the findings of your existing NIHR HSDR-funded SAMueL-2 work, and I am happy to confirm that your new research proposal to develop your methods further in support of wider improvements in reperfusion and hyperacute stroke care, using the powerful tools of prospective national audit data from SSNAP and the further development of your machine learning methods including Digital Twins, offers the prospect of contributing significantly to reductions in stroke-related disability and the planning of the most effective NHS provision. It is thus of considerable relevance and interest to the priorities of the NHS today especially the NHS England Long Term Plan, and I am very happy to lend my support to your Stage 2 application.

I wish you every success with your research.

Kind regards

Deborah Lowe

Dr Deb Lowe
National Clinical Director for Stroke, NHS England