

UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Klinik für Neuroradiologische Diagnostik und Intervention

Professor Dr. med. Jens Fiehler

Optimierte Modellentwicklung zur stratifizierten Prognosebestimmung bei posteriorem Apoplex in einem „Real World“ Datensatz– das pcBATASs–Prognosemodell

Dissertation

zur Erlangung des Grades eines Doktors der Medizin
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Maximilian Groffmann
aus Berlin

Hamburg 2023

(wird von der Medizinischen Fakultät ausgefüllt)

**Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 22.05.2024**

**Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.**

Prüfungsausschuss, der/die Vorsitzende: PD Dr. Lasse Dührsen

Prüfungsausschuss, zweite/r Gutachter/in: PD Dr. Uta Hannig

Inhaltsverzeichnis

I.	Abbildungsverzeichnis.....	V
II.	Tabellenverzeichnis.....	VI
1.1	Arbeitshypothese und zentrale Fragestellung	1
1.2	Einleitung.....	2
1.2.1	Grundlagen der Prognoseforschung im Gesundheitswesen	2
1.2.2	Entwicklung von Prognosemodellen	8
1.2.3	Multiple Imputation (MI)	16
1.2.4	Evaluation von Prognosemodellen.....	22
1.2.5	Prognoseforschung Posteriorer Schlaganfall.....	26
1.2.6	Klinische Endpunkte in der Schlaganfallforschung	27
1.2.7	Typ-1-PROGRESS-Studien posteriorer Schlaganfall	28
1.2.8	Typ-2-PROGRESS – Studien posteriorer Schlaganfall.....	29
1.2.9	Typ-3-PROGRESS-Studien posteriorer Schlaganfall	41
1.2.10	Typ-4-PROGRESS Studien posteriorer Schlaganfall	43
1.3	Material und Methoden.....	44
1.3.1	Patienten.....	44
1.3.2	Bildgebung.....	45
1.3.3	Bildanalyse.....	45
1.3.4	Klinische und andere paraklinische Basisparameter.....	47
1.3.5	Statistische Auswertung.....	48
1.4	Ergebnisse	57
1.4.1	Native Datenanalyse und Imputationsmodellentwicklung	57
1.4.2	Ergebnisse der deskriptiven Analyse des imputierten Datensatzes und Prognosemodellentwicklung.....	71
1.4.3	Ergebnisse Multikollinearitätsanalyse und Separationsanalyse	73
1.4.4	Ergebnisse: Modellparameter Regressionsmodelle.....	73
1.4.5	Ergebnisse der Ereigniswahrscheinlichkeitsdichtegraphen	75
1.4.6	Ergebnisse der Leistungsfähigkeit des Modells	78
1.4.7	Ergebnisse der Entscheidungskurvenanalyse	81
1.5	Diskussion	85
1.5.1	Prognosemodellentwicklung	86
1.5.2	Klinischer Nutzen des pcBATASs Prognosemodells	92
1.5.3	Limitation der Arbeit.....	96

1.6	Zusammenfassung	97
1.7	Abstrakt Englisch.....	99
1.8	Glossar	100
1.9	Literaturverzeichnis	103
1.10	Lebenslauf.....	123
1.11	Eidesstattliche Versicherung	126

I. Abbildungsverzeichnis

Abbildung 1: Konventionellen ML-Algorithmus	10
Abbildung 2: Native CT „Patient niedriger pcBATASs“	46
Abbildung 3: Kontrastmittel gestützte CT „Patient niedriger pcBATASs“	47
Abbildung 4: Algorithmus „klassisches logistischen Regressionsmodell“	52
Abbildung 5: Algorithmus „Ridge-Regressionsmodell“	54
Abbildung 6: md-Pattern- Graph.....	58
Abbildung 7: Kombiniertes Histogramm, Dichtediagramm, QQ-Plots 1	59
Abbildung 8: Kombiniertes Histogramm, Dichtediagramm, QQ-Plots 2.....	60
Abbildung 9: Kombiniertes Histogramm, Dichtediagramm, QQ-Plots 3.....	61
Abbildung 10: Kombiniertes Histogramm, Dichtediagramm, QQ-Plots 4.....	62
Abbildung 11: Flux-Graphen.....	64
Abbildung 12: Trace-Line-Plots 1	66
Abbildung 13: Trace-Line-Plots 2	66
Abbildung 14: Trace-Line-Plots 3	67
Abbildung 15: Trace-Line-Plots 4	67
Abbildung 16: Trace-Line-Plots 5	68
Abbildung 17: Trace-Line-Plots 6	68
Abbildung 18: Trace-Line-Plots 7	69
Abbildung 19: Stripplots 1.....	70
Abbildung 20: Stripplots 2.....	70
Abbildung 21: Stripplots 3.....	71
Abbildung 22: Ereigniswahrscheinlichkeitsdichtegraph 1	76
Abbildung 23: Ereigniswahrscheinlichkeitsdichtegraph 2.	76
Abbildung 24: Ereigniswahrscheinlichkeitsdichtegraph 3.	77
Abbildung 25: Flexible Kalibrierungskurven 1.....	80
Abbildung 26: Flexible Kalibrierungskurven 2.....	80
Abbildung 27: Flexible Kalibrierungskurven 3.....	81
Abbildung 28: Entscheidungskurvenanalyse 1	82
Abbildung 29: Entscheidungskurvenanalyse 2	83
Abbildung 30: Entscheidungskurvenanalyse 3	83

II. Tabellenverzeichnis

Tabelle 1: Deskriptive Auswertung des nativen Datensatzes.	57
Tabelle 2: Deskriptive Auswertung des imputierten Datensatzes	72
Tabelle 3: Modellparameter der klassischen logistischen Regressionsanalyse.....	73
Tabelle 4: Optimierte Beta-Koeffizienten des logistischen Regressionsmodells.....	74
Tabelle 5: Modellparameter der Ridge-Regressionsanalyse	74
Tabelle 6: Modellparameter der FLAC-Regressionsanalyse	75
Tabelle 7: Modellevaluation klassische logistische Regression.....	79
Tabelle 8: Modellevaluation Ridge-Regression	79
Tabelle 9: Modellevaluation FLAC-Regression.....	79

1.1 Arbeitshypothese und zentrale Fragestellung

Die Auswahl von geeigneten Patienten für die Schlaganfallbehandlung ist entscheidend, da verheerende Komplikationen den Nutzen der Behandlung beeinträchtigen können. Komplexe, bereits etablierte Prognosemodelle für isolierte Basilarisverschlüsse und für Kohorten mit unspezifischem posterioren Schlaganfall sind bislang nicht ausreichend oder nur teilweise evaluiert und validiert oder bilden nur unvollständig wichtige Einflussfaktoren auf den klinischen Verlauf erkrankter Patienten ab. Die Kombination aus präinterventionell-klinischen und den präinterventionell-neuroradiologischen Parametern pcASPECTS und BATMAN in einem hybriden Regressionsmodell ist ein geeignetes Werkzeug zur Prognoseabschätzung eines guten funktionellen Überlebens, definiert als mRS („modified Rankin Scale“) ≤ 2 , thrombektomierter Patienten bei posteriorem Schlaganfall am Tag 90 nach Diagnosestellung.

1.2 Einleitung

Dieser Abschnitt ist inhaltlich in drei Teile untergliedert. Zunächst wird allgemein in das Thema der Prognoseforschung im Gesundheitswesen eingeführt und aktuelle Evidenz sowie verfügbare Leitlinien zu diesem Sachverhalt zusammengetragen. Im zweiten Abschnitt werden die Grundlagen für die in dieser Arbeit verwendeten statistischen Methoden beschrieben. Es wird nicht in vollem Umfang auf die einzelnen technischen Grundlagen eingegangen, hierfür ist Sekundärliteratur zu verwenden. Es werden lediglich einzelne Punkte im Detail abgehandelt, die wesentlich zum besseren Verständnis des Material- und Methodenabschnitts beitragen. Im letzten Abschnitt wird die aktuelle Evidenz auf dem Gebiet der Prognoseforschung des posterioren Schlaganfalls dargelegt.

1.2.1 Grundlagen der Prognoseforschung im Gesundheitswesen

1.2.1.1 Einführung

In der Medizin beschreibt der Begriff Prognose die Risikobeurteilung künftiger gesundheitlicher Folgen von Menschen mit einer bestimmten Krankheit oder einem bestimmten Gesundheitszustand. Prognoseforschung ist also die Untersuchung der Beziehungen zwischen zukünftigen Ereignissen (Endpunkten) bei Menschen mit einem bestimmten Ausgangszustand (Ausgangspunkt) und dient dazu die Gesundheit zu verbessern (Hemingway et al., 2013). Prognoseforschung ist von enormer Relevanz, da zurzeit weltweit mehr Menschen mit einer oder mehreren Krankheiten beziehungsweise gesundheitlichen Beeinträchtigungen leben als je zuvor (Mathers and Loncar, 2006). Aus diesem Grund sind Regierungen der ganzen Welt zunehmend daran interessiert, die Prognoseforschung zu verbessern (Whitepaper, 2010). Ebenso konzentrieren sich Forschungsförderer und Forschende zunehmend darauf, neue Maßnahmen und Technologien aus der experimentellen Forschung in die klinische Praxis und anschließend in die Gesundheitspolitik zu überführen, um neue, qualitativ hochwertige Standards für eine bessere Versorgung zu ermöglichen, und um so die Prognose von Patienten mit Hilfe einer stratifizierten Medizin zu verbessern.

Stratifizierte Medizin oder Präzisionsmedizin (engl.: „precision, personalized, individualized medicine“) beschreibt im engeren Sinne individuelle Diagnostik - und

Therapiestrategien, die auf der Grundlage von Biomarkern sowie genetischen, phänotypischen oder psychosozialen Merkmalen auf die Bedürfnisse des einzelnen Patienten zugeschnitten sind und somit einen bestimmten Patienten von anderen Patienten mit ähnlichen klinischen Merkmalen sinnvoll unterscheidet (Rothwell et al., 2005, Trusheim et al., 2007). Das Ziel ist es klinische Endpunkte individuell zu verbessern und unnötige Nebenwirkungen bei denjenigen zu minimieren, die weniger wahrscheinlich auf eine bestimmte Behandlung ansprechen. So können zum Beispiel multivariable prognostische Modelle den Kliniker dabei unterstützen, den wahrscheinlichsten Krankheitsverlauf eines Patienten vorherzusagen, ein individuelles Therapieansprechen zu quantifizieren und zusätzlich transparente Entscheidungsfindungsprozesse im Dialog mit dem Patienten (engl. „informed decision making“) zu generieren (Moons et al., 2009). Personalisierte Medizin wird als zentral für den Fortschritt der Gesundheitsfürsorge angesehen (Hamburg and Collins, 2010).

1.2.1.2 PROGRESS

Obwohl tausende medizinische Studien jedes Jahr veröffentlicht werden, die prognostische Fragestellungen thematisieren, mangelt es auf diesem Forschungsgebiet an einer einheitlichen Terminologie und einer klaren Unterscheidung der verschiedenen Aspekte der Prognoseforschung (Hemingway, 2006). Um dieses Problem zu adressieren, entwickelte im Jahr 2013 die „PROGnosis RESearch Strategy“ (PROGRESS) Gesellschaft ein Klassifikationssystem für Prognosestudien (Hemingway et al., 2013, Riley et al., 2013, Steyerberg et al., 2013, Hingorani et al., 2013). Nach PROGRESS gibt es vier unterschiedliche Studientypen im Bereich der Prognoseforschung:

- 1) „Gesamtprognosestudien (engl. „Overall Prognosis“) oder PROGRESS-Typ-1-Prognoseforschung
- 2) Prognosefaktorforschung (engl. „prognostic factor“) oder PROGRESS-Typ-2 Prognoseforschung
- 3) Prognosemodellforschung oder PROGRESS-Typ-3 Prognoseforschung
- 4) absolute Therapieeffektforschung oder PROGRESS-Typ-4 Prognoseforschung

Jeder dieser einzelnen Studientypen fordert einen unterschiedlichen Ansatz in Bezug auf Studiendesign und Analyse, um verschiedene Gesichtspunkte der Prognoseforschung zu untersuchen.

