D2.2. Predictive modelling in Stroke

<table>
<thead>
<tr>
<th>Workpackage</th>
<th>WP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead beneficiary</td>
<td>TSA</td>
</tr>
<tr>
<td>Due date</td>
<td>31st of October 2016</td>
</tr>
<tr>
<td>Submission date</td>
<td>4th November 2016</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Dr. Gary Randall (TSA)</td>
</tr>
<tr>
<td>Abstract</td>
<td>This report describes the major predictive models found in clinical decision-making in the assessment of recurrent stroke risk. We will describe the usefulness of these models and which may be used in STARR.</td>
</tr>
</tbody>
</table>
## History of changes

<table>
<thead>
<tr>
<th>Rev. N</th>
<th>Description</th>
<th>Author(s)</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1st draft, ToC</td>
<td>Gary Randall (TSA)</td>
<td>02/08/2016</td>
</tr>
<tr>
<td>2</td>
<td>2nd draft</td>
<td>Gary Randall (TSA)</td>
<td>27/10/2016</td>
</tr>
<tr>
<td>3</td>
<td>Pre-final</td>
<td>Gary Randall (TSA)</td>
<td>4/11/2016</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Table of contents

Abbreviations / Definitions ................................................................. 3

Introduction .......................................................................................... 4

Predictive Modelling in Healthcare ....................................................... 6

Developing stroke risk models ............................................................... 7

Methodology ......................................................................................... 10

Stroke risk models .................................................................................

- Stroke Riskometer ........................................................................... 13
- Essen Stoke Risk Score ...................................................................... 16
- Stroke Prognostic Instrument I and II ............................................... 17
- Recurrence Risk Estimator ............................................................... 18
- Zuum ................................................................................................. 19

Selecting a model ..................................................................................

- Meta-analyses of popular models .................................................... 23
  - Lemmens (2013) ............................................................................ 23
  - Thompson (2014) ......................................................................... 24

INTERSTROKE – multivariate analysis of risk factors ............................ 26

Model Table ........................................................................................ 28

Conclusions ......................................................................................... 29

References ........................................................................................... 30

Appendix 1 ........................................................................................... 34
Abbreviations / Definitions

APAR – average population attributable risk

AURROC – area under the receiver operating characteristic (ROC) curve which plots the sensitivity (true positive rate) against the specificity (1 – false positive rate)) for the probability of an outcome. A good AURROC model might score 0.9, medium >0.75, while a chance model operating at chance has a value of 0.5

c-statistic - a commonly used performance measure to indicate the discriminative ability of linear regression models. For a binary outcome, c is identical to the area under the Receiver Operating Characteristic (ROC) curve.

GCS – Glasgow Coma Scale

MCI – myocardial infarction

mRS – modified Rankin Score

NIHSS - National Institutes of Health Stroke Scale

PAR – population attributable risk

TIA – transient ischaemic attack

TBI – traumatic brain injury
Introduction
The overall goal of the STARR project is to provide an integrated self management solution for people with stroke by bringing together experts in stroke, video analysis, physiological sensor monitoring, usability, lifestyle analysis, data fusion and platform development to work collaboratively with user communities and patient support organisations.

At the centre of this initiative is an effort to include some intelligence from the existing clinical literature on models of stroke risk into a technological environment to support better self-management by stroke survivors. This is motivated by widespread agreement and evidence that surviving a first stroke increases the risk of experiencing a second stroke. The level of this increased risk is however, unclear. A recent systematic review and meta-analysis\(^1\) found 13 studies over which they estimated the pooled cumulative risk of stroke recurrence. Fig. 1 shows the heterogeneous nature of the data. Results were condensed into cumulative risks of approximately 3,11, 26 and 39 % at 30 days, 1 year, 5 years and 10 years after stroke respectively. The variation in results found by Mohan\(^1\) may be explained by many factors including geographic location, ethnicity, sample size, case mix, basic methodology and the large time frame over which the studies had collected data (some back to the 1960s). The authors report that changes in secondary prevention over this time frame are likely to have strongly influenced the incidence of recurrence. Importantly, there is no set definition of a recurrent stroke in the literature. Examples can be found that range from what would normally be thought of as a TIA (neurological deficit lasting less than 24 hours\(^2\)) up to an exclusion period of 28 days only after which were stroke events included\(^3\).

![Fig. 1 Risk of stroke vs. Time since first stroke from Mohan (2011).\(^1\)](image-url)
However, despite the trend over time in the Mohan analysis\(^1\) being clear, the risk is not fully understood and so cannot be precisely described. The authors go on to declare that:

“To identify true populations at high risk for stroke recurrence, good-quality population-based studies using consistent criteria to define a stroke and a recurrence are needed. In particular, for studies reporting the cumulative risk of stroke recurrence in the first weeks and months after initial stroke, notification and analysis of all stroke recurrences without a defined exclusion period are important to understand fully the risk of recurrence during this period.”

This imprecision and variability has consequences for attempts to model and automate risk assessment and for the STARR project in general.

The following sections will outline how predictive risk modelling is carried out in healthcare and will then focus on stroke in particular. Several stroke models will be described in more detail along with three meta-level studies that are applicable. The first\(^4\) tried to determine the current best model and examined the merits of that question. The second\(^5\) examines how accurate several leading models are compared to clinician’s opinions when both are applied to a new cohort of data. The third\(^6\) is the very large recent multivariate analysis arising from the recent INTERSTROKE epidemiological study looking at which risk factors are actually the most predictive for recurrent stroke.

Note that STARR is concerned only with the risk of recurrent stroke, which is usually, but not exclusively, the risk of a second stroke. There is a more established and larger clinical knowledge base around the use of models to predict outcome after first stroke in the acute phase. These will not be fully described here. Similarly, the set of models behind some popular applications that predict future risk of stroke in the previously unaffected will not be completely addressed, except for the Stroke Riskometer Pro due to its prevalence, endorsement by the World Stroke Organisation and its user-friendly interface design.

