Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

SIMULATION OF STROKE CARE SYSTEMS

Thomas Monks

Faculty of Health Sciences
University of Southampton
Southampton, SO17 1BJ, UNITED KINGDOM

Kerry Pearn
Michael Allen

University of Exeter Medical School
University of Exeter
Exeter, EX1 2LU, UNITED KINGDOM

ABSTRACT

Stroke is major cause of disability internationally, the leading cause of disability in England, and the third most common cause of death worldwide. The good news is that there is growing evidence that simulation modeling can play an important role in understanding and designing improvements in acute stroke systems in order to reduce this disability burden. This paper presents an overview of simulation methodology to tackle logistical and capacity planning problems in stroke. Four contributions are made to accelerate studies in this area. First, a grounding in the basic processes and operational issues that occur in stroke pathways is given. Second, modeling approaches for single and multiple hospitals in emergency and rehabilitation settings are described along with guidance on selection of performance measures. Third, common data issues are highlighted. Last, a range of model simplifications are presented to mitigate potential data and complexity issues that are inherent to stroke systems.

1 INTRODUCTION

Stroke is major cause of disability internationally, the leading cause of disability in England, and the third most common cause of death worldwide. In the United Kingdom there are over 150,000 strokes each year with around 1.2m stroke survivors, many of whom live with post-stroke disability burdens . Total societal costs of stroke in the United Kingdom are estimated at £9 billion (€12.5; $13.7b) annually (Saka et al. 2009). Direct costs, treatment and long term care, represent £4.5 billion; or ~4.5% of the National Health Service's (NHS) budget. Indirect costs, from benefits payments and informal care represent £2.5b (Saka et al. 2009). To reduce the substantial human and economic burdens of surviving stroke healthcare systems must be highly responsive to the emergency an acute stroke represents (Monks et al. 2014) and also provide sufficient treatment capacity for lengthy rehabilitation. The good news is that there is growing evidence that simulation modeling, in particular discrete-event simulation (DES), can play an important role in understanding and designing improvements in acute stroke systems in order to reduce this disability burden (Bayer et al. 2010; Churilov and Donnan 2012; Churilov et al. 2013; Cordeaux et al. 2011; Lahr et al. 2013a, b; McClean et al. 2011; Monks et al. 2015; Monks et al. 2012; Monks et al. 2014; Pitt et al. 2012).

This paper presents an overview of simulation methodology to tackle logistical and capacity planning problems in stroke. Four contributions are made to accelerate studies in this area. First, a grounding in the basic processes and operational issues that occur in stroke pathways is provided. Second, modeling approaches for single and multiple hospitals in emergency and rehabilitation settings are described along with guidance on selection of performance measures. Third, common data issues are highlighted. Last, a range of model simplifications are presented to mitigate potential data and complexity issues that are inherent to stroke systems.

2 STROKE CARE SYSTEMS

Stroke is categorized into two types: ischaemic and haemorrhagic. The majority of strokes are the former (~85%) and occur when blood flow to part of the brain is interrupted, due to blockage of an artery by a thrombus (blood clot) while the latter refers to a bleed within the brain. In the community stroke can often be recognized using a simple diagnostic test based on the mnemonic FAST: Facial drooping, Arm weakness and slurring of Speech are all symptoms of stroke; while the 'T' refers simply to the time critical nature of the treatment for stroke.

We provide a simplified high level overview of a stroke pathway in Figure 1. In terms of hospital management these pathways can be thought of having three distinct phases: the hyper-acute (emergency), the acute and the rehabilitation. The pathway is typically initiated by someone other than the patient who witnesses the onset of stroke or finds a patient with suspected stroke.

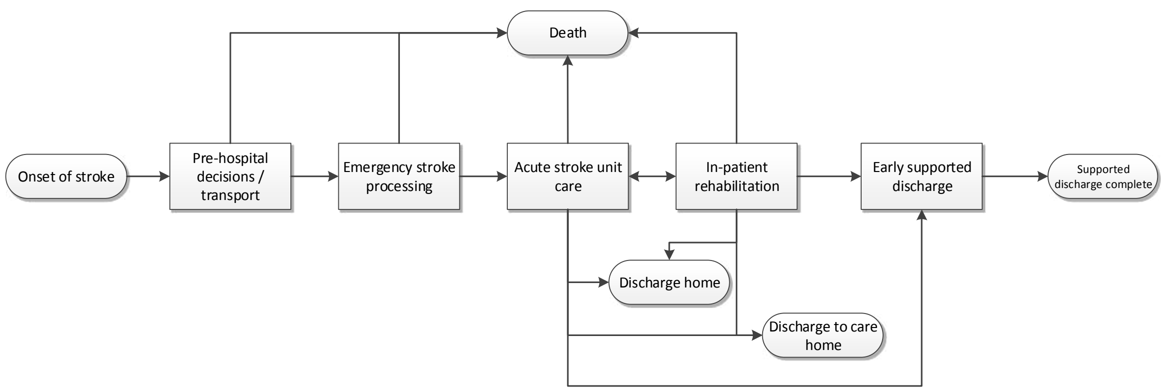

Figure 1: Overview of a stroke pathway.