1.2.1.3 Gesamtprognoseforschung

Die Gesamtprognose nach PROGRESS definiert den wahrscheinlichsten gemittelten Wert eines Prognoseendpunktes einer Gruppe von Individuen mit einer bestimmten Erkrankung oder einem bestimmten Gesundheitszustand im Kontext des gegenwärtigen Versorgungsmanagements (Hemingway et al., 2013). Das Studiendesign einer Gesamtprognosestudie ist das einer prospektiven Kohortenstudie. Gesamtprognosestudien helfen, die Relevanz von diagnostischen Tests oder Krankheitsdefinitionen und -klassifikationen zu verstehen sowie Konsequenzen durch die Einführung neuer diagnostischer Tests und Interventionen zu monitoren. Sie bilden für eine ganz Reihe verschiedener Personengruppen eine wichtige Informationsquelle und tragen zum Verständnis von Krankheitsdefinition, krankheitsspezifischen Konsequenzen und Krankheitsverlauf bei. Für Patienten sind sie zum Beispiel relevant, um den allgemeinen Verlauf ihrer Krankheit besser zu verstehen und um die damit assoziierten Einschränkungen besser einordnen zu können. Politiker können Gesamtprognosestudien nutzen, um Ressourcen im Gesundheitswesen zu planen und um die Qualität der allgemeinen medizinischen Versorgung zu überwachen. Investoren und Forscher können sinnvoller bestimmte Krankheitsbilder priorisieren und Populationen identifizieren, deren Gesamtprognose in Bezug auf die Allgemeinheit am stärksten verbessert werden muss, um systemrelevante Forschungsprojekte entwickeln zu können. Darüber hinaus helfen Gesamtprognosestudien dabei, Prognosefaktorstudien und Prognosemodellstudien besser zu interpretieren und im Gesamtkontext einzuordnen (Riley et al., 2019c).

1.2.1.4 Prognosefaktorforschung

Die Prognose einer bestimmten Population unterscheidet sich möglicherweise von der Gesamtprognose und die individuelle Prognose unterscheidet sich möglicherweise von der Prognose einer bestimmten Population. Daher ist es wichtig, Parameter zu

identifizieren, die mit Unterschieden zwischen Individuen und zwischen Populationen in Bezug auf die Prognose assoziiert sind. Ein Prognosefaktor auch als Prädiktor, prognostische Variable oder auch prognostischer Marker (auf dem Gebiet der Krebsforschung) bezeichnet, ist eine Variable, die in einem bestimmten Stratum mit einem definierten Gesundheitszustand zu einem definierten Zeitpunkt mit einem subsequenten klinischen Verlauf (Endpunkt) assoziiert ist. Verschiedene Werte eines prognostischen Faktors sind mit einer besseren oder schlechteren Prognose assoziiert. Vom Prognosefaktor terminologisch abzugrenzen ist der Risikofaktor, der im speziellen eine bestimmte Art von Exposition bezeichnet, die mit dem gehäuften Auftreten einer Erkrankung assoziiert ist, sowie das Konzept des „Confoundings“. Ein „Confounder“, zum Beispiel im Rahmen von Therapieeffekt-Observationsstudien, verzerrt den wahren Zusammenhang zwischen einer Intervention und dem Endpunkt, wenn die Werte des Confounders in der Interventions- und Kontrollgruppe nicht ausbalanciert sind. Führt man eine grobe Suchanfrage in Pubmed nach Artikeln mit den Termen „prognostischer Faktor“ oder „prognostischer Marker“ in Abstrakt oder Titel durch, ergeben sich in einem Zeitraum von 1990 – 2010 20759 Sucheinträge. Führt man mit gleicher Fragestellung eine Suche im Zeitraum 2010 – 2020 durch, erhält man 50821 Ergebnisse (Stand 06.Juli 2022). Dieser exponentielle Anstieg an Publikationen auf dem Gebiet der Prognosefaktorforschung im letzten Jahrzehnt ist mit einem hohen Risiko für Verzerrungen der prognostischen Zusammenhänge assoziiert (Altman et al., 1994, Holländer and Sauerbrei, 2007, Kyzas et al., 2005, Riley et al., 2003, Sekula et al., 2017, Simon and Altman, 1994). Darüber hinaus bieten wenige Studien zu prognostischen Faktoren klare Aussagen zu den Implikationen der neugewonnenen Erkenntnisse im Bereich der klinischen Praxis und zukünftiger Forschung mittels speziell randomisierter, kontrollierter Studien (Riley et al., 2013).

1.2.1.5 Prognosemodellforschung

„*The Median isn't the Message*“ (Gould, 1985): Dr. Gould starb im Alter von sechzig Jahren, zwanzig Jahre nach der Diagnose eines Mesothelioms, aufgrund dessen, ihm ursprünglich eine mittlere Lebenserwartung von acht Monaten prognostiziert worden war. Im Hinblick auf die Erforschung von Prognosefaktoren auf dem Gebiet der Gesundheitsforschung zeigt dies, dass die alleinige Erforschung von Gesamtprognose und Prognosefaktoren insuffizient ist. Ein einzelner prognostischer Faktor verbessert

zwar die Abschätzungen der allgemeinen Gesamtprognose eines Individuums in einer bestimmten Population, reicht aber selten aus, um spezifische, individualisierte Aussagen zu bestimmten klinischen Endpunkten zu treffen. Es hat sich gezeigt, dass Schlaganfallpatienten mit deutlichen Ischämiefrühzeichen in der Akutbildgebung (ein Merkmal, welches grundsätzlich mit einer schlechten Prognose verbunden ist), aber mit einer guten vaskulären Versorgung durch Kollateralkreisläufe, ein deutlich besseres Überleben aufweisen als alleine mithilfe der Ischämiefrühzeichen prognostiziert (van der Hoeven et al., 2016b). Individualisierte, diagnostische und therapeutische Handlungsempfehlungen können nur auf der Grundlage multipler prognostischer Faktoren (= Prädiktoren) getroffen werden und müssen im Rahmen der Entwicklung, Validierung und Implementierung multivariabler prognostischer Modelle miteinander kombiniert betrachtet werden (Riley et al., 2019c).

Im Wesentlichen beschreibt ein Prognosemodell die Beziehung zwischen dem Vorhandensein von klinischen und/oder paraklinischen Befunden (Input-, unabhängige Variablen, Kovariate, prognostischer Faktor oder Prädiktor) zu einem bestimmten Zeitpunkt und der Wahrscheinlichkeit, dass in einem definierten Zeitintervall ein klinischer Zustand eintritt (abhängige Variable). Mit einem solchen Modell wird die Wahrscheinlichkeit ausgedrückt, dass eine Krankheit vorhanden ist oder sich entwickelt beziehungsweise, dass allgemein ein bestimmter klinischer Endpunkt (engl. „outcome“) erreicht wird. Prognostische Modelle sollen dem Kliniker helfen, Krankheitsverläufe besser einschätzen zu können und Entscheidungsfindungsprozesse im Dialog mit dem Patienten optimieren (Harrell et al., 1996, Steyerberg et al., 2013). Weiterhin werden sie zur Verbesserung der Validität von randomisiert-kontrollierten Studien (engl. „randomized-controlled-trials“; „RCT“) genutzt (Hernández et al., 2004, Roozenbeek et al., 2009) Ebenso ist der Einsatz als Balancierungsinstrument im Hinblick auf die Prognose zwischen Interventionsgruppe und Kontrollgruppe bei Studienbeginn möglich. Weiterhin werden Prognosemodelle zur Adjustierung der „Case-mix-Variation“ im Bereich der Gesundheitsforschung verwendet, um die unterschiedlichen regionalen Krankheitsverläufe von Patienten in Zentren oder Ländern und Kontinenten zu erforschen (Jarman et al., 2010, Lingsma et al., 2011).

1.2.1.6 Absolute Therapieeffektforschung

Die stratifizierte Medizin beschreibt im Gegensatz zur klassischen "empirischen" Medizin die gezielte Entwicklung von Therapiestrategien (einschließlich pharmakologischer und nicht-pharmakologischer Behandlungen, sowie genereller therapeutischer Entscheidungen für Subpopulationen von ähnlichen Patienten (Rothwell et al., 2005, Trusheim et al., 2007). Die stratifizierte medizinische Forschung verwendet absolute Risiken. Wenn ein auf einer relativen Skala gemessener Behandlungseffekt (z. B. das relative Risiko) für alle Patienten gleich ist, dann ziehen diejenigen mit dem höchsten absoluten Risiko den größten absoluten Nutzen aus der Behandlung. Wenn ein Behandlungseffekt bei einer speziellen Gesamtpopulation uneinheitlich ist, können stratifizierte Faktoren bestimmt werden (wie Biomarkerschwellenwerte oder Genotypen), die das individuelle Ansprechen auf die Behandlung vorhersagen. Die klinische Anwendung solcher Tests ist jedoch derzeit gering und es mangelt häufig an starker Evidenz für den Therapieerfolg, da Mängel im Studiendesign, in der Analyse und in der Studienergebnisdokumentation zu potenziell falschen Schlussfolgerungen führen. Ein Problem besteht darin, dass die meisten Prognosefaktoren auf einen bestimmten Krankheitsverlauf, jedoch nicht gleichzeitig auch auf ein unterschiedliches Therapieansprechen hindeuten. Das heißt, sie identifizieren Gruppen von Patienten mit unterschiedlichen absoluten Outcome-Risiken, aber nicht Gruppen mit unterschiedlichen relativen Risiken für eine bestimmte Behandlungsstrategie (Clark, 2008). Es ist grundsätzlich schwierig, echte Prädiktoren für eine bestimmte Behandlungsstrategie aus einzelnen Forschungsprojekten zu ermitteln, da solche Untersuchungen in der Regel explorativ sind und ein hohes Potenzial für Fehler erster und zweiter Art beinhalten. Robuste Studiendesigns zur Identifizierung von Faktoren, die tatsächlich einen Behandlungserfolg vorhersagen, sollten idealerweise die vier folgenden Patientengruppen einschließen:

- 1) Interventionskohorte positiv für potenziellen Prädiktor
- 2) Interventionskohorte negativ für potenziellen Prädiktor
- 3) Kontrollkohorte positiv für potenziellen Prädiktor
- 4) Kontrollkohorte negativ für potenziellen Prädiktor

Somit kann der Unterschied im Behandlungseffekt zwischen Patienten, die Prädiktor positiv sind, und denen, die Prädiktor negativ sind, gemessen werden. In Beobachtungsstudien, in denen nur behandelte Patientenpopulationen analysiert

werden und daher kein Vergleich mit Kontrollgruppen möglich ist, kann ein echter Behandlungseffekt nur sehr schwer geschätzt werden. Bei dieser Art von Studiendesign muss man sicher ausschließen können, dass ein potenzieller Prädiktor für das Therapieansprechen, nicht auch einen prognostischen Einfluss auf den Krankheitsverlauf unabhängig von der Behandlung hat, um zu der Schlussfolgerung gelangen zu können, dass es sich um einen echten prädiktiven Marker der Behandlung handelt.

1.2.1.7 TRIPOD

Das Akronym „TRIPOD“ steht für „Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis“. Die TRIPOD-Erklärung umfasst eine 22-Punkte umfassende Checkliste, die die Berichterstattung über Studien verbessern soll, in denen Prognosemodelle entwickelt, validiert oder aktualisiert wurden. Die TRIPOD-Erklärung zielt unter anderem darauf ab, die Transparenz der Berichterstattung über eine Prognosemodellstudie unabhängig von den verwendeten Studienmethoden zu verbessern (Moons et al., 2015)

1.2.2 Entwicklung von Prognosemodellen

Die meisten Untersuchungen im Rahmen von Prognoseforschung beinhalten statistische Modellierungen, um Zusammenhänge von Endpunkt und Ausgangspunkt auszudrücken. Eine der häufigsten statistischen Methoden der klassischen Modellentwicklung ist die Regressionsanalyse (Wallisch et al., 2022): Eine Ergebnisvariable wird durch eine oder mehrere unabhängige Variablen ausgedrückt. Neben einfachen Regressionsalgorithmen werden in den letzten Jahren häufig komplexere Algorithmen wie maschinelles Lernen (ML) als vielversprechende Methodik zur Verbesserung von Prognosemodellen genutzt (Riley et al., 2019c). Um das Potenzial von komplexen Algorithmen mit flexiblen Modellentwicklungsmethoden jedoch vollständig ausschöpfen zu können, ist es notwendig, große Datensätze zu nutzen, die über Hunderte bis Tausende von Patienten umfassen. Nur dann haben die entwickelten Modelle das Potenzial, die interindividuelle Variabilität in Bezug auf Präsentation der Krankheit, Komorbiditäten und Prädisposition adäquat abzubilden (Arbabshirani et al., 2017, Collins, 2012). Durch das Zusammenspiel von wachsenden Datenmengen und einer exponentiell steigenden Rechenleistung von Computern,

erschwinglichen Rechen- und Speicherressourcen sowie einem breiten Softwareangebot konnten viele Unzulänglichkeiten behoben werden und zu bahnbrechenden Innovationen führen, wie etwa einer intelligenten Software zum Verstehen von Sprache und Bild (Hinton et al., 2012, Krizhevsky et al., 2012).

