PRECISE4Q

PREDICTIVE MODELLING IN STROKE

DELIVERABLE

Project Acronym: **Precise4Q**

Grant Agreement number: **777107**

Project Title: **Personalised Medicine by Predictive Modelling in Stroke for better Quality of Life**

D5.2 – Acute Study Protocol

Revision: 1.0

Authors and Contributors	Main author: Vince Madai (CUB); Contributors: Dietmar Frey (CUB), Michelle Livne (CUB)			
Responsible Author	Vince Madai		**Email**	vince_istvan.madai@charite.de
	Beneficiary	CUB	**Phone**	+4930450450569

Project co-funded by the European Commission within **H2020-SC1-2016-2017/SC1-PM-17-2017**		
Dissemination Level		
PU	Public, fully open	x
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

Revision History, Status, Abstract, Keywords, Statement of Originality

Revision History

Revision	Date	Author	Organisation	Description
0.1	15.3.19	Vince Madai	CUB	first version rationale
0.2	18.3.19	Michelle Livne	CUB	revision rationale
0.3	21.3.19	Vince Madai	CUB	first version study design
0.4	25.3.19	Vince Madai	CUB	revision rationale and study design
0.5	1.4.19	Dietmar Frey	CUB	review deliverable
0.6	8.4.19	Vince Madai	CUB	revision deliverable
0.7	23.4.19	Vince Madai	CUB	update new literature for rationale
0.8	25.4.19	Vince Madai	CUB	revision objectives
0.9	29.4.19	Michelle Livne	CUB	review deliverable
0.99	29.4.19	Dietmar Frey	CUB	review deliverable
1.0	30.4.19	Vince Madai	CUB	finalization deliverable

Date of delivery	Contractual:	30.4.19	Actual:	30.4.19
Status	final **x** /draft ☐			

Abstract (for dissemination)	This deliverable summarizes the rationale for the acute stroke study and a short description of the planned study design.
Keywords	acute stroke, study, precision medicine, predictive modelling

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Content

This deliverable does not contain figures

Executive Summary

The high mortality and the rates of disability make stroke a tremendous medical and socio-economic burden in Europe.
Precision medicine promises highly improved stroke outcome and better treatment. However, for the lately introduced new stroke treatment, mechanical thrombectomy (MT), precision medicine is largely untapped. First, precision medicine approaches need to follow current best practice guidelines, otherwise, their predictive value is highly limited. In stroke, this means that predictive modelling for precision medicine needs to be performed with current MT data. Most available databases, however, stem from the pre-MT era. Also, modern machine learning approaches need to be able to incorporate data from multiple sources in different formats. For the best performing algorithms like deep learning, we need to tailor specific architectures - so-called multistream architectures - which can integrate different inputs into one predictive model.
Further, prediction models need to be able to extract the necessary features automatically from the given data. Decision support systems integrated into the clinical workflow cannot rely on any manual input. Moreover, it is important to consider that precision medicine approaches will be used for every patient. however, it is known that study data does not necessarily represent the population of real-world patients due to constraints like the need for formal consent, specific populations frequenting university hospitals and others. It is important to show the capacity of predictive models to maintain their predictive performance in cohorts of real-world data. Lastly, the label used must be informative for patients and capture patient-centred quality-of-life (QoL) measures which are relevant for patients. Here, research into new QoL measures as outlined in the goals of P4Q is warranted and should involve patient-reported outcomes. Thus we will perform P4Q-AS to tackle these shortcomings and collect current MT treatment data to pave the way for precision medicine in stroke treatment.

1. Rationale for the Acute Stroke Study within P4Q

Ischemic stroke is the death of brain tissue due to the sudden lack of blood supply. Most strokes are the result of sudden obstruction of brain vessels by a blood-clot, i.e. thromboembolism. Approximately 900.000 Europeans suffer an ischemic stroke each year(Béjot et al., 2016). About 25% of men and 20% of women can expect to suffer a stroke if they live to be 85 years old(Meairs et al., 2006). The overall mortality of stroke is up to 30%(Béjot et al., 2016) making stroke a leading cause of death in developed countries. Moreover, up to half of the surviving patients remain permanently disabled(Hankey, 2003). The high mortality and the rates of disability due to stroke make it a tremendous medical and socio-economic burden in Europe.

1.1. Ischemic Stroke Types and Pathophysiology of Stroke (Progression)

Three major types of ischemic stroke can be distinguished: Cardio-embolic stroke, large vessel atherosclerotic stroke and lacunar stroke.
Cardio-embolic stroke is most commonly caused by atrial fibrillation, although other causes like endocarditis are possible. Large vessel atherosclerosis as the cause leads to the occlusion of a large brain vessel by a local thrombotic process (in contrast to embolism). Lastly, lacunar stroke is the occlusion of a small penetrating brain vessel, which is attributed to local vessel wall damage through lipohyalinosis.

In all stroke types, the occlusion of the blood vessel leads to the death of neurons in grey matter due to the lack of oxygen leading to a cascade of intracellular reactions due to lack of ATP. Briefly, the cascade is characterized by ionic imbalance, neurotransmitter release and inhibition of reuptake(Xing et al., 2012). Here, mainly glutamate plays a role, where glutamate leads to calcium influx, which in turn promotes excessive sodium and water influx leading to the so-called cytotoxic edema. Cell death is induced by damage to cell organs by calcium and by oxygen radicals produced by mitochondrial production. Lately, also cortical spreading depression was linked to damage in stroke(Lauritzen et al., 2011).