2.1 Hyper-Acute Stroke Pathways

The *hyper*-acute pathway is largely focused on responsive identification and treatment of a select group of stroke patients suffering a ischaemic stroke. In 1995 the National Institute of Neurological Disorders and Stroke (NINDS) trial provided evidence that early thrombolysis with recombinant tissue plasminogen activator (rtPA), generic name alteplase (brand name Activase), up to three hours from onset of symptoms was effective at reducing post stroke disability measured at 90 days. Put simply rtPA dissolves the blood clot and restores blood flow to the affected region of the patient's brain. It remains the most effective known treatment for acute ischaemic stroke. Interventions with rtPA must be done soon after onset of symptoms, as quite literally time is brain: each minute saved adds over a day of extra healthy life on average (Meretoja et al. 2014).

In the last two decades, international uptake of stroke thrombolysis has in general been poor. Most centres treat around 5% of all acute strokes and are given modest performance targets: for example, the English Department of Health's National Stroke Strategy has a target of 10%. This low performance is in part because in the short time available a sequence of events must occur that are all vulnerable to delays. A patient or witness must contact emergency services, travel to hospital, be assessed and then be processed in potentially very busy emergency and radiology departments. Figure 2 provides a more detailed overview of a hyper-acute pathway. Critical parts of the pathway are often considered to be the speed at which emergency services are contacted, where delays may be substantial (potentially due to first attending a general practitioner (GP) clinic or being transported to hospital by a carer/witness instead of emergency services) and the need for a CT scan to rule out haemorrhagic stroke.

Given the tight time constraints, research has focused on methods to extend the eligibility criteria for thrombolysis. The most notable results were from the European Cooperative Acute Stroke Study III

(ECASS-III) and the third international stroke trial (IST-3). ECASS-3 provided evidence to extend the treatment window from three to four and a half hours, results later confirmed in a pooled individual patient meta-analysis of all the positive and negative trials of alteplase, while IST-3 demonstrated clear benefit in patients over the age of 80. These trials along with numerous international thrombolysis databases confirmed a time dependent benefit of thrombolysis. This is to such an extent that the number of patients needed to be treated to achieve one additional patient with no disability effectively doubles with each 90 minute interval from onset (Emberson et al. 2014; Lees et al. 2010). This further emphasises that time is of the essence in every step of acute stroke care.

Although there are inherent difficulties in successful implementation of hyperacute pathways, there is clear evidence that highly responsive health systems can be implemented and treat large numbers of ischaemic stroke patients. This is typified by the London hyperacute centralisation (Morris et al. 2014) and Helsinki's twelve year programme of process improvement (Meretoja et al. 2012).

In 2008 London underwent a major reconfiguration of stroke services. The care that was provided by over 30 hospitals was centralised to eight hyperacute stroke units (HASUs). These HASUs have proved to be highly successful and the latest published figures showing thrombolysis rates of between 17 and 21% with a hospital arrival to treatment time (ATT) of between 29 and 48 minutes.

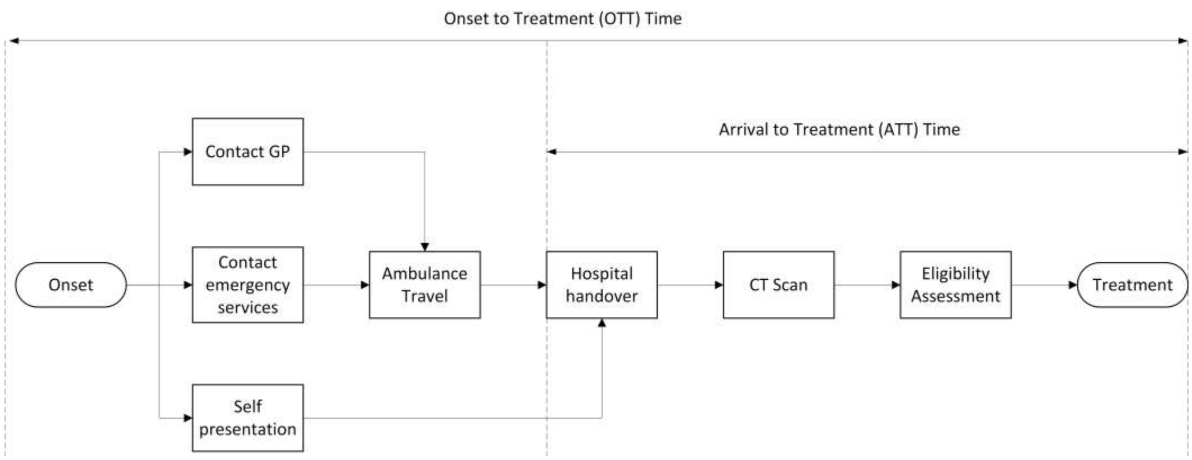

Figure 2: High level overview of a hyperacute stroke pathway.