1.2.2.1 Klassische Prognosemodelle

Klassische Algorithmen wie einfache univariate und komplexere multivariable Regressionsanalysen, werden häufig für die Entwicklung von Prognosemodellen genutzt (Kent et al., 2020). Die Entwicklung der logistischen Regressionsanalyse kann in ihren Ursprüngen bis in das neunzehnte Jahrhundert zurückverfolgt werden und ist vor allem in den letzten zwei Jahrzehnten zu einem sehr häufig verwendeten statistischen Instrument in der medizinischen Forschung geworden (Cramer, 2002, Oommen et al., 2011).

Die Bezeichnung „klassisches multivariables Regressionsmodell“ kann jedoch irreführend sein. Als rein statistisches Werkzeug, wird die logistische Regressionsanalyse grundlegend auf binäre Klassifizierungsprobleme angewendet. Die Methode wird aber auch als Modellentwicklungstechnik in komplexe ML-Algorithmen integriert (Makalic and Schmidt, 2011). Die Unterscheidung zwischen „klassischem Prognosemodell“ und „ML-basiertem Prognosemodell“ erfolgt daher nicht *a priori* durch die statistische Methode, mit der das Prognosemodell entwickelt wurde, sondern wird durch einen übergeordneten Algorithmus definiert (siehe Abbildung 1).

1.2.2.2 ML-basierte Prognosemodelle

Techniken des maschinellen Lernens (ML) werden zunehmend in einem medizinischen Kontext angewandt, insbesondere zur Prognosebestimmung (Obermeyer, 2016). Maschinelles Lernen ist ein Bereich der Computerwissenschaften, der nach Arthur Samuel (1959) definiert ist als "[...] *die Fähigkeit von Computern, zu lernen, ohne explizit programmiert zu werden* [...]". Konventionelle ML-Algorithmen entwickeln Modelle anhand eines Trainingsdatensatzes. Das trainierte Modell wird

anhand eines weiteren Testdatensatzes evaluiert, in dem es auf diesen zur Klassifikation beziehungsweise Analyse angewendet wird (Abbildung 1).

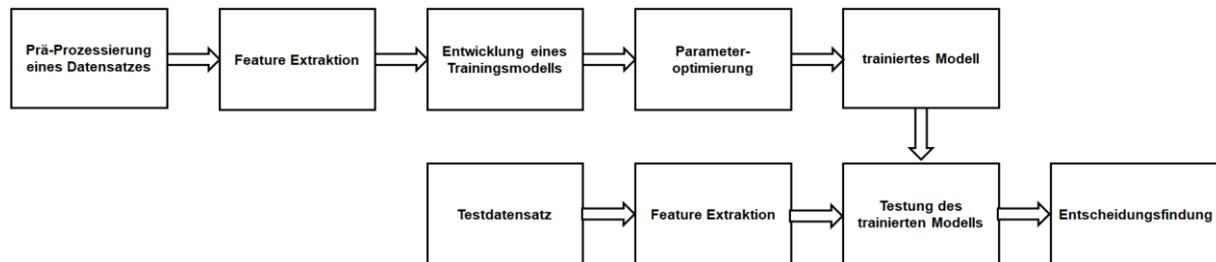


Abbildung 1: Allgemeines Entwicklungsdiagramm eines konventionellen ML-Algorithmus

Konventionelle ML-Algorithmen können in „Supervised Learning“ und „Unsupervised Learning“ unterschieden werden. „Supervised Learning“ beschreibt eine Gruppe von Algorithmen zum Erlernen der Beziehung zwischen unabhängigen Attributen und einem bestimmten abhängigen Attribut (dem Label). Dazu zählen z.B. Induktionsalgorithmen (Algorithmen, die eine bestimmte Anzahl an Fällen nutzen, um Modelle zu erzeugen, die über diese Instanzen hinaus verallgemeinerte Entscheidungen zulassen) (Kohavi, 1998). Der Begriff „Unsupervised Learning“ fasst Algorithmen zusammen, die Fälle ohne ein vorher festgelegtes abhängiges Attribut einordnen. Clustering-Algorithmen z. B. sind in der Regel unüberwacht (Kohavi, 1998).

Zu den konventionellen „Supervised Learning“-Techniken gehören Methoden wie „Support Vector Machines (Burgess, 1998), „Neural Networks“(Jain et al., 1996), „Random Forest“ (Afanador et al., 2016) und hybride Algorithmen, wie „Gradient Boosting Machine“(Natekin and Knoll, 2013), aber auch die multivariable logistische Regression (Rich Caruana et al., 2006).

1.2.2.3 Validierung

Bei der datenabhängigen Modellentwicklung ist es wichtig, den möglichen Umstand in Betracht zu ziehen, dass das Modell das Ergebnis des Zufalls sein kann (Sauerbrei et al., 2007). Daher sollten alle entwickelten Modelle intern und extern validiert werden. Validierung (auch als Transportabilität bezeichnet) beschreibt den Prozess der Beurteilung der Leistung (z.B. Diskriminationsfähigkeit, Kalibrierung, Brier-Score)

eines prognostischen Modells in (weiteren) Stichproben, die nicht primär zur Modellentwicklung verwendet worden sind. Ein häufiger Ansatz der internen Validierung ist es, einen zufällig gewählten Anteil des Ursprungsdatensatzes exklusiv für die Modelltestung zu nutzen. Grundsätzlich ist es jedoch statistisch effizienter, den kompletten Datensatz zur Modellentwicklung zu nutzen und anschließend mittels „Bootstrapping“ oder „Cross-validation“ die interne Validierung durchzuführen. Bootstrapping ist eine „Resampling-Methode“, bei der aus einer Stichprobe erneut viele Stichproben gezogen werden, um einen Testdatensatz von gleicher Größe wie im Ursprungsdatensatz zu erhalten (Efron, 1979). Von größerem Interesse als die interne Validität ist die externe Validität oder Verallgemeinerbarkeit (Altman and Royston, 2000). Externe Validität wird typischerweise in unabhängigen Validierungsstichproben mit Patienten aus einer unabhängigen, aber "plausibel verwandten" Population untersucht (Justice et al., 1999).

1.2.2.4 „Sparse Data Bias“

Auswirkungen einer Behandlung oder einer anderen Exposition auf bestimmte Endpunkte werden in der Regel durch relatives Risiko, Risikoraten oder Odds gemessen. Adjustierte Schätzungen dieser Größen werden in der Regel durch Maximum-Likelihood-Regression (z.B., logistische, Poisson- oder Cox-Modellierung) geschätzt. Effektschätzer können in diesem Zusammenhang ernsthafte Verzerrungen aufweisen, wenn die Daten keine adäquaten Fallzahlen für eine Kombination von Exposition und Ereignis aufweisen. Diese Verzerrung kann sogar bei recht großen Datensätzen auftreten und wird daher oft als "Verzerrung aufgrund spärlicher Daten" bezeichnet (engl. "sparse data bias"). Diese Verzerrung kann entweder durch adjustierte Regressionsmodellentwicklung für potenzielle Confounder entstehen oder durch sie verstärkt werden. Im Extremfall können die entwickelten Modelle artefaktbedingt sehr große oder unendliche große Werte der Schätzer generieren. Die folgenden Merkmale, die miteinander wechselwirken, tragen zur Verzerrung bei Maximum-Likelihood-Regressionsanalysen von Krankheitsereignissen in spärlichen Datensätzen bei:

1. Geringe Fallzahl im Verhältnis zur Anzahl der Prädiktoren (Fälle pro Variable = „EPV“), im speziellen gemessen als Anzahl der Ausfälle pro Variable für die Cox-Proportional-Hazards Hazard- und Poisson-Regression, und das Minimum

der Anzahl von Fällen und Nicht-Fällen pro Variable bei der logistischen Regression. Bei der bedingten logistischen Regression sollten nur die Zahlen innerhalb diskordanter übereinstimmender Gruppen gezählt werden.

2. Variablen mit engen Verteilungen oder mit Kategorien, die sehr unüblich sind.
3. Variablen, die zusammen das Ergebnis nahezu perfekt vorhersagen (z.B., wenn eine Kombination von diskreten Kovariaten nur unter den Studienteilnehmern mit einem bestimmten Ereignis auftritt). Dies wird auch als Separationsproblem bezeichnet.
4. Multikollinearität: Die unabhängigen Variablen in einem Modell korrelieren sehr stark miteinander.

1.2.2.5 „Overfitting und Underfitting“

Ein häufiges Problem bei der Entwicklung von Prognosemodellen ist, dass die Genauigkeit der Vorhersagen innerhalb externer Populationen abnimmt, d.h. die Modelle lassen sich nicht uneingeschränkt verallgemeinern (Charlson et al., 1987). Bei der Entwicklung eines Vorhersagealgorithmus ist der erste Schritt die Kontrolle der statistischen Überanpassung (engl. „overfitting“) oder Unteranpassung (engl. „underfitting“), um diesem Problem entgegenzuwirken. Ursache für das Problem ist, dass bei der Entwicklung von Regressionsmodellen häufig konventionelle Verfahren verwendet werden, insbesondere „ordinary least squares regression“ oder „standard maximum likelihood estimation“. Diese neigen jedoch dazu, Modellgleichungen zu generieren, die zu stark an den Entwicklungsdatensatz angepasst sind und zu extreme Vorhersagen liefern, wenn sie auf eine neue Population angewendet werden. D.h. einige Prädiktionen liegen zu weit vom Gesamtmittelwert entfernt vor. Zum Beispiel kann die Verwendung der „Standard maximum likelihood estimation“ bei der Entwicklung eines logistischen Regressionsmodells zu Wahrscheinlichkeiten führen, die zu nahe an dem Wert 0 für Personen mit niedrigem Risiko und zu nahe an dem Wert 1 für Personen mit hohem Risiko liegen (Riley et al., 2021). „Overfitting“ neigt dazu, die Wahrscheinlichkeit eines Ereignisses bei Patienten mit niedrigem Risiko zu unterschätzen und die Wahrscheinlichkeit eines Ereignisses bei Hochrisikopatienten zu überschätzen (Pavlou et al., 2015). Das Problem der Überanpassung wird zudem

größer, wenn der Stichprobenumfang der Modellentwicklungspopulation abnimmt, die Anzahl der in Frage kommenden Prädiktoren zunimmt und (bei binären oder zeitabhängigen Modellen) die Anzahl der Ereignisse in der untersuchten Population abnimmt (Riley et al., 2021). Bei Underfitting, welches zum Beispiel bei sehr kleinem EPV ($EPV < 4$) auftreten kann, kann es zu Problemen der Regularisierungen bei der Modellentwicklung kommen (Pavlou et al., 2016).

1.2.2.6 Penalisierung/Regularisierung

Wenn das EPV gering ist, kann es mittels klassischer Regressionsanalyse zu Overfitting, aber auch Underfitting kommen, was im Folgenden bei externer Validierung zu ungenauen Vorhersagen führt (Riley et al., 2021). Bei der Entwicklung von Prognosemodellen im Rahmen von Forschungsprojekten im Bereich seltener Krankheiten, seltener Entitäten eines Krankheitsbildes (z.B. posteriorer Schlaganfall) oder bei Forschungsprojekten mit sehr speziellen Fragestellungen kann das Problem auftreten, dass durch kleine Patientenkohorten mit wenigen Ereignissen ein sehr geringes EPV *a priori* vorliegt (z.B. $EPV < 10$). Daher ist es sinnvoll, die Anzahl der Prädiktoren vor Modellentwicklung oder während des Entwicklungsprozesses zu reduzieren. Simple Strategien zur Reduktion von Kandidatenprädiktoren sind univariable Screeningmethoden oder iterative Variableneliminationsverfahren. Beim univariablen Screening wird die Beziehung zwischen jedem Prädiktor mit dem Ergebnis einzeln untersucht und nur statistisch signifikante Zusammenhänge werden in das multivariable Risikomodelle aufgenommen. Der Term „iterative Variableneliminationsverfahren“ beschreibt die schrittweise Exklusion (z.B. Vorwärts- oder Rückwärtselimination) der Prädiktoren, die bei einem vorgegebenen p-Wert nicht statistisch signifikant im Modell vorliegen. Dadurch ist eine Reduktion der Komplexität des Modells und eine Senkung des EPV-Wertes möglich. Beide Ansätze haben jedoch gravierende Nachteile, wenn der Prädiktorenauswahlprozess nicht „stabil“ ist, d.h. dass kleine Änderungen in den Daten oder in der Methodik der Variablenauswahl dazu führen, dass nur bestimmte Prädiktoren in das endgültige Modell aufgenommen werden und andere nicht (Steyerberg et al., 2001a, Ye, 1998, Zou and Hastie, 2005b). Im Speziellen können sie bei der Entwicklung von logistischen Regressionsmodellen bei Vorliegen von „Sparse Data Bias“ zu inflationären Effektschätzungen, zu engen

Konfidenzintervallen der Regressionskoeffizienten und zum "Drop Out" von wichtigen Confoundern führen (Greenland et al., 2016).