Depending on the stroke type, smaller or larger areas of the brain are affected. In lacunar stroke, strokes are small - so-called lacunar lesions. This is explained by the fact that the affected penetrating brain vessels are small and usually only one is affected. In embolic or local thrombotic stroke, also called territorial stroke, the size of the affected area is dependent on the localisation of the blood-flow blocking thrombus - proximal occlusions leading to larger affected areas - and the amount of collaterals(Leng et al., 2016). Here, complete infarctions of whole vessel territories and even hemispheres can occur which are termed 'malignant infarction' due to their catastrophic outcomes(Hacke et al., 1996).

Stroke, however, is not a static, but rather a dynamic process. Animal and early positron-emission-tomography (PET) studies have suggested the "penumbra" model of acute ischemic stroke(Sobesky, 2012). The penumbra model includes that the final infarction is determined by the severity and by the duration of the ischemia. Here, stroke is characterized by an early core of infarction where neurons die quickly in the course of minutes and no therapeutic intervention is possible. This core, however, is surrounded by neurons in a dormant state, coined the penumbra. In this penumbra region, neurons are in a state which is characterized by reduced blood flow which is sufficient to prevent death in the near future but does not allow normal function of the cells. This state cannot be maintained forever, however, and over time the core spreads until also in the penumbra regions the neurons die. How

much of the penumbra is lost this way, is highly influenced by the grade of collateral flow, i.e. how much blood reaches the penumbra via indirect routes. Collateral flow shows high variance and it is established that higher collateral grades are associated with lower penumbral loss and better tissue fate(Bang et al., 2008; Jung et al., 2013). While stroke progression can be very quick, especially in malignant stroke, persisting penumbral tissue has been proven to exist up until 48 hours(Heiss et al., 1992).

These results together newer results from randomized studies, however, imply that the steady loss of neuronal tissue might not the norm for stroke patients(von Kummer, 2019). Rather, it is thought that there exists a stroke population with stable conditions which will benefit from stroke treatment many hours, even days after the stroke(von Kummer, 2019).

The best outcome for the tissue can be achieved by recanalization of the vessel with following reperfusion of the tissue. This can sometimes happen spontaneously, but it is naturally the goal of the causal acute ischemic stroke treatment options which are discussed in the following.

1.2. Acute Stroke Treatment

The goal of all causal stroke treatment is recanalization, i.e. the re-opening of the vessel, which is occluded by the thrombus, and following reestablishment of the blood flow with reperfusion. In current practice, first always neuroimaging, either CT or MRI, is performed to rule out haemorrhage. Then, two different treatment options exist. One, the dissolution of the blood clot using i.v. transcombinant plasminogen activator (t-PA). Two, the mechanical removal of the blood clot using intra-arterial mechanical devices.

1.3. Status of i.v.-Thrombolysis

Extensive research has been performed on i.v. thrombolysis using t-PA until today. Here, we will not reiterate this research in detail. Rather, we will summarize the current status-quo for the use of i.v.-t-PA. T-PA catalyzes the reaction from plasminogen to plasmin which is the strongest agent for blood clot breakdown. Current guidelines(Powers William J. et al., 2018) allow the usage of i.v. t-PA within 4.5 hours after established stroke onset. After 4.5 h, the number-needed-to-treat (NNT) drops below the number-needed-to-harm (NNH), e.g. by secondary bleeding, and general therapy of all stroke patients is not recommended. It is important to mention that the NNT decreases with time from stroke onset. For a patient with a good outcome, in the first 90 minutes 4-5 patients need to be treated. After more than 230 minutes, the NNT drops to 1:14.

It is important to mention that i.v. t-PA is given here according to the so-called time-clock paradigm. This means that only the clinical status, as well as contraindications and the time from stroke-onset, are taken into account when deciding on initiation of the therapy. Whether the patient has or does not have any salvageable penumbra, is not established. Treatment selection strategies are discussed in 1.5.

1.4. Status of Mechanical Thrombectomy

After several successful studies published in 2015(Papanagiotou Panagiotis and Ntaios George, 2018) mechanical thrombectomy (MT) has become the cornerstone of acute ischemic stroke treatment and can be considered standard of care for large vessel stroke (Turc et al., 2019). Here, the thrombus is

removed in a neuroradiological invasive procedure by specialised neuroradiologists. Whereas in the beginning only very selected patients received this procedure, today in some countries ⅕ of all patients receive MT with an average of 7% (Aguiar de Sousa et al., 2019). A reason for this is the relatively large time window where MT is possible in contrast to i.v.-thrombolysis. Current guidelines(Turc et al., 2019) allow MT with high evidence within the first six hours of stroke. Moderate evidence exists that MT is also beneficial in a time window between 6 and 24 hours after stroke onset. Here, MT should be performed according to the criteria of the DEFUSE-3 or DAWN trials. Overall, the NNT of MT is 1:8 which is one of the lowest numbers for stroke treatment so far(Church et al., 2017). Overall, It can be expected that the ratio of patients with stroke treated with MT will rise with new study results, new organizational models for stroke treatment and wider access to the procedure.

1.5. Stroke Outcome: Treatment Selection Strategies

Given the previously introduced penumbra model, and the fact that salvageable tissue can persist long after stroke onset, and the insufficient characteristic of the time clock approach to not take into account the penumbra, selection strategies for individual patient selection have been a focus of research since many years.

We cannot give a comprehensive review here, but we will focus on the major facts. A natural way to approach the so-called tissue clock paradigm is neuroimaging. However, PET is not feasible and available in the clinical setting. Thus, for both CT and MRI, perfusion imaging strategies have been developed. In these cases, the penumbra was coined tissue-at-risk and if tissue-at-risk was detected, this status was termed mismatch. Other paradigms like the mismatch between clinical deficit and small stroke lesion were introduced. The mismatch paradigm unfortunately never fulfilled its promise for i.v. thrombolysis. Heterogeneity of technical parameters, post-processing, and the inability of the scientific community to surmount these problems led to several failed studies so that today no general recommendation exists to use the mismatch concept for patient selection in i.v. thrombolysis. This was in contrast to the theoretical framework predicting an advantage for neuroimaging selection strategies.