This report is written largely in lay terms so as to be understandable by colleagues at the Stroke Association and by those in the STARR consortium with no prior experience of stroke.
Predictive Modelling in Healthcare

The trend towards more sophisticated modelling in healthcare is driven by a desire for better stratification of patients before being allocated to treatment. Clinicians, faced with large amounts of data, or significant amounts of partial data, need support so as to best optimise patient outcome. It is out of scope for this report to give an overview of the prevalence and efficacy of healthcare models across the sector generally. But we can say with confidence that due to advances in data collection and storage, alongside new analytic and modelling approaches i.e. the rich culture around big data and associated visualisation and machine learning technologies, healthcare modelling is a research topic that will come to have growing influence in years to come. Population growth, disease prevalence and healthcare costs make it imperative that we increase our understanding of how to make best use of cohort data.

This is evidenced by the growing amount of academic activity in the area, as healthcare become more influenced by risk stratification modelling from other sectors (e.g. actuarial forecasts in life insurance) and is also able to benefit from relatively recent advances in machine learning. For instance, Deepmind will address automated eye scan analysis and deliver improvements in radiotherapy planning. Some caution is advisable here though, as the old adage ‘garbage in – garbage out’ applies as the prominent failure of Google Flu Trends to accurately predict flu outbreaks, based purely on unregulated search terms and location date about users, readily testifies. Any data that is modelled has to be well described, verified and constrained.

More serious efforts include efforts to predict hospital readmissions. In 2013 alone, 36 peer-reviewed journal articles were published on the subject along with three additional review articles. Highlighting this rapidly growing interest are recent papers focused on simplified readmission scoring for elderly patients, the relationship between readmission and mortality rates, and a systematic review of tools for predicting severe adverse events. Prediction discussions associated with specific areas such as heart failure or within paediatric populations are also very active. As ever, the dual motivation for all these studies is both the improvement of patient care as well as optimising financial planning within hospitals. Nonetheless, although well intended, there are several key lessons that clinicians must be aware of when attempting to model complex phenomena. These include:-

More data does not necessarily mean more insight – data must be high quality and well described in order
to be useful. ‘Well described’ means the context in which data was gathered must be known, it must be reliable, repeatable, and ideally some measure of its accuracy should be recorded. Noisy data will produce an inaccurate model.

**Insight does not necessarily generate value** – gathering lessons from predictive modelling does not necessarily mean that an individual or a system will be able to improve their or its performance. There may be many personal or systemic issues that influence, or even prevent, the ability to extract value from insight.

**Interpreting data is hard and may not be informative** – data may be difficult to use or not lead to insight because it may be noisy, or more usually, it is incomplete. This is especially true for healthcare conditions such as stroke where the pathology and aetiology are not fully understood. This forces clinicians to make judgments and medical decisions using incomplete information every day and treatment decisions made on incomplete information and educated guesses are quite common. To complicate the picture, risk factor management may largely reduce stroke occurrence but it is likely that there are genetic factors too that increase an individuals susceptibility to cerebrovascular disease.

**Modelling is easier than implementing change** – for obvious reasons, modelling can be done in abstraction but implementing actual change involves many more actors acting in the real world and so it is much more complex to achieve

**Developing stroke risk models**

We will not describe at length the process of how new predictive models are developed in stroke care but it is informative to have an overview of a generic methodology that applies to many healthcare conditions not just stroke. Since this is conceptually non-trivial for the non-statistician we must also describe some key terms commonly found in the literature.

When thinking of risk and odds and when it is safe or unsafe to compare them, arguably the most respected source in healthcare analysis is the Cochrane collaboration / network of independent researchers. We can do no better than to quite directly from their handbook\(^ {15}\) that issues guidance on how to conduct meta-analyses of healthcare studies. From chapter 9.2.2.2 where effect measures for dichotomous outcomes are described, they distinguish odds and risk as follows;
“In general conversation the terms ‘risk’ and ‘odds’ are used interchangeably (as are the terms ‘chance’, ‘probability’ and ‘likelihood’) as if they describe the same quantity. In statistics, however, risk and odds have particular meanings and are calculated in different ways. ...

**Risk** is the concept more familiar to patients and health professionals. Risk describes the probability with which a health outcome (usually an adverse event) will occur. In research, risk is commonly expressed as a decimal number between 0 and 1, although it is occasionally converted into a percentage. In ‘Summary of findings’ tables in Cochrane reviews, it is often expressed as a number of individuals per 1000. It is simple to grasp the relationship between a risk and the likely occurrence of events: in a sample of 100 people the number of events observed will on average be the risk multiplied by 100. For example, when the risk is 0.1, about 10 people out of every 100 will have the event; when the risk is 0.5, about 50 people out of every 100 will have the event. In a sample of 1000 people, these numbers are 100 and 500 respectively.

**Odds** is a concept that is more familiar to gamblers. The odds is the ratio of the probability that a particular event will occur to the probability that it will not occur, and can be any number between zero and infinity. In gambling, the odds describes the ratio of the size of the potential winnings to the gambling stake; in health care it is the ratio of the number of people with the event to the number without. It is commonly expressed as a ratio of two integers. For example, an odds of 0.01 is often written as 1:100, odds of 0.33 as 1:3, and odds of 3 as 3:1. Odds can be converted to risks, and risks to odds, using the formulae:

\[
\text{risk} = \frac{\text{odds}}{1 + \text{odds}}, \quad \text{odds} = \frac{\text{risk}}{1 - \text{risk}}
\]

The interpretation of odds is more complicated than for a risk. The simplest way to ensure that the interpretation is correct is to first convert the odds into a risk. For example, when the odds are 1:10, or 0.1, one person will have the event for every 10 who do not, and, using the formula, the risk of the event is 0.1/(1+0.1) = 0.091. In a sample of 100, about 9 individuals will have the event and 91 will not. When the odds is equal to 1, one person will have the event for every one who does not, so in a sample of 100, 100 × 1/(1+1) = 50 will have the event and 50 will not.