2.2 Acute Care and Rehabilitation

Once a patient has either been ruled out or ruled in for thrombolysis they are admitted to an acute stroke unit (ASU). An ASU is a dedicated ward for patients with suspected stroke comprising of a multi-disciplinary team of specialist physicians and nurses. Acute care often involves an assessment of the patient, identification of the type of stroke suffered and in some cases early rehabilitation. Length of stay on an ASU is typically seven days for those with a final diagnosis of stroke. There are some caveats, however, as admission to an ASU might be severely delayed due to lack of beds. This will be for a number of reasons: the natural variation in stroke admissions and treatment duration, the inevitable overspill of patients other wards in a highly utilized hospital and transfer delays of patients to rehabilitation. Where direct admission is not possible stroke patients become outliers and are admitted to another hospital ward. Unless the hospitals stroke consultants visit their outlier patients, these patients will usually not have further access to a stroke specialist and will be seen by the resident, non-stroke, specialist on their ward.

In the UK acute care is often based in a large acute hospital while rehabilitation capacity is split over a number a of smaller community based hospitals. Length of stay in rehabilitation is substantially longer than an acute stay (average ~30-40 days;). Patients receive a number of treatments to aid rehabilitation from their stroke in this time including physical and speech therapy. Discharge from rehabilitation can be substantially delayed as patients may require special care packages from social services (a separate organization from the NHS) to be in place before they go home; a place in a care home or simply because there is a family dispute about how the patient will be cared for at home. In a minority of cases a patient's condition may deteriorate so they 'bounce-back' from rehabilitation to acute care units.

The difficulty in discharge from rehabilitation wards can lead to so called *bed-blocking* or in more familiar terms queueing; meaning operational difficulty for any upstream ASU. In the UK the performance of stroke services is measured by the proportion of stroke patients admitted to the stroke unit within four hours of hospital arrival and the proportion of stroke patients that spend 90% of their hospital stay on a stroke unit (a surrogate for good patient care), with large financial penalties for underperforming services.

Both acute and rehabilitation wards make use of early supported discharge (ESD) services which can be made up of both therapists and nursing staff. Given the available evidence of its effectiveness, ESD services often work with mild to moderate stroke severity patients leading to an average reduction in hospital length of stay of three days (Fearon and Langhorne 2012).

3 SIMULATION OF SINGLE TREATMENT PATHWAYS

3.1 Hyper-Acute Pathway Models

3.1.1 Objectives

There are now several published simulation studies analysing hyper-acute pathways and patient treatment with rtPA (e.g. Churilov et al. 2013; Lahr et al. 2013a; Monks et al. 2012). These studies aim to evaluate competing pre-hospital or in-hospital processes and their impact on treatment rates and post stroke disability. Depending on the focus of the study, such models include more or less detail on the pre-hospital ambulance protocols, such as the use of a 'scoop and run' protocol where the patient is transferred to the acute hospital with minimal pre-hospital treatment (Lahr et al. 2013a), or in-hospital reactions to the arrival of a suspected stroke patient, such as emergency department procedures and core and non-core workforce hour processes (Monks et al. 2012).

3.1.2 Choice of Performance Measures

The four main performance measure categories are summarized in Table 1. The key contribution of DES is to provide a stochastic framework for analyzing the likely treatment volume, speed and workload of the departments involved. Treatment volume might be presented as an average rate (total stroke treated with rtPA / total no. of discharge strokes) or, to exploit the stochastic results, as a histogram of the likely range that treatment rates may fall into, potentially enhanced as a Measure of Error and Risk (MORE) plot (Nelson 2008).

The results from a DES model can then be used as input parameters to either clinical models of population benefit or a health economic model to understand the longer term cost-effectiveness. Several simulation authors (Lahr et al. 2013a; Monks et al. 2012; Pitt et al. 2012) have opted to use the modified Ranking Scale (mRS) scores at 90 days treatment. The mRS score is a non-linear scale ranging from 0 (no disability) to 6 (death). Published meta-analysis of all of the thrombolysis clinical trials (Emberson et al. 2014; Lees et al. 2010) provide a base and time dependent odds ratios that can be used to convert the onset-to-treatment (OTT) times of simulated patients into estimates of patients with an mRS score of 0 or 1 *due to treatment* at 90 days. An alternative approach to odds ratios is to make use of the number needed to treat (NNT) statistics from 0-90, 91-180 and 180-270 minute OTT groups (4.5; 9.0 and 14.1 patients respectively). In these cases the simulation model only needs to count the number of patients in treated by OTT group and divide through by the respective NNT. More advanced outcomes might include long term costs of stroke and the impact of increased number of patients treated with rtPA along with gains in quality of life (c.f. Penaloza-Ramos et al. 2014).

Table 1: Hyeracute pathways Performance measure categories.