However, for MT, recent studies indicate a benefit using neuroimaging selection strategies. The DAWN trial successfully used the clinical-stroke lesion mismatch to perform MT in a time frame of 6-24 hours after stroke onset. The DEFUSE-3 trial successfully used the perfusion mismatch concept. These results are part of official guidelines (see 1.4) and are promising for the development of the tissue-clock approach in the future.

DWI-FLAIR mismatch is another type of mismatch which has a different selection goal then the previously discussed. Up to 25% of stroke patients arrive to the hospital with unknown time from stroke onset, so-called wake-up stroke. Here, under current guidelines no therapy is offered since the time from stroke-onset is unclear. However, MRI provides a solution. In stroke MRI, - amongst others - two different sequences, diffusion-weighted-imaging (DWI) and FLUID-attenuated-inversion-recovery(FLAIR), are used. DWI shows the infarct core, the cytotoxic edema after minutes of stroke. FLAIR, however, becomes positive only once the tissue transformed to vasogenic edema after due to the breakdown of the brain-blood-barrier. In many patients, the latter occurs only hours after stroke in contrast to the lesions visible in DWI. So, a mismatch between the two modalities is indicative of an earlier stroke which should still be treatable. Moreover, if used in this way the DWI-FLAIR-mismatch is actually closer to a tissue-clock approach since we do not know the exact time from stroke-onset. And in 2018, it was shown that it is safe to treat patients with wake-up stroke based on the DWI-FLAIR-mismatch and that patients benefit from i.v. t-PA(Thomalla et al., 2018).

Another selection strategy is to use the so-called ASPECT score. It is 10 scale score which can be derived from acute CT or MR imaging and can also be derived automatically through the so-called e-ASPECT score. Essentially, the lower the ASPECT score is the more tissue is infarcted in the area of the anterior

circulation. It has been shown that higher numbers of the ASPECT score are associated with favourable outcomes and lower scores are less likely to be(Phan et al., 2018). Preliminary reports, however, indicate that MT might also be favourable in patients with lower ASPECTS(Kaesmacher Johannes et al., 2019).

It should be mentioned, however, that the intra- and interrater agreement between raters both for DWI-FLAIR-mismatch as well as ASPECT score is not very high (Fahed Robert et al., 2018). Thus, more robust and objective criteria are warranted.

1.6. Predictive Modelling for Stroke Outcome Prediction

Generally, predictive modelling in stroke is still rare in the literature. In the published literature, a popular approach is the prediction of functional outcome.

In these cases, the most often used label is the so-called modified Rankin Scale (mRS). mRS is a seven point scale indicating a range from restitutio-ad-integrum (0) to death (6). While the scale has known challenges, it remains until today the scale of choice of the primary outcome measure. Broderick et al summarized the strengths and weaknesses of the mRS scale(Broderick Joseph P. et al., 2017): The mRS "covers the entire range of functional outcomes from no symptoms to death, its categories are intuitive and easily grasped by both clinicians and patients, its concurrent validity is demonstrated by strong correlation with measures of stroke pathology (eg, infarct volumes) and agreement with other stroke scales, and its use has demarcated effective and ineffective acute stroke therapies in trials with appropriately powered sample sizes. With a limited number of levels, the mRS may be less responsive to change than some other stroke scales; however, a single-point change on the mRS is clinically relevant. A limitation of the mRS has been the subjective determination between categories and the reproducibility of the score by examiners and patients.".

An improved alternative to the classic mRS is a utility-weighted (UW) version, the so-called UW-mRS. Here, the focus is on utility, i.e. patient-centred quality of life measures. A utility of 1 represents excellent health. Broderick summarizes several approaches where each of the mRS scores was assigned a utility score(Broderick Joseph P. et al., 2017):

Average utility score per mRS score

mRS	0	1	2	3	4	5	6
utility	1	0.91	0.76	0.64	0.32	0	0

Analyses showed that UW-mRS demonstrated advantages over the ordinal scale as well as dichotomized analyses. Another advantage of UW-mRS is the ability to calculate quality-adjusted-life-years (QALY) from the measure.

Thus, the UW-mRS solves the problem where the original mRS did not translate well to the question "what does that mean to the patient". Naturally, the advantages and disadvantages of the scale remain since only a numerical transformation is done. Despite the shortcomings of the mRS score, only a few attempts have been made to develop new scoring systems. This is surely also based on the notion that the validity of the mRS is sufficient and changes in its effect size can be validly used to determine the efficacy of new stroke treatments (Broderick Joseph P. et al., 2017). Other works have pointed out that all used scales, the NIHSS, the Barthel index and the mRS are all correlated to each other and thus capture different aspects of the latent variable "stroke recovery"(Saver Jeffrey L., 2011).

For predictive modelling, usually a dichotomized mRS outcome is used. This can be explained by the fact that simpler models can be used with a dichotomized outcome and multi-label classification is often not feasible with the - comparatively - low numbers of available patients in the medical domain. Also, for individual predictive modelling the interest is not to find all shifts between the groups of the ordinal scale like it is desirable for the comparison of treatment options. In contrast, we would like to know if the patient will have a desirable outcome or not.