The difference between odds and risk is small when the event is rare (as illustrated in the first example above where a risk of 0.091 was seen to be similar to an odds of 0.1). When events are common, as is often
the case in clinical trials, the differences between odds and risks are large. For example, a risk of 0.5 is equivalent to an odds of 1; and a risk of 0.95 is equivalent to odds of 19.

Measures of effect for clinical trials with dichotomous outcomes involve comparing either risks or odds from two intervention groups. To compare them we can look at their ratio (risk ratio or odds ratio) or their difference in risk (risk difference).

**Odds ratio** – the odds ratio is used to assess the risk of a particular outcome (or disease) if a certain factor (or exposure) is present. The odds ratio is a relative measure of risk, telling us how much more likely it is that someone who is exposed to the factor under study will develop the outcome as compared to someone who is not exposed. Odds ratios are a common way of presenting the results of a meta-analysis.

**Logistic regression** = logistic regression predicts the probability of particular outcomes and can be used in cases where multiple causal factors are at play. Results are reported as odds ratios and the simultaneous effect of possibly confounding other variables is dealt with by adjustment of the odds ratio. Logistic regression in healthcare can be used to determine the proportion of new cases that may develop in a given time period, i.e. the cumulative incidence.

**Proportional hazards regression** – this is an analytical method that assesses the *rate* of occurrence of phenomena instead of a proportion. This *incidence or hazard rate* is the number of new cases of disease per population at-risk per unit time. The *hazard function* is the probability that if a person survives to a certain time t, they will experience the event in the next instant. Proportional hazards regression estimates the *hazard ratio*.

In most cases, the interpretation of a hazard ratio is essentially the same as an odds ratio. However it’s worth noting that whilst an odds ratio is derived from calculating the odds of an event in the intervention and the control arms expressed as a ratio, the hazard ratio is derived from calculating the rate (number of events/time) in the intervention and the control arms expressed as a ratio. So there are some additional statistical considerations.

**Case-control** - In a case-control study, patients who have developed a disease are identified and their past exposure to suspected causal factors is compared with that of control patients who do not have the
disease. This allows estimation of odds ratios. Allowance is made for potential confounding factors by measuring them and making appropriate adjustments in the analysis. This adjustment may be done on an individual basis (e.g. pairing each case patient with a control of the same age and sex) or at the group level.

**Methodology**
Current medical understanding points to the likelihood of a secondary stroke as being primarily influenced by at least three things:-

1. The risk factors present before and after the first stroke.
2. The severity of the first stroke.
3. The acute treatment, long-term treatment and lifestyle changes after the first stroke.

Most models, since they are intended for use in the acute phase post first stroke, concentrate on the influence of points 1 and 2 above. The risk factors for stroke are described at length in D2.1. Suffice to say, that there is no universally agreed definitive set of factors, because stroke is not fully understood. As such, best practice in risk factor choice is that of *a priori* clinical knowledge being used to inform the selection. An alternative that must be avoided is data driven predictor selection, where data is used purely because it is available.

Perhaps the largest, highly cited and authoritative set of risk factors is given in Sacco et.al. (1997)\(^ {16}\) and is listed in Appendix 1. This groups factors into those that are modifiable or not and into those for which there are varying degrees of evidence. Beyond this, other variables are often used such as some measure of the level of impairment in the acute phase (e.g. NIHSS, mRS, GCS), or the level of pre-stroke independence etc. Similarly, due to the large increase in access to MRI since the late 1990s, some more recent models such as the Recurrence Risk Estimator are almost entirely driven by variables related to imaging as shown in Fig. 2 below;
Once factors are selected, the issue of how a score on a specific variable is recorded becomes critical. Not all variables (risk factors) have universally agreed definitions. For instance, what is high alcohol consumption? What is the treatment threshold for high systolic blood pressure? Transparency and precision around how a variable is recorded, its frequency, banding and metrics are all issues that affect model performance and more importantly, can seriously impair meta-level analysis across multiple models intending to address the same phenomena.

After factor selection, the generic steps in a modelling study can be summarised as:

- Find a population of concern
- Determine who to compare them against (if relevant for your design)
- Record subjects medical history, current status and outcomes over time
- Regress on multiple factors to determine the relative hazard associated with each factor

The review by Thompson et al (2014) attempts to characterise what makes an acceptable modelling study. In their opinion, a study of good quality should have the following:

**Internal Validity** – any data should be collected prospectively and not retrospectively. Any subjects lost to follow up should be fully documented. All variables must be explicitly and precisely defined. A methodology
to handle missing data should be easy to discern.

**Statistical Validity** – subjective categorisation of continuous variables is forbidden and any boundaries must be based on clinical reasoning

**Model Evaluation** – a model should be validated i.e. tested against a new population of previously unseen test cases. A failure to do this risks over-fitting to the training data and an inability to generalise to new cases. Any new test population should be fully described to make sure it is a good match to the model development cohort. The model’s ability to discriminate, and the range over which it can operate, should both be clear to the external reader.

The process is seen diagrammatically in Fig. 3 from [www.healthcatalyst.com](http://www.healthcatalyst.com) (reproduced with permission), which shows a generic flow chart for model development in the machine learning community.

![Fig. 3 Overview of best practice in predictive model development](image)

Fig. 3 Overview of best practice in predictive model development

A good model must be internally validated i.e. known to give valid predictions for subjects from the underlying, already categorised population. Preferably, the predictions also generalize to populations that are related but not categorised. The use of representative but different data sets at the formulation,
validation and testing stages is crucial. This is especially true in healthcare when data points are not as uniform or clean as those found in mainstream predictive modelling contexts (e.g. comparing some subset of a patient health record to a precise well defined signal in machine learning).