Performance measure category	**Description**
Treatment rates and speed	The time between onset of a stroke and treatment with rtPA has a clear relationship with post-stroke disability at 90 days. Treatment rates measured as percentage of all stroke treated with rtPA or percentage of ischemic stroke treated with rtPA.
Population disability benefit	Measured by a 90 day modified Rankin Scale (mRS) score of 0 (no disability) or 1 (minimal functional deficit). Some patients will always have a good outcome regardless of treatment. This measures identifies the additional patients due to treatment.
Workforce implications	E.g. the no. prioritized scans or urgent calls outs of stroke nurses and physicians (important in the case of suspected stroke)
Health economic outcomes	Cost per quality adjusted life year gained

3.1.3 Common Data Issues

There are common data issues when modeling in single-treatment pathways. When modeling pre-hospital processes in detail an important variable is the time taken from the onset of stroke to a call being received by emergency services (onset-to-call; OTC). OTC is not a mandatory field in any UK national dataset for stroke (in the UK this is the Sentinel Stroke National Audit Programme – SSNAP) and if collected locally may be biased to patients who were treated (a highly optimistic distribution). There are however, now some published data from simulation studies that might be adapted. Lahr et al. (2013a) provide details of a continuous empirical distribution for OTC (range 0 –2880 mins). Churilov et al. (2013) analyse the pre-hospital phase in detail and give a baseline proportion of OTC as ~20%. Carroll et al. (2004) report the median time from onset of symptoms to seeking medical help as 30 minutes.

In the UK, the process within an ED is often (but we emphasise not always) broken down into discrete time stamps within a local IT system. Strokes can be identified by diagnosis at discharge (International Classification of Diseases version 10 codes, or ICD-10, I60-I69). Data often include time that a CT scan is requested, however, data from radiology on the time of a report is less likely. Where this has been collected separately it has been found to be highly variable and may depend if a stroke physician is present at the scan or if a radiologist has been informed that a patient may be thrombolysis eligible.

Data is often most problematic outside of core working hours. In particular, treatment speed depends on the point at which on-call physicians are contacted, the (often unclear) speed of any tele-radiology used and the time that it takes on-call physicians (and possibly radiographers) to travel hospital. Unless an analyst gets very lucky, much of this data must be estimated by so-called expert opinion, small sample observation and use of utility distributions such as the triangular and beta distributions (useful for approximating travel times of on-call resources).

Perhaps the most important data issue is that of the number of stroke mimics. These are patients seen in ED or admitted to an ASU with suspected stroke, but for whom the final diagnosis is not stroke. For example, up to ~40% of seizures and complex migraines (e.g. with vision loss or aphasia) are initially diagnosed as stroke (Nau et al. 2010). As such changes in stoke pathways might divert mimics, whom before the change were effectively identified within ED, to specialist stroke services. Understanding these consequences is important in decision making and resource planning. Unfortunately, as many of mimics are screened out these data are usually not available locally or at best a lower bound. One option is to resort to published proportional 'over-diagnosis' of stroke ranging approximately between 20 and 30% (Yew and Cheng 2009). These figures could be used to proportionally inflate the mean number of strokes arriving by time of day. An additional measure is to analyse ambulance service data on suspected stroke patients and compare to a hospital dataset of discharged strokes. Such an analysis will not capture all suspected strokes, as some will 'self –present' and others may not have been 'diagnosed' by ambulance services.

3.1.4 Simplifications

Stroke patients are often processed within an emergency department (ED). The queueing time of stroke patients can be accurately modelled with a detailed ED model; however, even in large hospitals stroke and suspected stroke numbers are relatively low compared to general emergency patients (stroke patients account for about 1 in every 200 arrivals in the A&E department, and about 1 in 65 emergency admissions into the hospital). As such it often sensible to simplify ED queuing times to random variables as this greatly reduces model runtime while still providing results of sufficient accuracy.

Section 2.1 describes the pre-hospital process. It is only necessary to model this in detail if the objectives of the study focus on the pre-hospital phase. One disadvantage of opting for conceptualising this as a random variable is if at a later date one wishes to investigate pre-hospital processes then the model will need to be modified and re-verified and validated. Monks et al (2012) got around this problem by splitting the simulation into pre-hospital and in-hospital models. The in-hospital model Monks and colleagues describe was later successfully reused with a more detailed pre-hospital model studying ambulance time at scene and the use of rapid response units.

One complication in the time sensitive treatment of stroke is that treatment speed is correlated with time remaining (Pitt et al. 2012). This *deadline effect* is relatively straightforward to model at the ATT level (see Pitt et al, 2012 for an example), but it is highly unlikely that sufficient data will be available to model at the process level. As such models often do not include it, but acknowledge it as a limitation.

As stroke treatment simulations are just simplified to focus on stroke patients a particular challenge is balancing their needs against other patients groups using the same pathways, for example, the ED or CT scanners. Indeed an system optimised for stroke might unintentionally affect other patients care. Given this simplification, some informal analysis should be conducted to assess if changes disadvantage other patient groups.