Modernere Variablenselektionsmethoden vermindern dagegen zusätzlich die Varianz der Vorhersagen des entwickelten Modells (Regularisierungseigenschaften), wodurch der mittlere quadratische Fehler der Vorhersagen reduziert wird. Zum Beispiel werden in einem logistischen Regressionsmodell die Prädiktoreffekte (Odds Ratios) gegen 1 kalkuliert, sodass die vorhergesagten Wahrscheinlichkeiten in einer neuen (externen) Population weniger Variabilität aufweisen, d.h. in Richtung der mittleren Ergebniswahrscheinlichkeit im Datensatz laufen. Penalisierungstechniken werden daher grundsätzlich als wesentliche Werkzeuge für die Entwicklung von Vorhersagemodellen empfohlen (Ambler et al., 2012, Copas, 1983, Copas, 1997, Pavlou et al., 2015, Riley et al., 2021, Steyerberg et al., 2001b, Tibshirani, 1996, Van Houwelingen, 2001). Unter anderem werden LASSO-Regression (engl. „least Absolute Shrinkage and Selection Operator“) oder Elastic Net in diesem Zusammenhang verwendet (Tibshirani, 1996, Zou and Hastie, 2005a). Neben hybriden Selektions- und Regularisierungsverfahren existieren auch variablenselektionsunabhängige Regularisierungsmethoden wie klassisches, uniformes „Shrinkage“ oder Ridge Regression (Hoerl and Kennard, 1970, Hudda et al., 2019).

1.2.2.7 Ridge-Regression

In der medizinischen Forschung wird die klassische logistische Regressionsanalyse üblicherweise verwendet, um die Beziehung zwischen einem binären Ergebnis und einer Reihe von Kovariaten zu untersuchen. Klassisch bedeutet in diesem Zusammenhang die Schätzung der Regressionskoeffizienten mittels Maximum-Likelihood-Methodik. Bei einem Datensatz mit ähnlichen Prävalenzen für das binäre Ereignis und ausreichender Stichprobengröße wird dadurch die Inferenz erleichtert, d. h. die Interpretierbarkeit der Effektschätzer sowie die Genauigkeit der Vorhersagen unter Berücksichtigung der Kovariaten. Diese Eigenschaften verschwinden jedoch, wenn die Stichprobengröße klein oder die Prävalenz von Ergebnismöglichkeiten (für eine bestimmte Kombination von Expositionen beziehungsweise Kovariaten) niedrig ist, was zu verzerrten Koeffizientenschätzungen führen kann und konsekutiv nur sehr instabile Vorhersagen zulässt (Greenland et al., 2016, Pavlou et al., 2016). Bei der Modellentwicklung mittels Ridge Regression wird ein Regularisierungsbegriff (engl.

„penalty term“) der Log-Likelihood-Funktion hinzugefügt. Dies sorgt für eine Schrumpfung der Koeffizienten gegen 0, wodurch die Varianz der Schätzungen der Regressionskoeffizienten vermindert wird und dadurch die Vorhersage stabilisiert wird (Le Cessie and Van Houwelingen, 1992). Ridge-Regression ist von Vorteil bei Vorliegen von Separationsproblemen oder bei Vorliegen von starker Multikollinearität (Mansournia et al., 2018, Duzan and Mohamad Shariff, 2015). Aber auch unabhängig von Multikollinearität und vor allem generell im Zusammenhang mit „Sparse Data Bias“ kann Ridge-Regression eine sinnvolle Modellierungsstrategie darstellen (Wang et al., 2020, Greenland et al., 2016).

1.2.2.8 Firth-Regression und Firth-Modellmodifikationen

Die ursprüngliche Regressionsanalyse mittels klassischer Firth-Penalisation wird in jüngerer Zeit ebenfalls primär dafür genutzt, um Separationsproblemen zu begegnen, welche in besonderem Maße bei geringer Fallzahl auftreten können (Heinze and Schemper, 2002). Technisch wird bei der Penalisation nach Firth unter Verwendung des invarianten Jeffrey-Prior der Verzerrung erster Ordnung in den Regressionskoeffizienten entgegengewirkt (Firth, 1993). Bei „Sparse Data Bias“-Problemen stellen modifizierte Methoden nach Firth eine Alternative zur Ridge-Regression oder anderen Penalisierungsmethoden dar, wenn keine Variablenselektion gewünscht ist und unabhängig davon, ob ein Separationsproblem vorliegt oder nicht. Bei „FLIC“ (engl. „Firth’s logistic regression with intercept-correction“) wird das Intercept von der Penalisation ausgenommen, dadurch ergeben sich keine Verzerrungen durch den Jeffrey-Prior auf die Genauigkeit der Schätzungen des Regressionmodells. Bei „FLAC“ (engl. „Firth’s logistic regression with added covariate“) wird durch Augmentation beziehungsweise durch Ergänzung von Pseudodata der Verzerrung auf die geschätzten Wahrscheinlichkeiten durch den Jeffrey-Prior entgegengewirkt. FLIC und FLAC sind der klassischen Penalisation mittels Firth-Korrektur in Bezug auf Effektschätzungen und prognostische Genauigkeit überlegen (Puhr et al., 2017).

1.2.2.9 Modelloptimierung und Modellaktualisierung

Viele verschiedene Quellen können die Kalibrierung von Prognosemodellen verzerren. Es können methodenassoziierte Ursachen von nichtmethodenassoziierten Ursachen unterschieden werden. Overfitting oder Underfitting ist ein methodisches Problem (Steyerberg et al., 2018). Ein Algorithmus, der in einem Datensatz mit hoher Eventrate entwickelt wurde, kann aber auch systematisch zu überhöhten Risikoschätzungen führen, wenn er in einem Kollektiv niedriger Inzidenz/Prävalenz (niedrige Eventrate) angewendet wird (nicht-methodische Ursache). Darüber hinaus können Patientenpopulationen sich mit der Zeit aufgrund von Veränderungen in der generellen Patientenversorgung ändern und konsekutiv veränderte Inzidenz- und Prävalenzraten bedingen (Davis et al., 2017). Aus diesem Grund müssen Prognosemodelle über die Zeit regelmäßig aktualisiert werden („Modellaktualisierung“).

1.2.3 Multiple Imputation (MI)

Unvollständige Datensätze sind in der medizinischen Forschung häufig, was zu einem Verlust an statistischer Aussagekraft und potenziell verzerrten Ergebnissen führen kann, wenn sie nicht angemessen behandelt werden. Die Multiple Imputation (MI) ist eine in der Praxis weitverbreitetes, statistisches Verfahren zum Umgang mit fehlenden Daten (engl. „missing data“) (Hayati Rezvan et al., 2015).

1.2.3.1 Grundlagen der multiplen Imputation

Unvollständige Datensätze entstehen, wenn einige Variablen zum Beispiel nur bei einer Teilmenge der Probanden gemessen werden oder wenn Datensätze, die zu unterschiedlichen Zeitpunkten und/oder an unterschiedlichen Orten erhoben worden sind, zusammengefügt und anschließend gemeinsam statistisch ausgewertet werden sollen.

Es werden verschiedene Typen von Mustern fehlender Datenwerte unterschieden (van Buuren, 2018):

- 1) Univariat oder multivariat: Ein unvollständiger Datensatz ist univariat, wenn es nur eine Variable mit fehlenden Daten gibt.

- 2) Monoton oder nicht-monoton (oder allgemein): Ein unvollständiger Datensatz ist monoton, wenn die Variable Y_j eines fehlenden Datensatzes der Form einer $n \times p$ matrix „Y“, in der Art geordnet werden kann, dass, wenn Y_j fehlt, alle Variablen Y_k mit $k > j$ auch fehlen. Dies ist zum Beispiel bei longitudinalen Studien mit „Drop-out“ der Fall.
- 3) Verbunden oder nicht-verbunden: Ein unvollständiger Datensatz ist verbunden, wenn jeder beobachtete Datenpunkt mit jedem anderen beobachteten Datenpunkt verbunden ist und durch eine Sequenz von horizontalen oder vertikalen Bewegungen im „md.pattern“-Graph (engl. „missing data pattern graph“) erreicht werden kann (Buuren and Groothuis-Oudshoorn, 2011).

Weiterhin werden verschiedene Mechanismen unterschieden, durch die fehlende Werte eines Datensatzes miteinander in Beziehung stehen: Es werden drei verschiedene, Mechanismen beschrieben, die zu einem fehlenden Datensatz führen:

- 1) „Missing Completely At Random (MCAR)“
- 2) „Missing At Random (MAR)“
- 3) „Missing Not At Random (MNAR)“

MCAR ist definiert als die Wahrscheinlichkeit, dass eine Observation (einer Variablen), die lediglich vom Zufall abhängt, fehlt. MAR bezeichnet die Wahrscheinlichkeit, dass eine Observation (einer Variablen) fehlt, welche lediglich von anderen beobachteten Merkmalen abhängt. MNAR bezeichnet die Wahrscheinlichkeit, dass Daten (einer Variablen) fehlen, die von unbeobachteten und beobachteten Werten abhängig sind.

Rubins Klassifikation für fehlende Daten ist am einfachsten für eine einzelne unvollständige Variable zu beschreiben. Zum Beispiel sind Blutdruckdaten MAR, wenn sie primär in einem Kollektiv von älteren Patienten gemessen wurden (und das Alter als Kandidatenprädiktor in die Analyse miteinbezogen wird). Sie sind MNAR, wenn sich die Untersuchungskohorte zu jeweils hohen Anteilen aus Patienten mit bekanntem und nicht vorbekanntem Blutdruckhochdruck zusammensetzt (und die Information über den bereits bestehenden Bluthochdruck nicht mit in die Analyse einfließt), da Patienten mit bekannter arterieller Hypertonie ihren Blutdruck wahrscheinlich häufiger erfassen lassen als andere Personen desselben Alters.

Rubin et al. entwickelten darüber hinaus ein allgemeines Konzept für den Umgang mit fehlenden Daten (Rubin, 1987). Allgemein beruht es auf der Idee, dass jedes Subjekt

in einer Stichprobe durch ein neues, zufällig ausgewähltes Subjekt aus derselben Ausgangspopulation ersetzt werden kann. Bei der multiplen Imputation werden mehrere Schätzungen zur Ergänzung der fehlenden Werte einer Variablen genutzt, die anhand anderer, jedoch bekannter Charakteristika der Stichprobe ermittelt werden (Donders et al., 2006). Bei der multiplen Imputation werden M ($M > 1$) Kopien der ursprünglichen Stichprobe in m Imputationszyklen generiert und die fehlenden Daten mit Hilfe statistischer Modelle ergänzt. Unter Verwendung der Rubinschen Regeln (engl. "Rubin's rules") wird jeder fehlende Wert durch eine Reihe plausibler Werte ersetzt. Der Unsicherheit über den wahren Wert wird durch Berücksichtigung der Variabilität innerhalb des originalen Datensatzes und Berücksichtigung der Variabilität zwischen den Imputationen Rechnung getragen (Rubin, 1987).

Bei einer anschließenden deskriptiven und interferenzstatistischen Analyse können wichtige Schätzer (z.B. deskriptive Schätzer wie Mittelwert oder Median, Regressionskoeffizienten oder Evaluationsparameter von Prognosemodellen) der M imputierten Datensätze und ihre Standardfehler unter Verwendung der Rubinschen Regeln berechnet werden.

Warum ist es grundsätzlich sinnvoll, unvollständige Datensätze in die statistische Auswertung mit einzuschließen? Es wurde gezeigt, dass eine effiziente Schätzung der Ergebnisparameter komplizierter wird, wenn Basiswerte teilweise fehlen (White and Thompson, 2005, Jakobsen et al., 2017a). Praktisch wird durch das Löschen von Fällen im Rahmen einer „Complete Case Analysis“ die Variablenauswahl erschwert, wenn verschiedene Variablenwerte bei verschiedenen Fällen fehlen, außerdem wird die „statistische Power“ reduziert und ein potenzieller Bias in den resultierenden Ergebnissen erzeugt (Arnold and Kronmal, 2003). Alternativ formuliert ergibt sich eine mögliche Verzerrung durch die Entwicklung eines prognostischen Modells mit nur vollständigen Daten aus den folgenden zwei Punkten: 1.) die geschätzten Standardfehler für die Regressionskoeffizienten werden unnötig groß, da Informationen durch den Ausschluss von Probanden mit fehlenden Werten verloren gehen. 2.) Eine Unterscheidung zwischen MCAR und MAR mit einer ausschließlich kompletten Fallanalyse ist nicht möglich und die geschätzten Regressionskoeffizienten könnten somit verzerrt sein.