**Stroke risk models**

Predictive models in stroke fall into three main groups; those used in clinical (and non-clinical) settings to assess an individual’s risk of first stroke; those used in the acute setting to predict the level of impairment at a certain time point after first stroke; and those also used in acute and chronic settings to aid stratification of those at risk of recurrent stroke. These acute models arose because performing multiple acute diagnostic investigations for all suspected TIA and stroke patients might overwhelm a medical system and might not be feasible because of resource limitations. Simple and reliable risk estimation of recurrence might be beneficial to high-risk patients so that they can be admitted and investigated early. Strictly speaking, only the latter type of model is of interest to STARR because the project focus is on the sub-acute and chronic phase.

When writing STARR we imagined that predictive models were used across more of the stroke pathway than is in fact the case. In the first 6 months of the project we carried out informal interviews with various therapists in multiple countries; physiotherapists in the UK, Belgium and Spain; occupational therapists in the UK and France; and speech and language therapists (SaLT) in the UK. None of these specialties routinely use predictive models of the type considered here to gauge outcomes. This is to be expected because none of them are primarily interested in prevention of recurrent stroke, their focus is narrower, although of course stroke remains an over-riding concern. Instead, their focus is on improvements to the physiological, communication and functional independence status of their clients. Models are sometimes used in SaLT to aid initial classification of a patient.\(^\text{17}\)

The next section describes several of the leading candidate models for inclusion in STARR. Note that we have excluded models that were developed for specific narrow populations such as CHADS\(^2\)\(^\text{18}\) (for AF patients), California Risk Score\(^\text{19}\) and the ABCD, ABCD\(^2\)\(^\text{20,21}\) and most of their derivatives that use TIA only as a primary event, or that covered wider vascular risk such as QRISK\(^\text{22}\) and SCORE.\(^\text{23}\)

**Stroke Riskometer**

This tool is a fully validated predictive model of a users risk of having a stroke over a coming 5 or 10-year
period. It was developed by leading neurologists and is endorsed by major representative bodies of clinicians, including the World Stroke Organisation and the World Federation of Neurology. An assessment consists of answers to 20 questions pertaining to the following risk factors shown in Table 1. The answers are then weighted to give an overall risk score.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Modifiable</th>
<th>Non-modifiable</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (weight, height)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit &amp; vegetable consumption</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>Gender</td>
</tr>
<tr>
<td>Cognitive problems</td>
<td></td>
<td>Smoker</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td></td>
<td>Heart disease (peripheral)</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td>Stress</td>
</tr>
<tr>
<td>Blood pressure (systolic)</td>
<td>Ethnicity</td>
<td>Hypertrophy (left ventricle)</td>
</tr>
<tr>
<td>Blood pressure medication</td>
<td>Family history of stroke</td>
<td>Memory problems</td>
</tr>
<tr>
<td>TBI</td>
<td></td>
<td>TIA</td>
</tr>
</tbody>
</table>

Table 1. Modifiable and Non-modifiable risk factors assessed for in the Stroke Riskometer.

Note: The risk factors shown in bold could potentially be monitored by a self-management system although it is not clear yet which of those could be automatically monitored and if not, would instead rely on repeated real-time input by a user. This bold formatting to highlight factors that are potentially measureable by STARR will be used for all the other models we describe.

The Stroke Riskometer™ algorithm was derived from the well-known Framingham Stroke Risk Score (FSRS)²⁴ prediction algorithm and enhanced to improve accessibility and to include several additional major risk factors shown to be important for stroke, largely based on the first stage of the INTERSTROKE²⁵ study which is described later in this document. The FSRS is not described here because the Stroke Riskometer is essentially an extension of it; some of the extra risk factors used are the central targets in the new WHO Global Action Plan for the NCD 2013–2020²⁶.

The new app was recently validated by its authors²⁷ and was shown to have a very similar performance to the FSRS, although both models in fact performed poorly in predicting first stroke events i.e. they had low discrimination power when categorising users with both high and low known risk profiles. Discrimination is a measure of how well a model separates patients with the event in follow-up from those without.

However, the Stroke Riskometer is included here not because of its algorithm but because it is the first high-profile mobile app in stroke for general use by the wider population as well as high-risk individuals.
This is due to it reflecting new thinking in cardiovascular risk presentation that it is best to show an absolute risk of stroke as well as a matched baseline risk for comparison for each app user i.e. a user can see how they compare to someone of the same age and gender who has no risk factors.

Figure 4 above shows a selection of screens from the mobile application including different ways of displaying relative stroke risk as well as generic help about the condition. The Stroke Riskometer does allow repeated input of the whole survey in order to generate individual snapshots of risk but it does not show trends nor does it display the effect(s) of changing individual behaviours on risk.
Essen Stoke Risk Score

The Essen Stroke Risk Score (ESRS) was derived from a large study of more than 6000 cerebrovascular patients taking part in a Clopidogrel vs. Aspirin study after they had had a first stroke28. The ESRS uses a 10-point linear scale to broadly stratify a patient’s risk of stroke within one year.

<table>
<thead>
<tr>
<th>Age 65-75 (1 point)</th>
<th>Other cardiovascular disease (except MI and AF) (1 point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &gt; 75 (2 points)</td>
<td>Peripheral artery disease (1 point)</td>
</tr>
<tr>
<td>Arterial hypertension (1 point)</td>
<td>Smoker (1 point)</td>
</tr>
<tr>
<td>Diabetes mellitus (1 point)</td>
<td>Previous TIA / stroke (1 point)</td>
</tr>
<tr>
<td>Previous myocardial infarction (1 point)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Risk factors in the ESRS model

Online versions of the ESRS, such as found at www.pmidcalc.org present the tool and results as shown in Fig. 5 below.

Fig. 5 ESRS output expressed in relation to patients recorded in the REACH registry, a very large worldwide registry of atherothrombotic risk factors and ischemic events in which the ESRS was validated.29

An ESRS score of less than 2 is thought of as low risk and a score of greater than 3 is high. The model is thought of as being more useful when categorising the extremes of risk.
Stroke Prognostic Instrument I and II

Version 1 of the Stroke Prognostic Instrument SPI-I was developed in the early 1990s\(^3\) to predict recurrence within 90 days after TIA or minor stroke only. Just five factors with rankings based on relative values of regression coefficients were found to be sufficient for categorization of patients into three groups (0 to 2 points, 3 to 6 points, and 7 to 11 points) based on their being diabetic (3 points), having severe hypertension (2 points), older than 65 years (3 points), with coronary heart disease (1 point) and being known to have had a stroke and not a TIA (2 points). When tested against 4 different cohorts, the pooled estimate risk of the three groups was 9, 17, 24% respectively.