3.2 Acute Care and Rehabilitation Models

3.2.1 Objectives

Simulation studies of acute and rehabilitation services are typically focussed on capacity planning and are best tackled using discrete-event simulation. For example, what number of beds or therapists are required in order to minimise or maximise a measure of stroke pathway performance. More specifically study objectives will be to minimize admission delays to an acute stroke unit and/or transfer delays to a rehabilitation service in the most cost-effective manner. Decision variables can include number of resources, physical layout changes such as co-locating and pooling of acute and rehabilitation beds, and ESD policies such as extending services to patients with more severe strokes. Decision makers may also be interested in expected reductions in length of stay in rehabilitation services and their impact on required capacity.

3.2.2 Choice of Performance Measures

In the UK, the performance of stroke pathways is monitored by standard performance measures which can incorporated into DES models. The accepted surrogate for good stroke care is the so called 90% stay measure. This requires that 80% of all stroke patients spend 90% of the acute and super (acute + rehabilitation) stay on a stroke ward. Common reasons for not spending a full stay on stroke wards are delays in ED or a medical assessment unit, and insufficient capacity in the stroke wards. The latter means that some patients are admitted to alternative wards and become 'outliers'. In reality the 90% stay measure is also problematic because of good ESD services. If, for whatever reason, a patient has been substantially delayed in admission then timely discharge to ESD services means that the patient is likely to miss their 90% stay target.

Another NHS performance target useful for capacity planning is the proportion of stroke patients that are admitted the acute stroke unit within 4 hours of their arrival. More generally this might focus on average waiting time for an acute admission (which is linked to waiting time for transfer to one or more rehabilitation wards/ESD).

3.2.3 Common Data Issues

Perhaps the most common data deficiency encountered when building simulation models for capacity planning in stroke is the lack of a time-stamp to indicate when a patient is 'ready to be discharged' (here used to describe the time when a patient is both medically fit and has a discharge plan in place). This means that routine data often reflect total length of stay as opposed to 'treatment time' and 'queuing time'. The simplest way to mitigate this issue and estimate the distribution and parameters of treatment time is to use expert opinion. For example, by asking 'what percentage of length of stay is spent waiting for transfer'. If this option is taken modellers need to be very specific about which patients they are referencing and should also make extensive use of sensitivity analysis. The second option is to use model calibration techniques. That is, if a sample of the performance measure of interest is available, the input parameters are systematically varied until outputs agree with empirical observations. These approaches lead to multiple sets of input parameters that each provide matching output. Results should then be presented for each of these sets to adequately represent the uncertainty in the inputs.

A second problem is that stroke patients do not queue when an acute bed is unavailable, they are instead admitted to an alternative ward. There is often insufficient data to identify if these patients are repatriated, for example, first in first out, if decision rules are place to reject repatriation if only a short amount of treatment time remains, or indeed if patients resident on the ASU who are waiting to be transferred to rehabilitation are moved to an alternative ward to make space. Of course, some of this can be handled by assumptions agreed with simulation clients. However, in the next section we discuss simplifications to help specifically with this issue.

3.2.4 Simplifications

Although there may be some temptation to model both the hyper-acute, acute and rehabilitation phases of the pathway altogether it is often more sensible to model hyper-acute pathways separately as their objectives are more focused on rtPA treatment rates. If thrombolysis rates are important in capacity planning decisions then modellers should consider incorporating scenarios exploring alternative rtPA treatment rates.

It should be apparent from our introduction to stroke services, that simulation models of stroke pathways are typically bounded at discharge from rehabilitation. In reality patients length of stay in rehabilitation may be affected by delays in social services. Instead of modeling social services in detail, such delays in discharge from rehabilitation could be incorporated as a random variable (this would only apply to a proportion of the total patient population). Alternatively in studies where reductions in discharge delays are not relevant rehabilitation and discharge delays could be incorporated as a single random variable.

We noted that a common data issue is a paucity of data describing management of outlying stroke patients. Instead of attempting to model this management in detail, capacity constraints can be removed from a model (an infinite server approach). Infinite server simulation models predict the probability that a patient will be rejected (blocked) from a ward and eliminates the need to understand management of patients when capacity is insufficient. Such models are different from the real system, but are still a valuable decision aid for capacity planning.

4 SIMULATION OF MULTI-CENTRE TREATMENT PATHWAYS

4.1 Multi-Centre Objectives

In contrast to single treatment centre models that focus on competing treatment processes, multi-centre treatment pathway models focus on a facility location problem with stochastic demand. Such problems are particularly resonant in the UK, given NHS England's push for larger centralised HASUs. The decision variables in such simulation projects are therefore the number of facilities needed, their location, policies regarding pre-hospital transportation, policies for out of hours workforce and in-hospital treatment speed. An example of an alternative pre-hospital transportation policy is to make use of smaller HASU facilities during normal Monday to Friday working ours (while the appropriate workforce is available) and larger units, possibly located further away, outside of core workforce hours. Another example might be the use of a 'drip and ship' model where stroke patients are taken to their closest hospital for thrombolysis and then transferred to a larger 'comprehensive stroke unit' for longer term care (Qureshi et al. 2012).