Allgemein umfasst die Methode der Multiplen Imputation und anschließenden Analyse der komplettierten Daten drei Schritte:

- 1) Imputationsmodellentwicklung und mehrzeitige Imputation der Daten unter einem geeigneten Modell zur Erzeugung von M Kopien eines vollständigen Datensatzes
- 2) Analyse jedes Datensatzes separat, um die gewünschten Parameterschätzungen und Standardfehler zu erhalten
- 3) Kombination der Ergebnisse der m Analysen unter Verwendung der Rubinschen Regeln. Rubin, et al. und Schafer, et al. haben gezeigt, dass in der Regel 3-5 Imputationen ausreichen (Rubin, 2004, Schafer, 1999).

Bei der Konstruktion von Imputationsmodellen müssen die folgenden Punkte beantwortet werden:

- 1.) Ist eine Imputation fehlender Werte im vorliegenden Fall sinnvoll?
- 2.) Wahl des allgemeinen Imputationalgorithmus (Lee and Carlin, 2010)
- 3.) Wahl spezieller Imputationsmethoden für kategoriale/binäre Variablen und nicht-normalverteilte, kontinuierliche Variablen (Lee and Carlin, 2017, Lee et al., 2012, Yucel et al., 2011) in Abhängigkeit vom Imputationsalgorithmus.
- 4.) Auswahlmethode der in das Modell aufzunehmenden Variablen (Collins et al., 2001);
- 5.) ggf. Durchführung einer Post-Hoc-Imputationsdiagnostik
- 6.) spezielle Pooling-Methoden bei deskriptiv- und interferenzstatistischen Auswertungen.

1.2.3.1.1 Ist eine Imputation fehlender Werte sinnvoll?

In einer Vielzahl von Forschungsbereichen stellen sich Statistiker häufig die Frage ab welchem Anteil an fehlenden Werten die Anwendung von MI noch gerechtfertigt ist. In der Literatur wird ein Anteil von 5% an fehlenden Daten als untere Grenze vorgeschlagen, unterhalb derer MI einen vernachlässigbaren Nutzen bietet (Schafer, 1999). Obere „Cut-off's“ sind nicht klar definiert. In manchen Publikationen wird darauf hingewiesen, dass bei Analysen mit mehr als 10% fehlenden Daten eine Verzerrung der Ergebnisse wahrscheinlich ist und dass bei mehr als 40% an fehlenden Daten

wichtiger Schlüsselvariablen Ergebnisse nur als hypothesengenerierend betrachtet werden sollten (Dong and Peng, 2013, Jakobsen et al., 2017b). Arbeiten von Kleinke et al. haben jedoch gezeigt, dass bei höheren Anteilen an fehlenden Daten dennoch plausible Werte generiert werden können (Kleinke, 2018), sofern bestimmte Kriterien eingehalten werden:

- 1.) absolute Schiefe der Variablenverteilung (engl. skewness) $< 1,6$
- 2.) prozentualer Anteil an fehlenden Werten $< 50\%$ (Kleinke, 2017).

1.2.3.1.2 Wahl des Imputationsalgorithmus

Statistisch stehen verschiedene Möglichkeiten zur Entwicklung von Imputationsmodellen zur Verfügung. Zur Imputation multipler Variablen werden allgemein zwei Verfahren primär genutzt:

- 1.) „Joint Modelling“ -Techniken, wie „MultiVariate Normal Imputation (MVNI)“ (Ragunathan et al., 2001, van Buuren, 2007)
- 2.) „Multivariate Imputation by Chained Equations (MICE)“ (Schafer, 1997, He, 2010, Buuren and Groothuis-Oudshoorn, 2011).

Die Imputationen der fehlenden Werte mittels MICE werden durch Entwicklung einer Reihe von Regressionsmodellen basierend auf den vorhandenen Werten generiert. Für die Regressionen werden lineare, logistische, Poisson, verallgemeinerte Logit oder eine Mischung dieser Modelle genutzt, je nach Art der zu imputierenden Variablen (Ragunathan et al., 2001, van Buuren, 2007).

Bei MICE wird das Imputationsmodell für jede einzelne Variable definiert. Dies bedeutet, dass MICE das Problem, ein gemeinsames Verteilungsmuster für alle Variablen gleichzeitig zu finden, auf die Suche nach jeweils einer separat bedingten Verteilung für jede unvollständige Variable reduziert, was die MICE-Methodik zu einem sehr flexiblen Ansatz macht (van Buuren, 2007).

1.2.3.1.3 Spezielle Imputationsmodelle

Bei nicht linearen Beziehungen und Nicht-Normalverteilung kann für die multiple Imputation unter anderem das „Predictive mean matching“ (PMM) genutzt werden

(White et al., 2011). Bei PMM handelt es sich um eine semiparametrische Hot-Deck-Imputationsmethode (Schenker and Taylor, 1996). PMM ist weniger empfindlich gegenüber Modellfehlspezifizierungen wie nicht-linearen Assoziationen, heteroskedastischen Residuen und Nicht-Normalverteilungen (Schenker and Taylor, 1996, Morris et al., 2014). PMM ist insbesondere ein sehr robustes Verfahren zur Imputation bei hohem Anteil fehlender Werte (Morris et al., 2014), auch bei fehlenden Werten von bis zu 50% (Vink et al., 2014).

1.2.3.1.4 Auswahl der in das MI-Modell aufzunehmenden Variablen

Als allgemeine Regel gilt, dass die Nutzung aller verfügbaren Informationen Imputationen mit minimaler Verzerrung und maximaler Effizienz begünstigt (Collins et al., 2001, Xiao-Li, 1994). Daher sollte das Imputationsmodell, das zur Generierung plausibler Werte entwickelt wird, alle Variablen enthalten, die später auch analysiert werden sollen, ggf. einschließlich der abhängigen Variable, sowie jedwede Interaktion zwischen den Variablen und Variablen, die das prognostische Modell unterstützen (van Buuren et al., 1999, Tilling et al., 2016, White et al., 2011). Bei Letzteren handelt es sich um sogenannte Hilfsvariablen (engl. „auxillary variables“). Diese Variablen sind in der Regel nicht im späteren Prognosemodell enthalten, ihre Implementierung im Imputationsmodell bietet aber zwei wesentliche Vorteile:

- 1.) Auxillary Variables können hilfreich sein, um die MAR-Annahme, die als Bedingung bei der Implementierung eines MI-Modells mindestens erfüllt sein muss, plausibler zu machen
- 2.) „Auxillary Variables erhöhen die Wahrscheinlichkeit, unverzerrte Schätzungen zu generieren beziehungsweise, dass wichtige Informationen über fehlende Werte nicht verloren gehen (White et al., 2011).

Letzterer Effekt ist besonders stark bei Datensätzen mit einem hohen Anteil an fehlenden Informationen ausgeprägt (Madley-Dowd et al., 2019).

1.2.3.1.5 Posthoc-Imputationsdiagnostik

Der imputierte Datensatz sollte möglichst genau mit den vorhandenen Daten konvergieren. Zur Prüfung können „Trace-Line-Plots“ verwendet werden (Buuren and

Groothuis-Oudshoorn, 2011). In „Trace-Line-Plots“ sind Mittelwert und Standardabweichung der imputierten Werte gegen die Zahl der Imputationszyklen aufgetragen. Im Allgemeinen sollten sich die Trendlinien gut vermischen und bei den späten Iterationen frei von jeglichen Trends sein (van Buuren, 2018).

Im Allgemeinen würde man es vorziehen, wenn es sich bei den imputierten Daten um plausible Werte handelt, d.h. um Werte, die beobachtet worden sein könnten. Um sich ein Bild von der Plausibilität zu machen, kann man die Imputationen überprüfen und sie mit den beobachteten Werten vergleichen. Der Stripplot ist nützlich, um die Verteilungen in Datensätzen mit einer geringen Anzahl von Datenpunkten zu untersuchen (van Buuren, 2018). „Stripplot()“ des R-Package „MICE“ erzeugt eindimensionale Streudiagramme. Die Funktion trennt automatisch die beobachteten und imputierten Datenwerte metrischer Variablen farblich auf (van Buuren, 2018).

1.2.3.1.6 Spezielle Pooling Methoden

Die Anwendung der Rubinschen Regeln zur Bestimmung der Schätzer ist für kontinuierliche und dichotome Variablen leicht anwendbar, da klassischerweise nur eine einzige Wald-Statistik verwendet wird (Marshall et al., 2009). Bei der Kombination von multipler Imputation und Rückwärtsselektion zur Prädiktoridentifikation werden jedoch spezielle Verfahren, wie die MPR-Methode bevorzugt (Panken and Heymans, 2022).

1.2.4 Evaluation von Prognosemodellen

Die Evaluation von Prognosemodellen kann mit einer Vielzahl unterschiedlicher Methoden und Messgrößen bewertet werden. Zu den traditionellen Maßstäben für binäre Modelle gehören die Konkordanz- (oder c-) Statistik für die Unterscheidungsfähigkeit (Diskrimination), die Kalibrierungsstatistiken zur Überprüfung der Übereinstimmung vorhergesagter und beobachteter Endpunkte und der Brier-Score beziehungsweise standardisierte Brier-Score zur Angabe der Gesamtleistung des Modells (Steyerberg et al., 2010).

1.2.4.1.1 Diskrimination

Diskrimination analysiert die Fähigkeit eines Prognosemodells, zwischen den Patienten zu unterscheiden, die ein spezielles Ereignis während des Studienzeitraumes erfahren oder nicht. Diese Diskriminationsfähigkeit wird mittels „area under the receiver operating characteristic curve statistic“ (=c-Statistik) quantifiziert. Ein Wert von 0.5 bedeutet, dass die Einordnung durch das Prognosemodell zufällig geschieht. Ein Wert von 1 bedeutet, dass das Modell perfekt unterscheiden kann (Harrell et al., 1996).

1.2.4.1.2 Kalibrierung

Die Kalibrierung bezieht sich auf die Übereinstimmung zwischen beobachteten Ergebnissen und Vorhersagen (Hilden et al., 1978).

Es können verschiedene Hierarchien von Kalibrierung unterschieden werden (Van Calster et al., 2016a):

- 1) Durchschnittliche Kalibrierung („Mean Calibration“)
- 2) Schwache Kalibrierung („Weak Calibration“)
- 3) Moderate Kalibrierung („Moderate Calibration“)
- 4) Starke Kalibrierung („Strong Calibration“)

Bei der Prüfung auf durchschnittliche Kalibrierung wird bewertet, ob die beobachtete Ereignisrate gleich dem durchschnittlichen vorhergesagten Ereignis entspricht. Bei der Prüfung auf schwache Kalibrierung werden das Intercept der Kalibrierungsfunktion und die Steigung der Kalibrierungsfunktion analysiert (=Slope). Diese Parameter geben einen Hinweis darauf, ob eine Über- oder Unteranpassung und eine systematische Über- oder Unterschätzung der vorhergesagten Risiken vorliegen (Van Calster et al., 2019). Bei relativ kleinen Validierungskohorten sollte eine einfache Kalibrierungsbewertung gegenüber hierarchisch höheren Alternativen bevorzugt werden (Collins et al., 2016). Moderate Kalibrierung bezieht sich auf die klassische Definition von Kalibrierung: Ein Modell ist moderat kalibriert, wenn unter Patienten mit dem gleichen vorhergesagten Risiko die beobachtete Ereignisrate gleich dem vorhergesagten Risiko ist (Pepe and Janes, 2013, Steyerberg et al., 2010, Vickers and Cronin, 2010). Moderate Kalibrierung kann mittels flexibler Kalibrierungskurven oder

Kategorisierungen des vorhergesagten Risikos untersucht werden, vorzugsweise mit Angabe von Konfidenzintervallen (Austin and Steyerberg, 2014). Die strengste Definition der Kalibrierung verlangt, dass die vorhergesagten Risiken den beobachteten Ereignisraten für jedes einzelne Kovariatenmuster entsprechen (Pepe and Janes, 2013, Vach, 2013). Diese Definition einer starken Kalibrierung entkoppelt Kovariatenmuster voneinander, die mit demselben vorhergesagten Risiko verbunden sind. Die Forderung nach starker Kalibrierung ist aus klinischer Sicht sinnvoll: Wenn ein Modell moderat aber nicht stark kalibriert ist, kann es zu verzerrten Risikovorhersagen für einzelne Patienten kommen (Vach, 2013). Grundsätzlich hat ein schlecht kalibriertes Modell mit niedriger „area under the curve“ (AUC) einen geringeren klinischen Nutzen als ein Modell mit niedriger AUC aber guter Kalibrierung (Van Calster and Vickers, 2015). Ursachen, die zu einer schlechten Kalibrierung führen, können abhängig oder unabhängig von dem verwendeten Algorithmus der Modellentwicklung sein. Unabhängige Ursachen sind unter anderem stark variierende Inzidenz- oder Prävalenzraten der Modellentwicklungsdatensätze, die nicht repräsentativ für die Grundgesamtheit des Kollektivs sind. Darüber hinaus können sich Patientenpopulationen mit der Zeit aufgrund von generellen Veränderungen in der Patientenversorgung ändern (Davis et al., 2017). Abhängige Ursachen sind Modellentwicklungsstrategien, die zu komplex für den Entwicklungsdatensatz sind (hohes EPV, zu flexible Modellentwicklungsmethoden wie Random Forest) oder Messfehler bei der Datenerhebung (Van Calster et al., 2019).