To improve SPI-I, the inclusion of some of a large set of candidate factors were considered including physical functioning, cognitive status, infarct presence on imaging, ECG characteristics etc.\(^3\) But it was decided that it was having prior stroke and/or congestive heart failure as factors that would allow SPI-II to use the most easily obtained clinical variables that were well documented to influence the risk of recurrent stroke\(^3\)\(^2\)\(^,3\)\(^3\). The predictive window was also increased from 90 days to 2 years post first event.

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Adjusted relative risk for stroke or death within 2 years</th>
<th>Regression coefficient</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Congestive heart failure</td>
<td>1.8</td>
<td>0.57</td>
<td>3</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>1.7</td>
<td>0.55</td>
<td>3</td>
</tr>
<tr>
<td>Prior stroke</td>
<td>1.7</td>
<td>0.51</td>
<td>3</td>
</tr>
<tr>
<td>Age &gt; 70y</td>
<td>1.6</td>
<td>0.46</td>
<td>2</td>
</tr>
<tr>
<td>Stroke (not TIA)</td>
<td>1.5</td>
<td>0.38</td>
<td>2</td>
</tr>
<tr>
<td><strong>Severe hypertension</strong></td>
<td><strong>1.2</strong></td>
<td><strong>0.19</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>1.1</td>
<td>0.13</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Risk factors in the SPI-II model

Testing of SPI-II in 3 new cohorts confirmed that SPI-II performs better than SPI-I. Outcome rates for SPI-II in risk groups I to 3 were now 10%, 19%, and 31% and there was much clearer distinction between the low
and high risk groups compared to SPI-I. Performance overall was moderately good but variable and was thought to be affected by varying age groups, overall health levels and comorbidity status in the cohorts, all of which will introduce noise into the dataset. Importantly, differing means of gathering baseline data, where even self-report and casual enquiry were used, all contributed to weakening model performance.

Recurrence Risk Estimator
This recent model\textsuperscript{34} was built to use mainly imaging data to enable stratification of ischemic stroke patients for risk of recurrence up to 90 days post first event. The authors were motivated by the lack of a robust predictive tool for estimating short-term risk, pointing out that SPI-II and the Essen scores have a long-term focus.

Data from almost 1458 consecutive historical patients was analysed in which 60 strokes occurred during the follow-up period. Prior publications\textsuperscript{35,36,37} had shown that image features may be useful in determining prognosis after stroke.

<table>
<thead>
<tr>
<th>Imaging feature</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple acute infarcts\textsuperscript{20,21}</td>
<td>Multiple noncontiguous lesions that were hyperintense on DWI and hypointense on ADC maps</td>
</tr>
<tr>
<td>Simultaneous involvement of different circulations\textsuperscript{20}</td>
<td>Multiple acute ischemic lesions secondary to acute or subacute infarcts in both right and left anterior circulations or in both anterior and posterior circulations</td>
</tr>
<tr>
<td>Multiple infarcts of different ages\textsuperscript{19,25}</td>
<td>Ischemic lesions with hyperintense signal on DWI that met at least 2 of the following 3 combinations of signal changes on ADC and FLAIR images:</td>
</tr>
<tr>
<td></td>
<td>Hypointense on ADC, isointense on FLAIR (hyperacute)</td>
</tr>
<tr>
<td></td>
<td>Hypointense on ADC, hypointense on FLAIR (early acute)</td>
</tr>
<tr>
<td></td>
<td>Isointense on ADC, hypointense on FLAIR (late acute or subacute)</td>
</tr>
<tr>
<td>Infarct topography\textsuperscript{7,14,21}</td>
<td>Isolated cortical, isolated deep (lacunar), and subcortical with or without cortical involvement</td>
</tr>
<tr>
<td>Chronic infarcts\textsuperscript{22}</td>
<td>Clinically silent or symptomatic territorial lesions that were hyperintense on FLAIR and hyper- or isointense on DWI</td>
</tr>
<tr>
<td>Internal watershed infarcts\textsuperscript{23}</td>
<td>Rosary-like pattern of infarcts that were arranged in a linear fashion parallel to the lateral ventricle and located in the centrum semiovale or corona radiata</td>
</tr>
</tbody>
</table>

Table 4. Definitions of imaging features associated with stroke recurrence at 90 days and embedded in the RRE-90 model. ADC = apparent diffusion coefficient; DWI = diffusion-weighted imaging; FLAIR = fluid-attenuated inversion recovery.

The study confirmed this and found that the strongest predictors of recurrence were the stroke subtype, a prior history of stroke or TIA, and the distribution, age and spatial variety of areas of brain damage. Multiple investigators were needed to characterise imagery due to the risk of interrater error when
subjectively scoring images of this type. Varieties of the model were produced that did and did not include imaging as well as baseline clinical data. The model that included imaging variables outperformed the simpler model when internally validated, claiming an AURROC value of 0.8 (see next section). A strength of the RRE-90 is that it can be run using data that is available to the clinician immediately after initial stroke evaluation.

**The Recurrence Risk Estimator (RRE)**

*A Score for Prediction of Early Risk of Recurrence After Ischemic Stroke*

![RRE Image](image)

| Acute stroke MRI is available (within the first 72 hours) | Yes | No |
| History of TIA or stroke within the preceding month of index stroke | Yes | No |
| Multiple acute infarcts | Yes | No |
| Simultaneous infarcts in different circulations | Yes | No |
| Multiple infarcts of different ages | Yes | No |
| Isolated cortical infarcts | Yes | No |
| CCS etiologic stroke subtype | Large Artery Atherosclerosis | Yes | No |
| Cardioaortic Embolism | Small Artery Occlusion | Other Causes | Undetermined Causes |

**Fig. 6** – The web interface for the RRE-90 model online at www.nmr.mgh.harvard.edu/RRE

Figure 6 shows a completed RRE-90 survey in a clinician interface and below this a stratification score and associated 7 and 90 day risk levels.