First reports performing predictive modelling - using traditional linear regression and mRS as outcome measure- date back as far as 2006. These works and follow up works using more advance algorithms have established - at least for i.v.-thrombolysis - that non-treatment features exert the strongest effect on functional stroke outcome. Age and the baseline NIHSS have shown numerous times to be most predictive for functional outcome after stroke. Thus e.g., for comparative studies baseline heterogeneity of the groups needs to be adjusted for(Saver Jeffrey L., 2011).

For MT, only one published work has explored predictive modelling. Van Os et al. tested a variety of algorithm using the MR CLEAN study data for 3 months dichotomous outcome prediction. They found moderate prediction both by advanced algorithms as well as classic logistic regression approaches. The study did not include imaging as input. There is a lack of explorations into predictive modelling using MT data, especially integrating clinical data and imaging data.

1.7. Precision Medicine in Stroke

Precision medicine is a form of health care that emerged in the past years that relies on data, algorithms and precision molecular tools to offer individualized care for patients(Dzau and Ginsburg, 2016). Its goal is to give insight into mechanisms of disease, treatment and prevention. By treating the patient as an individual, the attending physician is able to consider variations in pathophysiology, genome and anatomical variances. This can improve outcomes and reduce healthcare costs. Precision medicine relies on the aggregation, integration and analysis of data in a computational “learning network”(Dzau and Ginsburg, 2016). It, therefore, requires interdisciplinary cooperation at the crossroads of medicine, statistics and computer science. One particularly promising approach is the use of machine learning artificial intelligence, particularly deep learning. Stroke has a complex pathophysiology comprising medical and environmental factors(Hinman et al., 2017) and is, therefore, a suitable candidate for precision medicine(Rostanski and Marshall, 2016). Different types of data like clinical and imaging data are available for ischemic stroke. Additionally, given its high prevalence, a lot of data is routinely acquired and can be made available. A precision medicine approach can thus integrate this data and offer higher performance for treatment decision making and outcome prediction.

A natural candidate for the application of these techniques is neuroimaging. Routinely, several images are acquired for each stroke patient. The common modalities are CT and MRI. Here, structural images including vessels and sometimes functional images, e.g. about perfusion, are acquired. While these images are nowadays visually analyzed to identify the presence of stroke or to estimate the perfusion deficit, their quantitative properties are not routinely assessed. A major reason for this is the lack of methods with which these changes can be automatically and quickly translated into meaningful features. This, however, is the prerequisite to finally use this highly informative available information - that is currently ignored in the clinical setting - to improve stroke treatment. Here, fully automated pipeline based on deep learning technology will automatically extract important stroke imaging

properties and translate them into measures for precision medicine. Together with clinical data and so far untapped sources of information such as genetic markers unprecedented accuracy can be achieved. However, so far predictive modelling approaches for stroke have not yet tapped into the potential of precision medicine. First, precision medicine approaches need to follow current best practice guidelines, otherwise their predictive value is highly limited. In stroke, this means that predictive modelling for precision medicine needs to be performed with current MT data. Most available databases, however, stem from the pre-MT era.

Also, modern machine learning approaches need to be able to incorporate data from multiple sources in different formats. For the best performing algorithms like deep learning, we need to tailor specific architectures - so-called multistream architectures - which can integrate different inputs into one predictive model.

Further, prediction models need to be able to extract the necessary features automatically from the given data. Decision support systems integrated into the clinical workflow cannot rely on any manual input.

Moreover, it is important to consider that precision medicine approaches will be used for every patient. however, it is known that study data does not necessarily represent the population of real-world patients due to constraints like the need for formal consent, specific populations frequenting university hospitals and others. It is important to show the capacity of predictive models to maintain their predictive performance in cohorts of real-world data.

Lastly, the label used must be informative for patients and capture patient-centred quality-of-life (QoL) measures which are relevant for patients. Here, research into new QoL measures as outlined in the goals of P4Q is warranted and should involve patient-reported outcomes.

1.8. Design of Predictive Architectures for Precision Medicine in Stroke

Taken together, the design of predictive architectures for precision medicine in stroke needs to:

- use current mechanical thrombectomy data
- use a wide variety of features from different sources to account for the heterogeneity and individuality of stroke
- be able to integrate a wide variety of features into a single model
- be independent of any manual input
- be validated using independent real-world mechanical thrombectomy data cohorts
- research into new QoL markers for stroke outcome with patient-reported outcomes

1.9. Rationale for Data Collection

With the current study we would like to target the points mentioned in 1.8.

By gathering current MT-data from two different hospitals, we will be able to capture the current state-of-the-art. We will use this data to develop modern multi-stream deep-learning architectures integrating a wide variety of clinical and imaging features into a single predictive model without the need for manual feature extraction. Moreover, we will gather commonly used outcome measures in stroke as well as general QoL outcome measures and will explore combining these into new labels.

Lastly, we will use the Open Stroke Data Platform to incentivize other centres to provide real world MT thrombectomy data to validate our predictive models.

2. Study Design

2.1. General information

Study Title	P4Q-AS		
Clinical Phase	Acute Treatment		
Design	Observational bicentral		
Participants	Ischemic stroke patients above 18 years of age who receive mechanical thrombectomy in routine care		
Sample Size	300		
Planned Period	M19-M36 (January 2020 - April 2021) 1 year + 3M FU		
Planned Recruitment period	M19-M33 (January 2020 - December 2020)		
	Objectives	Outcome Measures	Timepoint(s)
Primary	The development of new QoL markers for short and long term outcome after stroke and mechanical thrombectomy with patient-reported outcomes	mRS, Barthel index, NIHSS, QoL markers	3 months after stroke

Secondary	The development of novel precision medicine machine learning and deep learning architectures integrating various different clinical and radiological input variables for the prediction of QoL markers at 3 months post stroke	n.a.	n.a.