Trotz der Bedeutung für das Verständnis des Nutzens eines Modells wird über die Kalibrierung wenig berichtet: In einer systematischen Übersichtsarbeit wurde festgestellt, dass zwar 63% der publizierten Prognosemodelle ein Maß für die Diskriminierung enthielten, aber nur in 36% der analysierten Studien über den Faktor Kalibrierung berichtet wurde (Wessler et al., 2015).

1.2.4.1.3 Brier-Score/Allgemeine Leistungsfähigkeit

Der Brier-Score berücksichtigt gleichzeitig Diskrimination und Kalibrierung, weshalb er bei der Evaluation eines Prognosemodells im Vergleich zur einseitigen Bestimmung von Diskriminationsfähigkeit oder Kalibrierung von Vorteil sein kann. Ein niedrigerer

Brier-Score deutet grundsätzlich auf ein besseres Prognosemodell hin. Allerdings kann der klinische Nutzen mittels Brier-Score überschätzt werden, wenn das Prognosemodell miscalibriert ist. Bei niedriger Prävalenz favorisiert der Brier Score Prognosemodelle mit hoher Spezifität, obwohl der klinische Kontext möglicherweise eine hohe Sensitivität fordert (z.B. bei Screeningmaßnahmen). Ein Vorteil des Brier-Scores ist es jedoch, dass keine weiteren externen Informationen benötigt werden, wie es bei der Entscheidungskurvenanalyse der Fall ist (Assel et al., 2017). Eine Erweiterung des Brier-Scores stellt der standardisierte Brier-Score („Brier Scaled“) dar (Steyerberg et al., 2010). Dieser ist prävalenzunabhängig.

1.2.4.1.4 Entscheidungskurvenanalyse („DCA“)

Die Entscheidungskurvenanalyse (engl. „decision curve analysis“) ist eine Methode zur Bewertung und zum Vergleich von Vorhersagemodellen, die klinische Konsequenzen transparent darstellt. Bei der Analyse wird nur der Datensatz benötigt, in dem die Modelle getestet wurden. Mit der DCA können Modelle evaluiert werden, die entweder kontinuierliche oder dichotome Ergebnisse vorhersagen. In der DCA werden Prognosemodelle mit zwei standardisierten Vorhersagestrategien verglichen:

- 1) Alle Patienten werden als „richtig positiv“ betrachtet und werden behandelt oder
- 2) Alle Patienten werden als „richtig negativ“ betrachtet und es wird keine Behandlung durchgeführt.

Der Begriff "Behandlung" wird sehr weit gefasst, d.h. darunter werden nicht nur medikamentöse Therapien, interventionelle Eingriffe oder Operationen verstanden, sondern auch (spezielle) medizinische Beratung, weitere diagnostische Verfahren oder intensivere Überwachung subsumiert (Vickers and Elkin, 2006).

Entscheidungsanalytische Verfahren ermöglichen eine Bewertung des klinischen Nutzens eines Prognosemodells, erfordern aber häufig den Miteinbezug zusätzlicher externer Informationen, wie einen klinisch sinnvollen Schwellenwertbereich, in dem der mögliche relative Schaden durch eine vermiedene Intervention oder die „Nicht-Diagnose“ einer Erkrankung bei einem „richtig positiven“ Patienten gegenüber einem unnötigen Eingriff bei einem gesunden Patienten gerechtfertigt ist (Vickers et al., 2019).

1.2.5 Prognoseforschung Posteriorer Schlaganfall

Seit der bahnbrechenden Autopsieserie von Kubik und Adams im Jahr 1946 ist bekannt, dass akute Arteria (A.) basilaris (BA)-Verschlüsse meist tödliche Hirnstamminfarkte verursachen (Kubik and Adams, 1946). Ein Schlaganfall, bedingt durch einen Verschluss der A. basilaris (engl. „Basilar artery occlusion“; „BAO“), ist einer der verheerendsten Schlaganfall-Subtypen einhergehend mit einer hohen Morbiditäts- und Mortalitätsrate von bis zu 70% (Schonewille et al., 2009).

Allerdings sind nur etwa 1 % aller ischämischen Schlaganfälle durch einen Verschluss der Basilararterie bedingt (Meinel et al., 2019). Schlaganfälle mit genereller Beteiligung der hinteren Zirkulation sind wesentlich häufiger und machen etwa 20-30% aller ischämischen Schlaganfälle aus (Caplan et al., 2004, Bogousslavsky et al., 1988, Moulin et al., 1997). Zu Beginn des 21. Jahrhunderts lag die altersstandardisierte Inzidenz von Schlaganfällen in Europa zwischen 95 und 290/100.000 pro Jahr. Etwa 1,1 Millionen Einwohner Europas erleiden jedes Jahr einen Schlaganfall, wobei etwa 80 % der Fälle auf ischämische Schlaganfälle zurückzuführen sind. Außerdem wird die absolute Zahl der Schlaganfälle aufgrund der alternden Bevölkerung in den kommenden Jahren drastisch ansteigen: Bis 2025 werden 1,5 Millionen Menschen in Europa jedes Jahr an einem Schlaganfall erkranken (Bejot et al., 2016). Davon werden circa 240.000-360.000 Patienten pro Jahr in Europa einen ischämischen Schlaganfall aufgrund eines Verschlusses des postero-vertebrobasilären Systems erleiden. Trotzdem wurden bis vor Kurzem und ungeachtet der Innovationen in der Behandlung des ischämischen Schlaganfalls der letzten Jahre Patienten mit Schlaganfällen der hinteren Zirkulation nicht in die großen randomisiert-kontrollierten Studien miteingeschlossen (Berkhemer et al., 2015, Bracard et al., 2016, Campbell et al., 2015, Goyal et al., 2015, Jovin et al., 2015, Khoury et al., 2017, Mocco et al., 2016, Muir et al., 2017, Saver et al., 2015), was zu Schwierigkeiten führte für diese Entität evidenzbasierte Handlungsempfehlungen zu definieren.

Die Wirksamkeit der endovaskulären Therapie (EVT) bei BAO wurde mittlerweile in mehreren Beobachtungsstudien und auch in einigen kürzlich veröffentlichten kontrolliert, randomisierten Studien (engl. „randomized controlled trial“; RCT) untersucht (Kang et al., 2018, Liu et al., 2020, O'Donnell et al., 2010, Singer et al., 2015, Zi et al., 2020). Einige der RCT's deuten darauf hin, dass Patienten mit akutem

Verschluss der A. basilaris, die mit endovaskulärer Therapie plus medizinischer Standardtherapie ("SMT") behandelt werden, ein besseres funktionelles Überleben zeigen als Patienten, die nur mit SMT behandelt werden (Liu et al., 2020, Zi et al., 2020, Jovin et al., 2022). In diesen Studien konnten jedoch keine festen Kriterien etabliert werden, die Patienten identifizieren, die am ehesten von einer EVT profitieren würden. Eine angemessene Patientenauswahl auf der Grundlage der klinischen und paraklinischen, speziell neuroradiologischen Bildgebung ist jedoch für die erfolgreiche EVT essenziell. Bereits definierte Kriterien müssen folglich weiter verfeinert werden (Sang et al., 2021).

1.2.6 Klinische Endpunkte in der Schlaganfallforschung

Ein häufig verwendeter, klinischer Endpunkt auf dem Gebiet der posterioren Schlaganfallforschung ist der mRS (engl. "modified Rankin Scale") (Jovin et al., 2022, Langezaal et al., 2021, Liu et al., 2020, Xu et al., 2022).

1.2.6.1 Modified Rankin Scale (mRS)

Der mRS bewertet die funktionale Beeinträchtigung und Abhängigkeit von Schlaganfallpatienten basierend auf ihrem Grad der Behinderung. Die Skala reicht von 0 bis 6, wobei 0 auf keine Symptome hinweist und 6 auf den Tod hindeutet. Der mRS wird häufig verwendet, um die Wirksamkeit von Schlaganfallbehandlungen zu messen und Patienten in verschiedene Behinderungsstufen einzuteilen. Der mRS ermöglicht es Forschern, den funktionellen Status von Schlaganfallpatienten zu beurteilen und Veränderungen über die Zeit zu verfolgen. Die Skala umfasst die folgenden Stufen:

- 0) Keine Symptome - Keine Beeinträchtigung.
- 1) Keine signifikante Behinderung - In der Lage, alle üblichen Aktivitäten zu bewältigen, obwohl es leichte Symptome geben kann.
- 2) Leichte Beeinträchtigung - In der Lage, sich selbstständig anzuziehen, zu waschen und auf die Toilette zu gehen, aber einige Aktivitäten können eingeschränkt sein
- 3) Mäßige Beeinträchtigung - Benötigt Hilfe bei einigen täglichen Aktivitäten und kann eingeschränkt mobil sein.

- 4) Mäßig schwere Beeinträchtigung - Abhängigkeit von anderen Personen in Bezug auf die meisten täglichen Aktivitäten.
- 5) Schwere Beeinträchtigung - Benötigt konstante Pflege und Überwachung, noch nicht im Sterbeprozess.
- 6) Tod - Verstorben.

Der mRS ermöglicht es Forschern, den funktionellen Status der Patienten vor und nach einer Behandlung zu bewerten. Eine Verbesserung des mRS-Wertes nach einer Behandlung deutet auf eine Verringerung der Behinderung hin, während eine Zunahme des mRS-Wertes auf eine Verschlechterung des Zustands hinweisen kann. Der mRS wird häufig verwendet, um Studienpopulationen in verschiedene Kategorien einzuteilen, um den Grad der Behinderung in klinischen Studien zu standardisieren (Rankin, 1957).

1.2.7 Typ-1-PROGRESS-Studien posteriorer Schlaganfall

In diesem und in den folgenden Abschnitten wird die aktuelle Evidenz auf dem Gebiet der Prognoseforschung des posterioren Schlaganfalls gezielt zusammengetragen. Dabei wird die aktuelle Studienlage nach dem PROGRESS-Klassifikationssystem organisiert.

Insgesamt 20-30% aller ischämischen Insulte sind Schlaganfälle der hinteren Zirkulation (Bogousslavsky et al., 1988, Moulin et al., 1997, Caplan et al., 2004), wobei die Prävalenz im asiatischen Raum mit bis zu 40% etwas höher ist (Lee et al., 2001, Lee et al., 2006). Die Angaben zur Mortalität bei posteriorem Schlaganfall variieren stark. Nimmt man ältere Studien (vor dem Jahr 2015) basierend auf Registerdaten aus dem nicht asiatischen Raum als Grundlage, liegt die relative Mortalität zwischen 3,6% und 18,6% (Caplan et al., 2004). Die Mortalität für Schlaganfälle beim A. basilaris-Verschlusssubtyp liegt durchschnittlich wesentlich höher. Hier wurde eine relative Mortalität zwischen 6% und 58% beobachtet (Barlinn et al., 2012, Chandra et al., 2011, Costalat et al., 2011, Eckert et al., 2005, Jung et al., 2011, Kashiwagi et al., 2010, Lindsberg and Mattle, 2006, Lutsep et al., 2008, Miteff et al., 2011, Nagel et al., 2009, Pfefferkorn et al., 2010, Renard et al., 2008, Sairanen et al., 2011, Schonewille et al., 2005, Schonewille et al., 2009, Yu et al., 2010). Betrachtet man die kürzlich veröffentlichten, randomisiert-kontrollierten Studien zwischen 2019 und 2022 ergibt

sich eine relative Mortalität von ca. 38% (379/988) für den BAO-Subtyp nach 90 Tagen (Xu et al., 2022).