The size of these discrete facility location problems might vary from a local three site system to reconfiguration up to a 30 site change as was seen in London or even a national model of all centres in a country. Central to this location problem is the tension between pre-hospital travel time and in-hospital arrival to treatment time (ATT). Fewer units mean longer average travel times to the closest HASU, but might mean a greatly reduced average ATT (along with a shorter tailed ATT distribution) due to higher treatment volumes, streamlined processes and funding to appropriately resource out of hours workforces. Centralising stroke services will increase the workload burden of the ambulance service by increasing the pre-hospital travel times and also requiring their services to repatriate patients from the HASU to their local hospital stroke ward for rehabilitation after the acute phase. In large problems, for tractability, it is unlikely that models will incorporate the detailed in-hospital processes found in single treatment pathway models. Instead models may have assumptions about speed improvements following centralisation or instead be used to conduct a *threshold analysis* in order to assess the feasibility of ATT speeds required to exceed current treatment rates.

An important contribution of simulating multi-centre treatment pathways is the ability to evaluate both the patients that gain from centralisation, in terms of being treated and treatment speed, to smaller number of HASUs and those patients that lose out (generally those living further away from the high performing HASU). This 'winners and losers' analysis is of most significance when the geographic region simulated contains rural areas.

4.2 Choice of Performance Measures

Decision making concerning reconfiguration of multi-centre stroke pathways is extremely difficult, especially as changes may shift the benefits of time sensitive treatment to different geographic locations. As such, simulation models may make use of a variety of performance measures to quantify trade-offs between decision variables each of which might be assigned a different decision 'weight'. Similar to single-centre simulation models, a model may predict OTT and clinical outcomes, but in addition may output measures indicating average number of attendances at a HASU, the probability different HASUs hitting some feasibility threshold, such as 1000 admissions per year or maximum OTT.

4.3 Common Data Issues

Multi-treatment centre problems are at their heart a facility location problem analysing the trade-off between improved in-hospital treatment times and change in pre-hospital treatment times. As such models require data on the location of a demand (onset of stroke). In some limited cases actual site of demand might be available (from an ambulance service); however it is more likely that patient records will contain the patients home location (for example a UK postcode or super output area) which will be highly correlated with the location of an ambulance is dispatched to pick-up a suspected stroke patient, but may only explain 80-90% of events due to those strokes occurring away from the patients home location.

Random variables to represent travel times from demand nodes to a HASU can be constructed from geographic information software, but will require some adjustment to correctly mimic ambulance 'blue light' travel times. Adjustment may also be needed for transient affects across the day (to mimic rush hour traffic), weekdays versus weekends or seasonal periods simulated. This is possible if a sample of ambulance travel time data is available; however, sample size will quickly diminish as more subgroups are investigated.

4.4 Simplifications

The main simplification already mentioned is that in-hospital processes are simplified to one random variable, namely ATT. If a decision is made to centralise stroke care this could then be followed up by a number of single-treatment centre studies aiming to optimize in-hospital processes for thrombolysis.

4.5 Additional Complexities

Consider the reconfiguration of healthcare system consisting of 10 treatment centres that will be reduced to seven HASUs. To fully enumerate the solution requires full simulation of 252 competing alternatives. If this were 15 sites reduced to eight HASUs full enumeration is 6435 alternatives. To make this problem tractable it is sensible to apply Ranking and Selection (R&S) methods such as Kim-Nelson (Kim and Nelson 2007). When taken to a national scale the challenge becomes one beyond full computation: if 50 hyper-acute stroke units were to be selected from 160 national acute hospitals then 10^{42} possible solutions exist. In these strategic planning exercises, heuristic methods such as greedy or genetic algorithms are required to find good solutions in this vast solution space.

One limitation is that R&S tends to be a focus on a single performance measure and as seen above there may be several of interest in multi-centre problems. One solution is to select a single measure as the most important and then analyse the best 5-10 alternatives following R&S. Alternatively a composite score, a product of different performance measures, may be used for optimisation. Again multiple solutions may be presented to stakeholders, as many solutions may have a similar score and stakeholders may have good reasons, outside of the scoring system, for preferring one solution (which may not be quite the best solution identified based on the scoring system).An important consideration is the weighting applied to each of these performance measures when using it to determine the optimal configuration. Different score definitions will produce a different optimal solution.

Modeling also does not address ethical questions that may be posed during such an analysis. Is it acceptable, for example, to take a utilitarian approach which seeks to maximise the number of patients benefiting from the healthcare system even if the solution adopted is known to disadvantage one particular group of patients.

5 CONCLUSIONS

This paper presents methodology for modeling simulation of stroke care systems. Its overall aim is to provide modellers key insights to conceptualise their models and anticipate some of the common difficulties faced. We provide an overview of three important areas in stroke care systems: hyper-acute single-treatment pathways, acute and rehabilitation services and multi-centre treatment systems. An aspect of stroke care we do not explicitly cover are systems to manage patients suffering a transient ischemic attack (TIA) sometimes referred to as a mini-stroke. The risk of suffering a major stroke within a week of a TIA is substantial. As such, provision of a rapid assessment and treatment service for suspected TIA, by admission to hospital or by outpatient clinic, is essential to manage this risk. As in acute stroke, simulation offers decision makers a powerful approach to analyze and improve the organization of TIA services and is just as important.