2.2. Objectives and Outcome Measures

Objectives	**Outcome Measures**	**Timepoint(s) of evaluation**
Primary Objective The development of new QoL markers for short and long term outcome after stroke and mechanical thrombectomy with patient reported outcomes	mRS, Barthel index, NIHSS, QoL markers	3 months after stroke

Secondary Objective The development of novel precision medicine machine learning and deep learning architectures integrating various different clinical and radiological input variables for the prediction of QoL markers at 3 months post stroke	n.a.	n.a.

2.3. Study Design

1 year one-armed non-interventional observational study.

2.4. Participants Description

2.4.1. Study Participants

Participants are routinely treated acute stroke patients at two centres, the Charité university hospital Berlin as well as the Johanna-Etienne-hospital in Neuss, Germany.

2.4.2. Inclusion criteria

- age >= 18 years
- acute ischemic stroke, radiologically proven
- first-time stroke
- treatment with mechanical thrombectomy
- signed informed consent

2.4.3. Exclusion criteria

- patients with diagnosed stroke and MT treatment, but a different diagnosis is established post hoc

2.5. Study Procedures

All study procedures will be performed according to GCP and only after obtaining ethics approval (see point 2.5.2).

2.5.1. Recruitment

Recruitment will be performed by professional study personnel initiated by the study coordinator. Recruitment will take place at the ER services of both Charité Berlin and the Johanna-Etienne-Hospital Neuss. Patients arriving with acute ischemic stroke and the routine clinical decision to perform mechanical thrombectomy will be screened and if eligible their informed consent will be obtained by either the patient or their next of kin.

2.5.2. Regulatory clearance

The study will need approval by the ethics commitee of Charité University Hospital Berlin as well as the ethics committee of the Heinrich-Heine-University Düsseldorf. The ethics committees will review the study protocol, the Study Informative Sheet (SIS), the Informed Consent (IC) and the Informed Questionnaire (IQ) of the study. The Study Informative Sheet explains the study to the participants, the expected effects, possible complications and what the study implies to the patient. In the Informed Questionnaire of the study, several questions are asked to the participant to assess if he/she understood the informed consent and its implications.

In case of legal incapacity, a representative or tutor of the patient will sign SIS, IC and IQ.

2.5.3. Baseline assessments

Baseline clinical variables will be assessed as outlined in WP3 with harmonized variable names.

2.5.4. Data Management

Data will be collected through an electronic case report form (eCRF) and for data after dismissal in addition with an electronic patient-reported outcome framework (ePRO). As outlined in Deliverable D5.1 appropriate frameworks have been selected and are ready to use. The features of the ePRO and the eCRF frameworks are described in D5.1.

From there, study data will be transferred to the data repository (Data warehouse) for further processing and predictive modelling as outlined in the deliverables of WP2. A preliminary database scheme can be found in the appendix.

2.5.5. Withdrawal of Participants

During the course of the study, a participant may withdraw early from it at any time and request deletion of their data. In addition, the Investigator may discontinue a participant from the study at any time if the Investigator considers it necessary. Provide justification for any procedures and observations that will be required following a complete withdrawal.

2.5.6. Definition of End of Study

The end of the study is defined as the follow-up of the last patient in the 3M FU visit.

2.6. Study Intervention

No interventions are planned in P4Q-AS.

2.7. Statistical Analysis

2.7.1. Description of statistical methods

Collected features and label data will be used to build predictive models. Here, baseline algorithms like logistic regression will be used as well as advanced algorithms like decision trees, tree boosting and artificial neural nets. Here, Python will be used applying standard packages like sci-kit learn, tensorflow, keras and pytorch.

Data will be described with mean and standard deviation in case of normal distributions or median and IQR for continuous variables. For other variables, histograms will be used. Exploratory data analysis will be performed with Python and R.

2.7.2. Sample Size calculation

For predictive modelling, the sample size needs to be estimated based on prior experience. Sample size calculations like for frequentist comparisons are not readily available. For a study of the planned size, the planned architectures and the primary objective of prediction of dichotomized outcome a sample size of 300 can be deemed sufficient.

2.8. Ethical and regulatory considerations

The Investigator will ensure that this study is conducted in accordance with the Ethical regulations, following clearance described in section 2.5.2.

3. References

Aguiar de Sousa, D., von Martial, R., Abilleira, S., Gattringer, T., Kobayashi, A., Gallofré, M., Fazekas, F., Szikora, I., Feigin, V., Caso, V., Fischer, U., 2019. Access to and delivery of acute ischaemic stroke treatments: A survey of national scientific societies and stroke experts in 44 European countries. Eur. Stroke J. 4, 13–28. https://doi.org/10.1177/2396987318786023

Bang, O.Y., Saver, J.L., Buck, B.H., Alger, J.R., Starkman, S., Ovbiagele, B., Kim, D., Jahan, R., Duckwiler, G.R., Yoon, S.R., Viñuela, F., Liebeskind, D.S., 2008. Impact of collateral flow on tissue fate in acute ischaemic stroke. J. Neurol. Neurosurg. Psychiatry 79, 625–629. https://doi.org/10.1136/jnnp.2007.132100

Béjot, Y., Bailly, H., Durier, J., Giroud, M., 2016. Epidemiology of stroke in Europe and trends for the 21st century. Presse Médicale, QMR Stroke 45, e391–e398. https://doi.org/10.1016/j.lpm.2016.10.003

Broderick Joseph P., Adeoye Opeolu, Elm Jordan, 2017. Evolution of the Modified Rankin Scale and Its Use in Future Stroke Trials. Stroke 48, 2007–2012. https://doi.org/10.1161/STROKEAHA.117.017866