**Zuum**

The Zuum healthtracker application (http://zuum.wustl.edu) is a state of the art standalone health monitor deriving from the long-standing consensus of opinions of a Harvard-led group of academic medical practitioners and epidemiologists. The app is managed and marketed by Washington University School of...
Medicine and is aimed at the lay public. Standalone in this sense means that Zuum is not connected in real time to any external sensors. Instead, it engages the user in an intuitive and simple lifestyle survey along with some questions on their medical history when relevant, and outputs their relative risk of developing several chronic conditions. This approach is claimed to have been validated as effective as a motivational tool and the app has a long-standing website equivalent at www.yourdiseaserisk.wustl.edu. The addition of clinical knowhow and updated knowledge both to the website and the app are largely based on a group consensus process, which places emphasis on study design and gives greatest weight to combined analyses of prospective cohort studies. The list of included chronic conditions has grown over the years and currently consists of heart disease, stroke, diabetes, lung cancer and colon cancer, as well as breast cancer for women and prostate cancer for men. The tool allows repeated sessions with the survey so that a user can track any trends in their risks and allow trends in the habits. There are three design aspects that make Zuum relevant to the STARR project. They are:

1. User friendly design
An example is shown below in the baseline survey data collection screen. Zuum has a simple non-cluttered visual display with minimal colours. Shallow menus allow entry of constrained input, meaning better navigation for the user and easier data processing by the app. Contrast this use of lay language with that of the RRE-90 shown previously.

Fig. 7. Streamlined, patient friendly design in Zuum.

2. Dynamic interaction
A strong selling point of Zuum is that it runs multiple models in the background and so can show the effects
of lifestyle change in real time. Initial risk results are displayed in simple categorical terms (e.g. low, average, high) without the use of numbers or percentages. A user can then metaphorically ask the question “What if I changed my behaviour in relation to risk factor A?” Fig Y shows how weight loss reduces risk of stroke, heart disease and colon cancer for this user.

![my results](image)

**Fig. 8.** Feedback showing the effect of lifestyle change on risk. Original risk is grey, modified risk in orange.

3. Simple, specific advice
Zuum provides personalised and simple advice directly related to risk factors and of how to influence them.

![lower my risk](image)

**Fig. 9.** User friendly advice is embedded in Zuum.

Direct communication by STARR with the Zuum developers has revealed details of the modelling behind the app. For stroke, the factors and weightings are shown in Table 5.
<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Relative risk score</th>
<th>Risk Factor</th>
<th>Relative risk score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1</td>
<td>Hypertension</td>
<td>3</td>
</tr>
<tr>
<td>Family history (immediate)</td>
<td>2</td>
<td>Diabetes</td>
<td>2</td>
</tr>
<tr>
<td>African American)</td>
<td>2</td>
<td>Total cholesterol</td>
<td></td>
</tr>
<tr>
<td>Waist size</td>
<td></td>
<td>Binary</td>
<td></td>
</tr>
<tr>
<td>Women &gt;35in</td>
<td>1</td>
<td>Told have high cholesterol</td>
<td>1</td>
</tr>
<tr>
<td>Men &gt;40in</td>
<td>2</td>
<td>Range mg/dL</td>
<td></td>
</tr>
<tr>
<td>Smoking 1-14 per day</td>
<td>1</td>
<td>&gt;=240</td>
<td>1</td>
</tr>
<tr>
<td>Smoking 15-25 per day</td>
<td>2</td>
<td>HDL cholesterol &lt; 40 mg/dL</td>
<td>2</td>
</tr>
<tr>
<td>Smoking &gt;25 per day</td>
<td>3</td>
<td>Physical activity of 30 minutes per day or 3 hours per week</td>
<td>-2</td>
</tr>
<tr>
<td>Quit smoking &lt;10 years</td>
<td>1</td>
<td>Fruit/Vegetables &gt;= 5 servings per day</td>
<td>-1</td>
</tr>
<tr>
<td>Cereal/Fibre whole grain &gt;= 3 servings per day</td>
<td>-1</td>
<td>Alcohol &gt;= 3 drinks per day</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5. Risk factors and relative scores (above) and the mapping to relative risks from scores (below) for stroke in Zuum.

<table>
<thead>
<tr>
<th>Key</th>
<th>RR</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoring</td>
<td>(-) decr risk</td>
<td>(+) incr risk</td>
</tr>
<tr>
<td>1 = weak</td>
<td>0.7 - &lt; 0.9</td>
<td>1.1 - &lt; 1.5</td>
</tr>
<tr>
<td>2 = moderate</td>
<td>0.4 - &lt; 0.7</td>
<td>1.5 - &lt; 3.0</td>
</tr>
<tr>
<td>3 = strong</td>
<td>0.2 - &lt; 0.4</td>
<td>3.0 - &lt; 7.0</td>
</tr>
<tr>
<td>4 = very strong</td>
<td>&lt;0.2</td>
<td>&gt; 7.0</td>
</tr>
</tbody>
</table>

Table 5 shows that of the models we have described, Zumm uses by far the highest number of lifestyle factors that could theoretically be changed and monitored by STARR technologies.
Selecting a model

STARR is not resourced or contractually required to build its own predictive model, nor to do any new analysis of the relative predictive power of individual stroke risk factors (which is in effect the same thing as building a new model). The call text explicitly mentioned the use of existing models only. This leaves us to progress by checking comparative reports in the literature and to act if possible in light of the results of those. As we will see, such reports are not hugely supportive of the performance of current models and so our approach is to take inspiration where possible from the models but not to rely solely on their performance. A more pragmatic method is to record those factors we have access to and that are known to be responsible for stroke. With this in mind we report on two meta-studies comparing model performance as well as on a large epidemiological study to find precisely which factors are most responsible for stroke.