1.2.8 Typ-2-PROGRESS – Studien posteriorer Schlaganfall

1.2.8.1 Allgemeine Prognosefaktoren

Es gibt gut dokumentierte Prognosefaktoren für das Krankheitsbild Schlaganfall. Auf der Grundlage einer groß angelegten internationalen Fall-Kontroll-Studie wurde geschätzt, dass 90% der ischämischen Schlaganfälle auf zehn Faktoren zurückgeführt werden können: Bluthochdruck, Rauchen, das Taillen-Hüft-Verhältnis, Diät-Risiko und Diät-Risiko-Score, Bewegung, Diabetes mellitus, Alkoholkonsum >30Tage/Monat oder Alkoholexzesse, psychosozialer Stress und Depression, kardiale Vorerkrankungen und das Verhältnis von Apolipoprotein B zu A1 (O'Donnell et al., 2010).

In der Literatur werden zusätzlich weitere Risikofaktoren beschrieben.

1.2.8.2 Geschlecht

Männliches Geschlecht scheint mit einem höheren Risiko assoziiert zu sein, an einem Schlaganfall mit Beteiligung der hinteren Zirkulation zu erkranken als an einem anterioren Schlaganfall (Zeng et al., 2015).

1.2.8.3 Alter

Höheres Alter erwies sich als Prädiktor für eine erfolglose Rekanalisation (Meinel et al., 2019). Zusätzlich zeigen mehrere multivariate Prognosemodelle eine Assoziation von höherem Lebensalter und schlechtem funktionellen Überleben beziehungsweise jüngerem Lebensalter und gutem funktionellen Überleben bei BAO-Patienten (Greving et al., 2012), weshalb bei Patienten mit hohem Lebensalter (>75 Jahren) und BAO eventuell auf eine Thrombektomie verzichtet werden sollte (Ahmed et al., 2022). In weniger spezifizierten posterioren Schlaganfallpopulationen scheint der Prognosefaktor Alter möglicherweise nicht mit einem schlechten beziehungsweise guten funktionellen Überleben assoziiert zu sein (Kniep et al., 2022b, Mokin et al., 2016).

1.2.8.4 National Institutes of Health Stroke Score

Die "Schlaganfallskala des National Institutes of Health (engl. "National Institutes of Health Stroke Score", NIHSS) ist die in der modernen Neurologie am weitesten verbreitete Skala zur Bewertung von neurologischen Defiziten: Mehr als 500 000 medizinische Fachkräfte wurden bereits für die Anwendung der Skala über eine webbasierte Plattform zertifiziert. Jede klinische Studie der neurovaskulären Forschung erfordert eine Bewertung des neurologischen Schweregrads. Die NIHSS ist Goldstandard für die Bewertung des Schweregrads von Schlaganfällen (Lyden, 2017). Die NIHSS erfasst allerdings nur ungenau die Symptomvielfalt bei posteriorem Verschluss (Sato et al., 2008, Inoa et al., 2014). Trotzdem scheint der Parameter als Prognosefaktor eine geeignete prognostische Aussagekraft in Bezug auf ein gutes funktionelles Überleben nach 90 Tagen zu haben (Kniep et al., 2022b). Auch die Bestimmung des NIHSS-Scores nach 24 Stunden hat sich als potenzieller Parameter für die Prognoseabschätzung eines guten funktionellen Überlebens nach Thrombektomie bei posteriorem Schlaganfall erwiesen (Kniep et al., 2022a).

1.2.8.5 Störungen des Glukosestoffwechsels

Einige Studien deuten darauf hin, dass ein vorbestehender Diabetes mellitus eine stärkere Assoziation zum Schlaganfall der hinteren Zirkulation als zu anterioren Verschlüssen aufweist (Subramanian et al., 2009, Miyamoto et al., 2010, Zeng et al., 2015). In einer Metaanalyse aus dem Jahr 2020 zeigten Xun et al., dass ein bestehender Diabetes mellitus mit einem schlechten funktionellen Überleben nach Thrombektomie assoziiert ist (Xun et al., 2021). Allerdings handelt es sich bei den eingeschlossenen Studien primär um reine BAO – Kohorten. Eine Arbeit von Broocks, et. al. deutet ebenfalls darauf hin, dass der Glukosemetabolismus einen Einfluss auf den Krankheitsverlauf von Patienten mit Schlaganfall im hinteren Kreislauf hat (Broocks et al., 2022b).

1.2.8.6 Störung des Fettstoffwechsels

Die gleiche Metaanalyse von Xun et al. aus dem Jahr 2021 ergab Hinweise darauf, dass eine bestehende Hyperlipidämie möglicherweise einen positiven Effekt auf den

Krankheitsverlauf haben könnte (Xun et al., 2021). Die Ergebnisse von Greving et. al. unterstützen diese Beobachtung (Greving et al., 2012).

1.2.8.7 Vorhofflimmern

Vorhofflimmern als Risikofaktor spielt wahrscheinlich bei posteriorem Schlaganfall eine untergeordnete Rolle und ist stärker mit Schlaganfällen im vorderen Stromgebiet verknüpft (Miyamoto et al., 2010, Subramanian et al., 2009, Zeng et al., 2015).

1.2.8.8 Pathomechanismus/Ätiologie

Es kann schwierig sein, die Ätiologie bei akutem Schlaganfall zu bestimmen, aber es gibt klinische und bildgebende Merkmale, die bei der Bestimmung helfen können. Ein proximaler/mittlerer Basilarisverschluss mit Katheterspitzendurchgängigkeit, das Fehlen einer konvexen Kante im Bereich der Okklusion, ausgedehntere Verschlüsse und Gefäßwandverkalkungen in der computertomographischen Angiographie (CTA) können auf intrakranielle, atherosklerotische Verschlüsse und nicht auf ein embolisches Geschehen hindeuten (Lee et al., 2017, Lin et al., 2020). Chung et al. zeigten, dass es in Abhängigkeit von der zugrunde liegenden Pathoätiologie zu unterschiedlichen Krankheitsverläufen bei posterioren Verschlüssen kommt. Speziell atherosklerotische Schlaganfälle der großen Arterien und kardioembolische Schlaganfälle sind mit einem schlechteren funktionellen Überleben assoziiert. Die Autoren fordern die Etablierung zielgerichteter Behandlungsstrategien in Abhängigkeit von der jeweiligen Pathoätiologie und die Berücksichtigung der pathomechanistischen Heterogenität bei der Konzeption klinischer Studien (Chung et al., 2015). Wu et. al. konnten diesen generellen Zusammenhang nicht bestätigen. Allerdings beobachteten sie in einer Subanalyse bei Patienten <60 Jahren, dass ein embolischer Schlaganfall mit einer besseren funktionellen Unabhängigkeit assoziiert war (Wu et al., 2021).

1.2.8.9 Thrombuslokalisation

Prädilektionsstellen für das Auftreten thrombotischer Verschlüsse im posterioren, zerebralen Kreislauf sind die extrakranielle Vertebralarterie (ECVA), BA, die

intrakranielle Vertebralarterie (ICVA) und die posteriore, zerebrale Arterie (PCA) (Glass et al., 2002). Verschlüsse der BA oder ICVA sind mit einem schlechteren funktionellen Überleben assoziiert, Verschlüsse der ECVA oder PCA mit einer günstigeren Prognose (Glass et al., 2002).

1.2.8.10 Thrombuslänge und Thrombuslast

Die Thrombuslänge oder Thrombuslast ist ein wichtiger Parameter, der den Rekanalisationserfolg und das funktionelle Überleben bei anterioren Schlaganfällen beeinflusst (Sillanpaa et al., 2012, Tan et al., 2009, Yan et al., 2016, Legrand et al., 2013, Soize et al., 2015). Ein Zusammenhang zwischen Thrombuslänge und funktionellem Überleben ist auch bei posterioren Verschlüssen beschrieben worden (Brandt et al., 1996). Eine generell erhöhte Thrombuslast ist mit einem schlechten funktionellen Überleben assoziiert (Da Ros et al., 2016).

1.2.8.11 Neuroimaging Scores

1.2.8.11.1 Allgemein

Eine symptom-basierte Diagnostik oder Triage kann aufgrund unspezifischer klinischer Merkmale bei Schlaganfällen der hinteren Zirkulation schwierig sein (Fabritius et al., 2017), weshalb eine bildgebungsbasierte Diagnostik bei posterioren Schlaganfällen sogar noch wichtiger sein kann als bei anteriorem Schlaganfall. Trotz des Problems der erschwerten Diagnostik eines posterioren Verschlusses (Merwick and Werring, 2014) und der Schwierigkeit, eine hohe Zahl an Patienten in klinischen Studien zu rekrutieren, wurden eine Reihe neuroradiologischer Scores entwickelt, die den Kliniker dabei unterstützen sollen, evidenzbasierte Handlungsempfehlungen umzusetzen. Zu diesen Scores zählen der „Posterior Circulation Acute Stroke Prognosis Early CT Score“ (pcASPECTS) (Puetz et al., 2008b), der „Posterior Circulation Collateral Score“ (PC-CS) (van der Hoeven et al., 2016b) und der „Basilar Artery on Computed Tomography Angiography“ (BATMAN) Prognosescore für den Verschluss der Arteria Basilaris (Alemseged et al., 2017), sowie der pASCO (pcASPECTS + PC-CS) (Broocks et al., 2022c).

1.2.8.11.2 pcASPECTS

Der pcASPECTS-Score ist der am häufigsten verwendete Score und prognostiziert akkurat jene Patienten, bei denen eine Behandlung absolut wirkungslos ist. Er ist fehleranfällig bei speziellen Subpopulationen mit moderatem pcASPECT (Sang et al., 2021). Die Notwendigkeit eines Scores, der universell zuverlässig, d.h. auch bei speziellen Subpopulationen funktioniert, ist enorm wichtig. Die jüngsten enttäuschenden Ergebnisse der BASIC-Studie machen die Notwendigkeit optimaler Auswahlkriterien für eine mechanische Rekanalisationstherapie noch zwingender, um einen effektiven Nutzen durch eine interventionelle Behandlung zu erzielen (Langezaal et al., 2021).

1.2.8.11.2.1 Pathophysiologische Grundlagen

Der „Alberta Stroke Program Early CT Score“ (ASPECTS) wurde ursprünglich als Messinstrument entwickelt, um frühe ischämische Veränderungen der CT- Bildgebung bei Patienten mit akutem ischämischen Schlaganfall der mittleren Hirnarterie (engl. "middle cerebral artery; MCA") zu quantifizieren (Barber et al., 2000, Pexman et al., 2001). Als frühe ischämische Veränderungen im CT werden

- 1.) Hypodensitäten
- 2.) Verlust der Grau-Weiß-Differenzierung
- 3.) Schwellung der Kortikalis und Verlust der Hirnsulkusstrukturierung

bezeichnet. Diese Veränderungen sind durch das zytotoxische Ödem und konsekutive Hirnschwellung bedingt. Bei der zytotoxischen Ödembildung handelt es sich um einen präterminalen Prozess vor dem eigentlichen Zelltod, bei dem sich extrazelluläres Na^+ und andere Kationen in Neuronen und Astrozyten intrazellulär anreichern, was zum Teil auf ein Versagen der energieabhängigen Extrusionsmechanismen zurückzuführen ist. Der Einstrom von Kationen wiederum führt zum Einstrom von Anionen zur Aufrechterhaltung der elektrischen Neutralität. In Kombination bedingen diese Phänomene den Einstrom von Wasser, was zu einer osmotischen Ausdehnung der Zelle, dem zytotoxischen Ödem, führt. Das zytotoxische Ödem selbst verursacht nicht direkt die makroskopische Hirnschwellung, aber durch die Bildung des zytotoxischen Ödems verarmt der extrazelluläre Raum an Na^+ , Cl^- und Wasser,

wodurch ein neuer elektrochemischer Gradient für diese Moleküle über der kapillären Blut-Hirn-Schranke entsteht. Bei entsprechenden Veränderungen der Kapillardurchlässigkeit führt der durch das zytotoxische Ödem geschaffene neue elektrochemische Potenzialunterschied zur transkapillären Bildung eines ionischen Ödems. Dieser zweite Prozess wird als Hirnschwellung deutlich (Simard et al., 2007).

1.2.8.11.2.2 Scoring

Der pcASPECTS setzt sich aus 10 Punkten zusammen. Jeweils ein Punkt wird abgezogen für frühe ischämische Veränderungen in der CT im linken oder rechten Thalamus-, Kleinhirn- oder hinteren Hirnarterien-Territorium (PCA) und je 2 Punkte für frühe ischämische Veränderungen in einem beliebigen Teil des Mittelhirns oder der Pons. Ein pcASPECTS-Score von 10 bedeutet das Fehlen einer sichtbaren Ischämie der hinteren Zirkulation, ein Wert von 0 bedeutet frühe ischämische Veränderungen in allen pcASPECTS-Regionen (Puetz et al., 2008b).