Church, E.W., Gundersen, A., Glantz, M.J., Simon, S.D., 2017. Number needed to treat for stroke

thrombectomy based on a systematic review and meta-analysis. Clin. Neurol. Neurosurg. 156, 83–88. https://doi.org/10.1016/j.clineuro.2017.03.005

Dzau, V.J., Ginsburg, G.S., 2016. Realizing the Full Potential of Precision Medicine in Health and Health Care. JAMA 316, 1659–1660. https://doi.org/10.1001/jama.2016.14117

Fahed Robert, Lecler Augustin, Sabben Candice, Khoury Naim, Ducroux Célina, Chalumeau Vanessa, Botta Daniele, Kalsoum Erwah, Boisseau William, Duron Loïc, Cabral Dominique, Koskas Patricia, Benaïssa Azzedine, Koulakian Hasmik, Obadia Michael, Maïer Benjamin, Weisenburger-Lile David, Lapergue Bertrand, Wang Adrien, Redjem Hocine, Ciccio Gabriele, Smajda Stanislas, Desilles Jean-Philippe, Mazighi Mikaël, Ben Maacha Malek, Akkari Inès, Zuber Kevin, Blanc Raphaël, Raymond Jean, Piotin Michel, 2018. DWI-ASPECTS (Diffusion-Weighted Imaging–Alberta Stroke Program Early Computed Tomography Scores) and DWI-FLAIR (Diffusion-Weighted Imaging–Fluid Attenuated Inversion Recovery) Mismatch in Thrombectomy Candidates. Stroke 49, 223–227. https://doi.org/10.1161/STROKEAHA.117.019508

Hacke, W., Schwab, S., Horn, M., Spranger, M., Georgia, M.D., Kummer, R. von, 1996. "Malignant" Middle Cerebral Artery Territory Infarction: Clinical Course and Prognostic Signs. Arch. Neurol. 53, 309–315. https://doi.org/10.1001/archneur.1996.00550040037012

Hankey, G.J., 2003. Long-Term Outcome after Ischaemic Stroke/Transient Ischaemic Attack. Cerebrovasc. Dis. 16, 14–19. https://doi.org/10.1159/000069936

Heiss, W.-D., Huber, M., Fink, G.R., Herholz, K., Pietrzyk, U., Wagner, R., Wienhard, K., 1992. Progressive Derangement of Periinfarct Viable Tissue in Ischemic Stroke. J. Cereb. Blood Flow Metab. 12, 193–203. https://doi.org/10.1038/jcbfm.1992.29

Hinman, J.D., Rost, N.S., Leung, T.W., Montaner, J., Muir, K.W., Brown, S., Arenillas, J.F., Feldmann, E., Liebeskind, D.S., 2017. Principles of precision medicine in stroke. J Neurol Neurosurg Psychiatry 88, 54–61. https://doi.org/10.1136/jnnp-2016-314587

Jung, S., Gilgen, M., Slotboom, J., El-Koussy, M., Zubler, C., Kiefer, C., Luedi, R., Mono, M.-L., Heldner, M.R., Weck, A., Mordasini, P., Schroth, G., Mattle, H.P., Arnold, M., Gralla, J., Fischer, U., 2013. Factors that determine penumbral tissue loss in acute ischaemic stroke. Brain 136, 3554–3560. https://doi.org/10.1093/brain/awt246

Kaesmacher Johannes, Chaloulos-Iakovidis Panagiotis, Panos Leonidas, Mordasini Pasquale, Michel Patrik, Hajdu Steven D., Ribo Marc, Requena Manuel, Maegerlein Christian, Friedrich Benjamin, Costalat Vincent, Benali Amel, Pierot Laurent, Gawlitza Matthias, Schaafsma Joanna, Mendes Pereira Vitor, Gralla Jan, Fischer Urs, 2019. Mechanical Thrombectomy in Ischemic Stroke Patients With Alberta Stroke Program Early Computed Tomography Score 0–5. Stroke 50, 880–888. https://doi.org/10.1161/STROKEAHA.118.023465

Lauritzen, M., Dreier, J.P., Fabricius, M., Hartings, J.A., Graf, R., Strong, A.J., 2011. Clinical Relevance of Cortical Spreading Depression in Neurological Disorders: Migraine, Malignant Stroke, Subarachnoid and Intracranial Hemorrhage, and Traumatic Brain Injury. J. Cereb. Blood Flow Metab. 31, 17–35. https://doi.org/10.1038/jcbfm.2010.191

Leng, X., Fang, H., Leung, T.W.H., Mao, C., Miao, Z., Liu, L., Wong, K.S., Liebeskind, D.S., 2016. Impact of collaterals on the efficacy and safety of endovascular treatment in acute ischaemic stroke: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry 87, 537–544. https://doi.org/10.1136/jnnp-2015-310965

Meairs, S., Wahlgren, N., Dirnagl, U., Lindvall, O., Rothwell, P., Baron, J.-C., Hossmann, K., Engelhardt, B., Ferro, J., McCulloch, J., Kaste, M., Endres, M., Koistinaho, J., Planas, A., Vivien, D., Dijkhuizen, R., Czlonkowska, A., Hagen, A., Evans, A., Libero, G.D., Nagy, Z., Rastenyte, D., Reess, J., Davalos, A., Lenzi, G.L., Amarenco, P., Hennerici, M., 2006. Stroke Research Priorities for the Next Decade – A Representative View of the European Scientific Community. Cerebrovasc. Dis. 22, 75–82. https://doi.org/10.1159/000093098