Meta-analyses of popular models
There are two recent meta-analyses of the current set of predictive models of recurrent stroke. Both papers are from reputable clinicians from leading academic stroke groups in Europe.

Lemmens (2013)
The review by Lemmens et al (2013) checked the quality of published risk scores based on the characteristics of various derivation and replication studies. 17 possibly eligible scores were found in the literature of which only 2 had stroke as the unique index event. TIA only, found in 10 studies such as the ABCD score and its derivatives, was much more common. Inconsistent or poorly described methodologies in some aspects of baseline assessment and many aspects of follow-up were common. The models also varied in the length of the future time window in which the prediction applied. Other models were not externally validated, or failed on validation (e.g. the Framingham Stroke Risk Score, Dutch TIA, LiLAC). Table W below is an extract from their model comparison table with the models using only TIA as the index event having been removed.
Determining which is the ‘best’ of the remaining scores is difficult. Having both TIA and stroke as an index event adds noise to any model. Applying a model to a predictive window for which it was not explicitly designed will also reduce power. The presence or not of patients with atrial fibrillation is yet another confounding factor. Finally, few studies fully characterized either the index or follow up stroke event in terms of it being ischaemic or hemorrhagic. The authors state succinctly that:

“It can be assumed that risk factors of a second cerebrovascular event differ between early and late recurrence, as well as between TIA and stroke as the index event”.

In terms of absolute performance, very few of the models in any context had a sufficiently high c-statistic (>0.8) on which to confidently base clinical decision-making. Low discriminability means that stratifications boundaries are vague. There was moderately good performance in predicting early risk after TIA, although this has since been disputed\(^42\), and there was some support for including imaging to better predict risk after stroke at 90 days. Such limited positive findings only further reduce the potential model base for STARR. However, the authors caution that the very heterogeneity of stroke, its multiple causes and outcomes, mean that a universal best-fit model may not be plausible or desirable. The question of what is the ‘best’ model may itself be redundant. They conclude with a cautionary final note:

“When evaluating risk models, changes in diagnostic and therapeutic avenues need to be investigated to identify a correlation with improved patient care; thus far, this has not been clearly determined”.

In other words, these authors claim that the models are not yet proven to be clinically useful, despite predictive performance or their validation status.

Thompson (2014)
The second review and study\(^5\) included predictions of both stroke and MCI, although the latter are not discussed here. Again several models were assessed for the quality of their development, and for how many evaluation studies could be found for them, but in addition the models were also evaluated against a new cohort of patients from the Edinburgh Stroke Study\(^43\). Crucially, both model performance and that of clinicians was assessed in predicting the outcome of the new cohort. Clinicians were asked to provide a gut
feeling of the chance of recurrent stroke within one year in patients and their results were compared to
discriminations made by the models. Clinicians (n = 13) were a mix of fully trained and partially trained
stroke physicians and neurologists.

Twelve model development studies were found. The many limitations of how well the models were initially
built are shown in Figure 10 below. The authors were critical of this imprecision and sloppiness, especially
around the unnecessary categorisation of continuous variables and the low prevalence of events per
variable.

![Figure 10. Quality aspects of model development from Thompson (2014).](image)

In terms of performance in pre-existing evaluation studies, only the ESRS and the SPI-II were reported as
having multiple evaluation studies whose results could be adjusted and pooled for comparison. Results are
shown in Figure 11 overleaf.
Fig. 11. Meta-analysis of AUROCC values for ESRS and SPI-II extracted from Thompson (2014).

The results of this pooled analysis, with mean AUROCC values of around 0.6 are generously described as ‘modest’ by the authors. When applied to the new Edinburgh Stroke Study cohort of 542 patients, the AUROCC values for the clinicians, ESRS and SPI-II were 0.53, 0.56 and 0.58 respectively. The authors concluded that discrimination performance for both models and the informal opinions given by clinicians was poor, barely above chance levels.

**INTERSTROKE – multivariate analysis of risk factors**

The INTERSTROKE study is an important, recent high-profile study in the global epidemiology of stroke that was published in *The Lancet* in July 2016. At its centre was a large case-controlled study comparing the presence of risk factors in patients who were within 5 days of their acute first stroke, against age and sex-matched individual with no history of stroke. The total subject population was of almost 27,000 people from 32 different countries. Both ischaemic and haemorrhagic strokes were included and severity was categorised at baseline and at one month on the modified Rankin score. Candidate risk factors had earlier been identified in a previous INTERSTROKE study that in turn had been derived from the related INTERHEART project. Risk factor assessment was by structured questionnaire or self-report or known medical history and all factors were precisely defined in terms of any cut-off points e.g. the banding of waist-to-hip ratio scores into three tertiles T1, T2 and T3 (with e.g. boundaries for men at 0.91 and 0.97), or the categorisation of levels of alcohol consumption (none, low, moderate, high).
INTERSTROKE’s aim was to detect possible associations between the risk factors and all types of stroke. There were further sub-studies around sex, age and geographic location that do not concern us here. The study methodology was to build many logistic regression models to see which best factors accounted for variation in outcome between the carefully matched pairs. The factors in question were hypertension, smoking, diabetes mellitus, physical activity, diet, psychosocial factors, abdominal obesity, alcohol, cardiac causes, and apolipoprotein level. Estimates for odds ratios (OR) and the accompanying 99% confidence intervals (CI) were reported for every risk factor and their combinations. The top line result was reported in terms of Population Attributable Risk (PAR) which is the reduction in incidence that would be observed if the population were entirely unexposed to the factor or factors, compared with its current exposure pattern. The PAR can lead to a total of more than 100% when individual risk factors are simply added, so a more cogent measure, known as average PAR (APAR), which does not exceed 100%, was also reported. APAR is found by adding all risk factors to the model in every possible permutation which has the advantage that it allows measurement of the independent proportion of PAR that each risk factor contributes to the overall PAR for all risk factors.