We also do not cover future developments of acute stroke care. Recent trial results (Fransen et al. 2014) are showing promise of augmenting medical thrombolysis with mechanical thrombectomy (where the clot is retrieved by a medical device). Although more clinical evidence is needed in this area, modeling again offers the potential to analyze implementation strategies.

All methodology presented makes use of a DES framework and as such focuses on the micro-level stochastic operational aspects of stroke care. Future research may consider the macro-level issues of stroke care and systems. High level system dynamics models could incorporate both patient cohorts with risk factors for major stroke, patients within hospital environments, and patients living in the community with the consequences of stroke. Such models might, for example, consider improved diagnosis of atrial fibrillation, a heart condition that increases the risk of a severe stroke, or management of TIA patients and predict the resulting impact on both acute and community services.

Stroke is a leading cause of adult disability worldwide. Simulation of stroke care systems offers both insight and an approach to drive improvement in this important area of healthcare.

ACKNOWLEDGMENTS

This article presents independent research funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Wessex. KP and MA are funded by NIHR CLAHRC South West Peninsula. The views expressed in this publication are those of the author(s) and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

REFERENCES

Bayer, S., C. Petsoulas, B. Cox, A. Honeyman, and J. Barlow. 2010. "Facilitating Stroke Care Planning through Simulation Modelling." *Health Informatics Journal* 16 (2):129-143.

Carroll, C., J. Hobart, C. Fox, L. Teare, and J. Gibson. 2004. "Stroke in Devon: Knowledge Was Good, but Action Was Poor." *Journal of Neurology Neurosurgery and Psychiatry* 75 (4):567-571.

Churilov, L., and G. A. Donnan. 2012. "Operations Research for Stroke Care Systems: An Opportunity for the Science of Better to Do Much Better." *Operations Research for Health Care* 1 (1):6-15.

Churilov, L., A. Fridriksdottir, M. Keshtkaran, I. Mosley, A. Flitman, and H. M. Dewey. 2013. "Decision Support in Pre-Hospital Stroke Care Operations: A Case of Using Simulation to Improve Eligibility of Acute Stroke Patients for Thrombolysis Treatment." *Computers & Operations Research* 40 (9):2208-2218.

Cordeaux, C., A. Hughes, and M. Elder. 2011. "Simulating the Impact of Change: Implementing Best Practice in Stroke Care." *London Journal of Primacy Care* 4:33-37.

Emberson, J., K. R. Lees, P. Lyden, L. Blackwell, G. Albers, E. Bluhmki, T. Brott, G. Cohen, S. Davis, G. Donnan, J. Grotta, G. Howard, M. Kaste, M. Koga, R. von Kummer, M. Lansberg, R. I. Lindley, G. Murray, J. M. Olivot, M. Parsons, B. Tilley, D. Toni, K. Toyoda, N. Wahlgren, J. Wardlaw, W. Whiteley, G. J. del Zoppo, C. Baigent, P. Sandercock, and W. Hacke. 2014. "Effect of Treatment Delay, Age, and Stroke Severity on the Effects of Intravenous Thrombolysis with Alteplase for Acute Ischaemic Stroke: A Meta-Analysis of Individual Patient Data from Randomised Trials." *The Lancet*.

Fearon, P., and P. Langhorne. 2012. "Services for Reducing Duration of Hospital Care for Acute Stroke Patients." *Cochrane Database Syst Rev* 9:Cd000443.

Fransen, P. S., D. Beumer, O. A. Berkhemer, L. A. van den Berg, H. Lingsma, A. van der Lugt, W. H. van Zwam, R. J. van Oostenbrugge, Y. B. Roos, C. B. Majoie, and D. W. Dippel. 2014. "Mr Clean, a Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in the Netherlands: Study Protocol for a Randomized Controlled Trial." *Trials* 15:343.

Kim, S.-H., and B. L. Nelson. 2007. "Recent Advances in Ranking and Selection." In *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*, edited by 162-172. 1351576: IEEE Press.

Lahr, M. M. H., D.-J. van der Zee, G.-J. Luijckx, P. C. A. J. Vroomen, and E. Buskens. 2013a. "A Simulation-Based Approach for Improving Utilization of Thrombolysis in Acute Brain Infarction." *Medical Care* 51 (12):1101-1105 1110.1097/MLR.1100b1013e3182a1103e1505.

Lahr, M. M. H., D.-J. van der Zee, G.-J. Luijckx, P. C. A. J. Vroomen, and E. Buskens. 2013b. "Thrombolysis in Acute Ischemic Stroke: A Simulation Study to Improve Pre- and in-Hospital Delays in Community Hospitals." *PloS one*.

Lees, K. R., E. Bluhmki, R. von Kummer, T. G. Brott, D. Toni, J. C. Grotta, G. W. Albers, M. Kaste, J. R. Marler, S. A. Hamilton, B. C. Tilley, S. M. Davis, G. A. Donnan, and W. Hacke. 2010. "Time to Treatment with Intravenous Alteplase and Outcome in Stroke: An Updated Pooled Analysis of Ecass, Atlantis, Ninds, and Epithet Trials." *The Lancet* 375 (9727):1695-1703.