Phan, K., Saleh, S., Dmytriw, A.A., Maingard, J., Barras, C., Hirsch, J.A., Kok, H.K., Brooks, M., Chandra,

R.V., Asadi, H., 2018. Influence of ASPECTS and endovascular thrombectomy in acute ischemic stroke: a meta-analysis. J. NeuroInterventional Surg. neurintsurg-2018-014250. https://doi.org/10.1136/neurintsurg-2018-014250

Powers William J., Rabinstein Alejandro A., Ackerson Teri, Adeoye Opeolu M., Bambakidis Nicholas C., Becker Kyra, Biller José, Brown Michael, Demaerschalk Bart M., Hoh Brian, Jauch Edward C., Kidwell Chelsea S., Leslie-Mazwi Thabele M., Ovbiagele Bruce, Scott Phillip A., Sheth Kevin N., Southerland Andrew M., Summers Deborah V., Tirschwell David L., 2018. 2018 Guidelines for the Early Management of Patients With Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke 49, e46–e99. https://doi.org/10.1161/STR.0000000000000158

Rostanski, S.K., Marshall, R.S., 2016. Precision Medicine for Ischemic Stroke. JAMA Neurol. 73, 773–774. https://doi.org/10.1001/jamaneurol.2016.0087

Saver Jeffrey L., 2011. Optimal End Points for Acute Stroke Therapy Trials. Stroke 42, 2356–2362. https://doi.org/10.1161/STROKEAHA.111.619122

Sobesky, J., 2012. Refining the mismatch concept in acute stroke: lessons learned from PET and MRI. J. Cereb. Blood Flow Metab. 32, 1416–1425. https://doi.org/10.1038/jcbfm.2012.54

Thomalla, G., Simonsen, C.Z., Boutitie, F., Andersen, G., Berthezene, Y., Cheng, B., Cheripelli, B., Cho, T.-H., Fazekas, F., Fiehler, J., Ford, I., Galinovic, I., Gellissen, S., Golsari, A., Gregori, J., Günther, M., Guibernau, J., Häusler, K.G., Hennerici, M., Kemmling, A., Marstrand, J., Modrau, B., Neeb, L., Ossa, N.P. de la, Puig, J., Ringleb, P., Roy, P., Scheel, E., Schonewille, W., Serena, J., Sunaert, S., Villringer, K., Wouters, A., Thijs, V., Ebinger, M., Endres, M., Fiebach, J.B., Lemmens, R., Muir, K.W., Nighoghossian, N., Pedraza, S., Gerloff, C., 2018. MRI-Guided Thrombolysis for Stroke with Unknown Time of Onset. N. Engl. J. Med. https://doi.org/10.1056/NEJMoa1804355

Turc, G., Bhogal, P., Fischer, U., Khatri, P., Lobotesis, K., Mazighi, M., Schellinger, P.D., Toni, D., de Vries, J., White, P., Fiehler, J., 2019. European Stroke Organisation (ESO) – European Society for Minimally Invasive Neurological Therapy (ESMINT) Guidelines on Mechanical Thrombectomy in Acute Ischaemic StrokeEndorsed by Stroke Alliance for Europe (SAFE). Eur. Stroke J. 4, 6–12. https://doi.org/10.1177/2396987319832140

von Kummer, R., 2019. Treatment of ischemic stroke beyond 3 hours: is time really brain? Neuroradiology 61, 115–117. https://doi.org/10.1007/s00234-018-2122-1

Xing, C., Arai, K., Lo, E.H., Hommel, M., 2012. Pathophysiologic Cascades in Ischemic Stroke. Int. J. Stroke 7, 378–385. https://doi.org/10.1111/j.1747-4949.2012.00839.x

4. Appendix - Data Base Scheme

Introduction

Our database schema consists of

1. Usual tables with one or more attributes. Prefix: **T**
2. Catalogue tables which represent categorical values of specific properties. Prefix: **T_CAT**
3. Mapping tables which connect a record in a usual table to a categorical value in a catalogue table. Prefix: **T_MAP**

Attributes can have prefixes as well:

- **PK:** primary key, unique identifier (in record's own table) of a record

- *FK:* foreign key, reference to a key (often the PK) in a different table

The *PK_code* attribute often found in catalogue tables can for example be a common abbreviation such as *TIA.*

Special tables:

CAT_truth_values:
This table is used to represent values which are typical in a study questionnaire.
Value examples are *Not measured, not asked, not sure, not known, yes, no.*

- **PK_code:** varchar
- property: varchar
- value (or description): varchar

Attributes which are a varchar foreign key (*FK_truth_value_code* would be a more correct but less understandable name probably) to this table are called ***presence: TruthValue*** in the diagram and in the following document.

units: Table to represent different types of units. Not used at the moment.

Tables:

patients:

- **PK_id**: integer

demographics:

- **PK_id**: integer
- firstname: varchar
- lastname: varchar
- yearOfBirth: integer
- monthOfBirth: integer
- dayOfBirth: integer
- countryOfBirth: integer
- age: integer
- ethnicity: varchar
- dateOfDeath: Date
- sex: Boolean
- weight: float
- height: integer
- *FK_patient_id*: integer - **reference to T_patients**

We have not decided yet which information about a patient's age we would store and therefore included both attributes (***birthdate*** and ***age***) in the schema. A similar decision has to be made regarding ***countryOfBirth*** and ***ethnicity***

Value semantics:
sex:

studies:

This table is used in order to assign multiple patients to one or more studies.

- **PK_id:** integer
- name: varchar

participations:

- **PK_id:** integer
- id_in_study: integer - **The internal id of the referenced patient in the referenced study**
- FK_patient_id: integer
- FK_study_id: integer

projects:

This table allows patients to be assigned to one or more projects.