Fig. 12. INTERSTROKE’s main result showing the relative predictive power of each of the ten main risk factors in the study.

The authors report that “Collectively, these risk factors accounted for 90-7% of the PAR for all stroke worldwide (91-5% for ischaemic stroke, 87-1% for intracerebral haemorrhage), ..... Hypertension was more associated with intracerebral haemorrhage than with ischaemic stroke, whereas current smoking, diabetes, apolipoproteins, and cardiac causes were more associated with ischaemic stroke (p<0-0001).” Individual
data per risk is shown in Figure 12 above.

As expected, the sum of PARs for each risk factor exceeded the total given for the composite of factors. The authors emphasis that this supports the claim that “... stroke is a consequence of numerous causal risk factors, rather than a single cause ...”, and that “Generally, multiple risk factors are needed to act together to be sufficient to cause an acute stroke.” APAR results were lower for individual risk factors than PAR but the relative ranking of each was preserved. These results are mutually supportive with PAR reflecting the effect of complete removal of an individual factor and APAR being more in line with the effect of changing multiple risk factors at once.

The consequences for STARR are that we should attempt to simultaneously influence as many of these risk factors as possible, whether or not current predictive models capture their precise interaction.

Model Table
This table summarises aspects of the strongest applicable common models that STARR will consider, along with the Stroke Riskometer and Zuum models whose functionality should also be taken into account by STARR.

<table>
<thead>
<tr>
<th>Model</th>
<th>Models recurrent risk in chronic population</th>
<th>Number of modifiable risk factors included</th>
<th>Number of modifiable factors measurable by STARR</th>
<th>Requires hospital medical record data.</th>
<th>Recommended design for STARR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke Riskometer</td>
<td>No</td>
<td>8</td>
<td>8</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Essen Stroke Risk Score (ESRS)</td>
<td>Yes</td>
<td>2</td>
<td>2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Stroke Prognostic Instrument (SPI-II)</td>
<td>Yes</td>
<td>1</td>
<td>1</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Recurrence Risk Estimator (RRE-90)</td>
<td>No</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Zuum</td>
<td>No</td>
<td>9</td>
<td>9</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Conclusions

The sobering fact is that recurrent stroke cannot be readily predicted by man or machine or algorithm. Add to this that none of the prediction models described here are in general use, either because their predictive performance is too poor or because the models are too hard to use for reasons of time or cost.

Nonetheless, the aetiology of a significant portion of strokes is well understood as evidenced by the INTERSTROKE study. So STARR must assume a middle ground, where it takes inspiration from measuring as many common risk factors as possible, as urged by INTERSTROKE; and it also gives non-specific guidance by acknowledging the relative weighting of those factors as they are expressed in the models. This should be packaged in an easy to understand dynamic interface, as in Zuum, where the idea is not to give risks in absolute terms (5% higher, 10% lower etc.) but instead to give the general direction of outcome from the modified risk. The key requirement is to motivate self-management in the sub-acute and chronic phases. Therefore how the general guidance is presented is almost as important as the reliability of that guidance itself (with the proviso that great caution is followed). For instance, large fluctuations in blood pressure and high maximum blood pressure itself are clearly important predictors and we can confidently express this to a user. But minor daily variations in factors should not lead to an alarmed user and the language and caveats used around the expression of risk are important. At all times, the quality of the evidence for stroke interventions must be kept in mind – this is described at length in D2.1. Encouraging self-awareness and trend-awareness in lifestyle is the main aim of the STARR project. By products of this should be a reduction in co-morbidities and a great culture of drug adherence (the latter driven by technological prompting). By monitoring as many risk factors as possible it is likely, but not completely guaranteed, that we will reduce risk of further stroke since the very large correlations found in INTERSTROKE and likely to have a causal basis in reality. That we cannot fully model that process yet, for many of the reasons described above, many concerning heterogeneity of populations and methodologies, is no reason to not develop a state of the art monitoring system such as STARR.
References


5. Thompson, DD, Murray, GD, Dennis, M, Sudlow, C, & Whiteley, WN. 2014, 'Formal and informal prediction of recurrent stroke and myocardial infarction after stroke: a systematic review and evaluation of clinical prediction models in a new cohort' BMC Medicine, v12, 58


Appendix 1

Risk Factors


Stroke. 1997;28:1507-1517

Risk Factors for First Ischemic Stroke

Well-documented risk factors

Modifiable, value established

Hypertension
Cardiac disease
Atrial fibrillation
Infective endocarditis
Mitral stenosis
Recent large myocardial infarction
Cigarette smoking
Sickle cell disease
Transient ischemic attacks
Asymptomatic carotid stenosis

Potentially modifiable

Diabetes mellitus
Hyperhomocysteinemia
Left ventricular hypertrophy
Nonmodifiable

Age
Gender
Hereditary/familial factors
Race/ethnicity
Geographic location

Less well-documented risk factors

Potentially modifiable

Elevated blood cholesterol and lipids
Cardiac disease
Cardiomyopathy
Segmental wall motion abnormalities
Nonbacterial endocarditis
Mitral annular calcification
Mitral valve prolapse
Valve strands
Spontaneous echocardiographic contrast
Aortic stenosis
Patent foramen ovale
Atrial septal aneurysm
Use of oral contraceptives
Consumption of alcohol
Use of illicit drugs
Physical inactivity
Obesity
Elevated hematocrit
Dietary factors
Hyperinsulinemia and insulin resistance
Acute triggers (stress)
Migraine
Hypercoagulability and inflammation
Fibrin formation and fibrinolysis
Fibrinogen
Anticardiolipin antibodies
Genetic and acquired causes
Subclinical diseases
Intimal-medial thickness
Aortic atheroma
Ankle-brachial blood pressure ratio
Infarctlike lesions on MRI

**Socioeconomic features**

*Nonmodifiable*

Season and climate