McClean, S., M. Barton, L. Garg, and K. Fullerton. 2011. "A Modeling Framework That Combines Markov Models and Discrete-Event Simulation for Stroke Patient Care." *ACM Transactions on Modelling and Computer Simulation* 21 (4):1-26.

Meretoja, A., M. Keshtkaran, J. L. Saver, T. Tatlisumak, M. W. Parsons, M. Kaste, S. M. Davis, G. A. Donnan, and L. Churilov. 2014. "Stroke Thrombolysis: Save a Minute, Save a Day." *Stroke*.

Meretoja, A., D. Strbian, S. Mustanoja, T. Tatlisumak, P. J. Lindsberg, and M. Kaste. 2012. "Reducing in-Hospital Delay to 20 Minutes in Stroke Thrombolysis." *Neurology* 79 (4):306-313.

Monks, T., M. Pearson, M. Pitt, K. Stein, and M. A. James. 2015. "Evaluating the Impact of a Simulation Study in Emergency Stroke Care." *Operations Research for Health Care* IN PRESS.

Monks, T., M. Pitt, K. Stein, and M. James. 2012. "Maximizing the Population Benefit from Thrombolysis in Acute Ischemic Stroke: A Modeling Study of in-Hospital Delays." *Stroke* 43 (10):2706-2711.

Monks, T., M. Pitt, K. Stein, and M. A. James. 2014. "Hyperacute Stroke Care and Nhs England's Business Plan." *BMJ* 348.

Morris, S., R. M. Hunter, A. I. G. Ramsay, R. Boaden, C. McKevitt, C. Perry, N. Pursani, A. G. Rudd, L. H. Schwamm, S. J. Turner, P. J. Tyrrell, C. D. A. Wolfe, and N. J. Fulop. 2014. "Impact of Centralising Acute Stroke Services in English Metropolitan Areas on Mortality and Length of Hospital Stay: Difference-in-Differences Analysis." *BMJ* 349.

Nau, K. C., T. Crocco, J. Biola, and H. Larrabee. 2010. "Is It Stroke, or Something Else?" *Journal of Family Practice* 59 (1):26-31.
Nelson, B. L. 2008. The More Plot: Displaying Measures of Risk & Error from Simulation Output. Proceedings of the 2008 Winter Simulation Conference, 7-10 Dec. 2008.
Penaloza-Ramos, M. C., J. P. Sheppard, S. Jowett, P. Barton, J. Mant, T. Quinn, R. M. Mellor, D. Sims, D. Sandler, and R. J. McManus. 2014. "Cost-Effectiveness of Optimizing Acute Stroke Care Services for Thrombolysis." *Stroke* 45 (2):553-562.
Pitt, M., T. Monks, P. Agarwal, D. Worthington, G. A. Ford, K. R. Lees, K. Stein, and M. A. James. 2012. "Will Delays in Treatment Jeopardize the Population Benefit from Extending the Time Window for Stroke Thrombolysis?" *Stroke* 43 (11):2992-2997.
Qureshi, A. I., S. A. Chaudhry, G. J. Rodriguez, M. F. Suri, K. Lakshminarayan, and M. A. Ezzeddine. 2012. "Outcome of the 'Drip-and-Ship' Paradigm among Patients with Acute Ischemic Stroke: Results of a Statewide Study." *Cerebrovascular diseases extra* 2 (1):1-8.
Saka, Ö., A. McGuire, and C. Wolfe. 2009. "Cost of Stroke in the United Kingdom." *Age and Ageing* 38 (1):27-32.
Yew, K. S., and E. Cheng. 2009. "Acute Stroke Diagnosis." *American Family Physician* 80 (1):33-40.

AUTHOR BIOGRAPHIES

THOMAS MONKS is a Senior Research Fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modeling. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His specialty is simulation of healthcare systems of unscheduled and emergency care. He is co-chair of the UKs Simulation Workshop 2016 and co-chair of the UKs simulation special interest group. His email address is thomas.monks@soton.ac.uk.

KERRY PEARN is an Associate Research Fellow in PenCHORD, University of Exeter Medical School. She obtained a BSc in Mathematics and Sports Science and has worked as an Operational Researcher for 14 years applying an array of heuristic and optimisation techniques to a wide range of decision making tasks, spanning systems related to the environment, healthcare and agriculture. Her e-mail address is K.Pearn@exeter.ac.uk.

MICHAEL ALLEN is a Senior Modeler in PenCHORD, University of Exeter Medical School, University of Exeter. His research focuses on the application of computer simulation modeling to improving the quality and efficiency of health systems. Projects have included whole hospital bed modeling, mental health referral systems, labour ward staffing, operating theatre throughput, surgical referral-to-treatment waiting times, community nurse travel distances, neonatal network capacity, and ambulance handover in emergency departments. His e-mail address is m.allen@exeter.ac.uk.