- **PK_name**: integer

The relation between projects and patients is stored using the **MAP_patients_projects** table.

events:

This table is used for events/incidents/occurences such as Admissions or Reinfarctions (see EventTypes below)

- **PK_id:** integer
- datetime:
- time_label: If no timestamp is provided a time_label such as 'day 1' or 'admission' can be given
- *FK_event_type_code:* varchar - Reference to an event type (see below)
- *FK_patient_id:* integer - Reference to the patient

CAT_event_types:

Event types such as:

- Admisson
- StrokeOnset
- Complication
- Edema
- Reinfarction
- Pneumonia
- Symptomatic Bleeding
- Discharge

Attributes of this table are:

- **PK_code:** integer
- description: varchar

images: Table for metadata of radiological images

- **PK_id:** integer
- time: timestamp
- file_location: varchar
- *FK_patient_id:* integer references patient
- An *image* can be linked to one or more *image_properties* by the **MAP_images_image_properties** mapping table.

CAT_image_properties:

I.e.: Modality, Space, Type (Sequence), Quality, Processing Steps

- **PK_code:** varchar
- property: varchar
- value (or description): varchar

An example entry in this table would be ('T1', 'type', 'T1 weighted')

findings:

This table in combination with *CAT_finding_types* and *CAT_territories* is used to store information about radiological findings.

The *findings* table stores references to a finding_type, a territory and the patient. It also stores a *TruthValue* as the *presence* attribute indicating wether or not a finding is present (or not measured for example).

Quantifiable findings can be stored using the *radiological_scores* table (see below).

CAT_finding_types:

Examples: Occlusion, Lesion, Ischemia, Vessel malformation, Bleeding, Microangiopathy

- **PK_code:** varchar
- name: varchar - a name describing the finding

radiological_scores:

Used for saving properties such as a lesion volume.

- measurement: varchar - property which was measured
- value: float
- *FK_finding_id:* reference to a record in the *findings* table

CAT_territories:

Provides categorical options for storing the territory of a *finding*, a *vascular_intervention* or a *cerebrovascular_event*

- **PK_code:** integer
- category: varchar - I.e. vessel, side, lobe, segment
- name: varchar - value of categorical propertie, i.e. MCA, ACA, right, left, frontal, occipital

It is **important to understand** that the exact region of a finding (or something else) is stored by mapping multiple *CAT_territories* records to one *finding* record using the **MAP_findings_territories** table.

For example given the following *CAT_territories* table

(MCA, vessel, MCA),
(r, side (transversal), right)

storing 'MCA right' for a finding with id 7 would require storing two entries in the *MAP_findings_territories* table: (7, MCA), (7, r)

One could of course also create entries in *T_CAT_territories* for every possible combination of options.

cerebrovascular_events:
Stores information about the acute and previous events

- **PK_id:** integer
- datetime: timestamp
- acute: boolean - whether this record described the acute event resulting in hospitalization
- localisation_unclear: boolean - whether the location can be determined / is stored
- symptom_length: varchar - string describing the symptom length (3m, 1h, 2d)
- *FK_cerebrovascular_event_type_code:* reference to *CAT_cerebrovascular_event_types*

The affected territory is encoded by entries in the **MAP_cerebrovascular_events_territories** mapping table.

CAT_cerebrovascular_event_types:
Encodes events such as Infarct, TIA, ICB, SAB, SVT

- **PK_code:** varchar
- name: varchar - name, description or abbreviation

neurological_deficits:
Textual descriptions of neurological deficits

- description: text
- duration: intervall
- *FK_patient_id:* reference to affected patient

clinical_parameter_score:
Scores for numerical clinical parameters such as NIHSS, mRS, Oxygen saturation, TOAST, heart rate
The parameter type is encoded by the *CAT_clinical_parameter_types* table (see below).

- **PK_id:** integer
- datetime: timestamp
- time_label: varchar - i.e. day1, admission, etc.
- value: float
- *FK_clinical_parameter_type_code:* varchar - reference to parameter type
- *FK_patient_id*: integer - reference to patient

CAT_clinical_parameter_types:
i.e. NIHSS, mRS, Oxygen saturation, TOAST, heart rate

- **PK_code:** varchar
- name: varchar - name, description or abbreviation

medication:

- **PK_id:** integer
- started: timestamp
- stopped: timestamp
- dosis: float
- drug_name: varchar
- regarding_latest_stroke: boolean - wether the treatment is used for the latest stroke
- *FK_patient_id*: integer - reference to patient

MAP_diagnoses:

- **PK_id**: integer
- presence: TruthValue

- *FK_patient_id:* integer
- *FK_disease_code:* varchar

CAT_diseases:
list of diseases. used for referencing in *T_diagnoses*
- **PK_code:** varchar
- name: varchar - name, description or abbreviation

MAP_patients_risk_factors:
- **PK_id**: integer
- presence: TruthValue
- *FK_patient_id:* integer
- *FK_risk_factor _code:* varchar

CAT_risk_factors:
list of risk factors. used for referencing in *T_MAP_patients_risk_factors.*
- **PK_code:** varchar
- name: varchar - name, description or abbreviation

vascular_interventions:
- **PK_id**: integers
- presence: TruthValue
- datetime: timestamp
- *FK_patient_id:* integer - references patient
- *FK_territory_code*: varchar - references territory (see above)
- *FK_treatment_type_code:* varchar - references territory (see above)

CAT_treatment_types:
list of treatment types. i.e. lysis (ia, iv), stents, arterectomy
- **PK_code:** varchar
- name: varchar - name, description or abbreviation