1.2.8.11.3 Kollateralengraduierungssysteme

1.2.8.11.3.1 Pathophysiologische Grundlagen

Die Entwicklung und die Auswirkungen des ischämischen Schadens beim akuten Schlaganfall hängen von der Schwere und Dauer der zerebralen Ischämie ab, die wiederum entscheidend von der kollateralen Zirkulation beeinflusst wird (Sato et al., 1991, Jones et al., 1981, Yuh WTC, 1992, Mueller DP, 1991). Im speziellen konnte gezeigt werden, dass das Vorhandensein und die Größe der ischämischen Penumbra durch die kollaterale Zirkulation beeinflusst wird (Yuh et al., 1999). Funktionell wird als zerebraler Kollateralkreislauf das Netzwerk von Gefäßen bezeichnet, das den zerebralen Blutfluss stabilisiert, wenn die primäre arterielle Versorgung, bedingt durch Thromben, Thromboembolien oder einer hämodynamische Beeinträchtigung oder einer Kombination dieser Pathomechanismen versagt (Liebeskind, 2003). Es können arterielle von venösen Kollateralkreisläufen unterschieden werden. Arteriell kann man weiterhin grob zwei Subkreisläufe unterscheiden: a) die extrakranielle, arterielle Versorgung und b) die intrakranielle Versorgung, wobei das intrakranielle Kollateralom in primären und sekundären Kreislauf unterteilt wird. Zu den primären intrakraniellen

Kollateralen gehören die arteriellen Segmente des Willis'schen Kreislaufs, während die Arteria ophthalmica und die leptomeningealen Gefäße zu den sekundären intrakraniellen Kollateralkreisläufen beitragen (Liebeskind, 2003).

Durch den Blutfluss zwischen den Hemisphären über die vorderen kommunizierenden Arterien und durch die Umkehr des Flusses im proximalen Abschnitt der vorderen Hirnarterie wird eine primäre kollaterale Unterstützung im vorderen Teil des Circulus willisii gewährleistet. Im hinteren Teil des Circulus willisii fungieren die proximalen hinteren Hirnarterien als zusätzliche interhemisphärische Primärkollateralen. Die hinteren kommunizierenden Arterien (engl. „posterior communicating artery, Pcom“) können den kollateralen Blutfluss in beide Richtungen zwischen dem anterioren und hinteren Kreislauf unterstützen. Weiterhin liefern im hinteren Kreislauf distale Äste der großen Kleinhirnarterien kollaterale Verbindungen zu den vertebrale und basilären Segmenten der hinteren Zirkulation. Kortikal fungieren leptomeningeale und durale Arteriolen als kollaterale Perforatoren für primäre Gefäße in dieser Region.

Weitere Anastomosen zwischen den distalen Segmenten der großen Hirnarterien tragen ebenfalls zu einem kollateralen Blutfluss bei. Die Anzahl und Größe dieser anastomosierenden Gefäße ist am größten zwischen den vorderen und mittleren Hirnarterien, sowie zwischen den mittleren und hinteren Hirnarterien. Zwischen den vorderen und hinteren Hirnarterien sind anastomosierende Gefäßverbindungen weniger stark ausgeprägt.

Venöse Kollateralen unterstützen den Abfluss des zerebralen Blutes, wenn die arteriellen Hauptwege verschlossen sind oder eine venöse Hypertonie entsteht. Die Anatomie der venösen Kollateralkreisläufe ist sehr variabel und erlaubt die Umleitung des Blutes über zahlreiche Abflüsse nach extraaxial. Wichtige venöse kollaterale Feeder sind die Vena Labbé, der sphenoparietale Sinus und die superfizielle mittlere zerebrale Vene (engl. „superficial middle cerebral vein, SMCV“) (Liebeskind, 2003).

Die erhebliche Variabilität der Anatomie des primären zerebralen Gefäßbettes führt zu einer großen anatomischen Asymmetrie und somit dazu, dass eine ideale vasale

Konfiguration in nur einer Minderheit der Fälle vorliegt. Es wurde zum Beispiel gezeigt, dass die Arteria communicans anterior bei 1% der Patienten fehlt, in 10% der Fälle fehlen die proximalen vorderen Hirnarterien oder sind hypoplastisch angelegt. Eine Hypoplasie der beiden hinteren Pcoms liegt in circa in 30% der Fälle vor (Hans Lippert 1985). Speziell beim akuten Schlaganfall wurde beobachtet, dass einige Kollateralkreisläufe weniger häufig vorhanden sind: wie z.B. der Plexus tectalis (Kollateralen zwischen den supratentoriellen Ästen der A. posteriores cerebri und infratentoriellen Ästen der superioren Kleinhirnarterie); der Orbitalplexus (Kollateralkreislauf zwischen der A. ophthalmica, fazialen Arterien, zwischen den mittleren meningealen Arterien des Oberkiefers und den Ethmoidalarterien), sowie das Rete mirabile caroticum, die die A. carotis interna mit der A. carotis externa kollateralisiert (Liebeskind, 2003).

Der Mechanismus kollateraler Rekrutierung ist ein dynamischer Prozess, dessen Ausprägung kurzfristig vom Kaliberdurchmesser und der Durchlässigkeit der primären Kollateralen abhängt, die *a priori* für eine sofortige Umleitung des zerebralen Blutflusses in ischämische Regionen funktionell zur Verfügung stehen. Maßgeblich bestimmende Parameter sind ein verminderter Blutfluss in den stromabwärts gelegenen Gefäßen (Meyer and Denny-Brown, 1957). Langfristig wird dieser Prozess auch durch die Verfügbarkeit der sekundären Kollateralen bestimmt, deren funktionelle Integration eine gewisse Zeit beansprucht und u.a. durch Faktoren wie die Angiogenese im Randbereich ischämischer Regionen gesteuert wird (Wei et al., 2001). Im Tiermodell konnte gezeigt werden, dass eine arterielle Hypertension den Prozess der Kollateralisierung verlangsamt und eine Anastomosenbildung mit deutlich schmalerem Kaliber und geringeren Kapazitäten zur Folge hat (Coyle and Heistad, 1991). Grundsätzlich ist die Darstellung kollateraler Kreisläufe mit den gegenwärtigen diagnostischen Methoden schwierig (Liebeskind, 2003). Ursprünglich wurde die Digitale Subtraktionsangiografie (DSA) zum Assessment der kollateralen Gefäße genutzt. Diese wurde jedoch aufgrund der ubiquitären Verfügbarkeit und der geringen Invasivität von der CTA-Bildgebung im neuroradiologischen Routinebetrieb abgelöst (Arnold et al., 2004, Brandt et al., 1999, Cross et al., 1998, Brandt et al., 1995, Singer et al., 2015). Ein Problem dieses diagnostischen Strategiewechsels zeigt folgendes Beispiel: Sekundäre leptomeningeale und piale Kollateralen zwischen den Hauptästen der Vertebral- und Basilararterie oder Anastomosen zwischen distalen MCA- und PCA-

Ästen, sowie Anastomosen zwischen den spinalen Arterien und den vertebralen bzw. basilären Ästen können nicht zuverlässig evaluiert werden (van der Hoeven et al., 2016a). Bestimmte kollaterale Flussreserven können jedoch auch mittels CTA – Bildgebung gut dargestellt werden, wie zum Beispiel die posterioren, kommunizierenden Arterien, die hinteren inferioren Kleinhirnarterien oder die fetalen Pcoms. Diese sind wesentlicher Bestandteil protektiver Kollateralkreisläufe bei posterioren Schlaganfällen: Wenn der Stamm der Arteria basilaris okkludiert, kommt es zu einem Perfusionsdruckabfall an der Einmündung der hinteren Hirnarterien (PCAs), was zu einer retrograden Füllung über die Pcoms führt (falls sie vorhanden sind). Diese Flussumkehr kann die arterielle Versorgung distal des Thrombus aufrechterhalten und gewährleistet so die Blutversorgung in der ischämischen Region, speziell durch Aufrechterhaltung der Perfusion in den perforierenden Arterien und potenziell auch in den superioren zerebralen Arterien (engl. „superior cerebral artery, SCA“)-Ästen. Ein zweiter kollateraler Flussweg über die hinteren inferioren Kleinhirnarterien (PICAs), die von der vorderen Spinalarterie versorgt werden, kann ebenfalls als alternativer Perfusionsweg dienen (Lindsberg et al., 2015): Zwei fetale Pcoms (mit fehlenden P1-Segmenten) können zudem bei einem BAO schützend wirken, da das ischämiegefährdete Gebiet kleiner ist, weil die perforierenden Arterien zum Mittelhirn und Thalamus über den vorderen Kreislauf versorgt werden (Hong et al., 2009, Lindsberg et al., 2015). Ein weiterer Kreislauf kann zusätzlich als kollaterale Reserve bei posteriorem Schlaganfall dienen: Wenn die Vertebralarterien an ihrem Ursprung durchgängig bleiben, ist ein Fluss durch die hinteren inferioren Kleinhirnarterien in die anterioren inferioren Kleinhirnarterien (engl. “anterior inferior cerebellar artery, AICA“), die SCAs und die perforierenden Arterien der hinteren zerebralen Zirkulation möglich (Lindsberg et al., 2015).

1.2.8.11.3.2 Scoring

In der Literatur sind verschiedene Bewertungssysteme zur Beurteilung der kollateralen Reserve bei Schlaganfällen der hinteren Zirkulation beschrieben. Im Folgenden werden einige ausgewählte Gradierungssysteme besprochen. Der Kollateralenfluss im posterioren Kreislauf wurde als Prognosemarker bei akutem Basilararterienverschluss untersucht. Der primäre Kollateralweg des hinteren Kreislaufs zum vorderen Kreislauf erfolgt über die posteriore kommunizierende Arterie (Pcom). In frühen Studien wurde

mittels DSA ein günstiges Überleben von Patienten mit kollateraler Füllung der A. basilaris beobachtet (Cross et al., 1998).

1.2.8.11.3.3 Posterior Circulation Collateral Score (PC-CS)

In einer retrospektiven Studie mit Patienten, die in das BASICS-Register aufgenommen wurden, entwickelten Hoeven et al. einen 10-Punkte-Score zur Quantifizierung des Kollateralflusses in den Pcoms und den Kleinhirnarterien. Jeder perfundierten Kleinhirnarterie [posteriore inferiore Kleinhirnarterie, anteriore inferiore Kleinhirnarterie und Superior Kleinhirnarterie] wird 1 Punkt pro Hemisphärenseite zugeordnet. Für jede vorhandene Pcom wird 1 Punkt vergeben, wenn der Durchmesser kleiner ist als das ipsilaterale P1-Segment der posterioren Zerebralarterie, und 2 Punkte, wenn der Durchmesser größer ist als das ipsilaterale P1-Segment der PCA. Der PC-CS wurde trichotomisiert in schlecht (0-3), mittel (4-5) und gut (6-10) kollateralisiert. Patienten mit einem schlechten PC-CS hatten häufiger ein schweres Defizit (Koma, Tetraplegie oder Locked-in-State) als Patienten mit einem mittleren oder guten PC-CS. Beim Vergleich eines guten mit einem schlechten PC-CS ergab sich ein statistisch signifikant geringeres Risiko (RR, 0,74; 95% CI, 0,58-0,96) für ein schlechtes funktionelles Überleben (mRS von 4 oder höher). Außerdem war das Vorhandensein von einem oder mehreren offenen Pcom-Arterien mit einem statistisch signifikant niedrigeren Risiko (RR, 0,79; 95% CI, 0,66-0,95) für ein schlechtes funktionelles Überleben assoziiert. Wenn mindestens eine Pcom ein größeres Kaliber als das ipsilaterale P-1-Segment der PCA aufwies, war das Risiko für ein schlechtes funktionelles Überleben in der multivariaten Analyse ebenfalls geringer (RR, 0,76; 95% CI, 0,61-0,96) (van der Hoeven et al., 2016b).

1.2.8.11.3.4 Collateral Score (CS)

In einer anderen retrospektiven Studie wurde speziell der prognostische Nutzen von Pcom-Kollateralen untersucht. Goyal und Kollegen definierten ein Graduierungssystem ausschließlich aufgrund des Vorhandenseins oder Nichtvorhandenseins von Pcoms. Bei Fehlen von bilateralen Pcoms wird ein Score von 0 vergeben, ein Score von 1 für das Vorhandensein von einer Pcom und ein Score von 2 für das Vorhandensein von bilateralen Pcoms. Die Scores wurden dichotomisiert

in schlecht (0 oder 1) oder gut (2). Im Vergleich zu Patienten mit einem schlechten CS hatten Patienten mit einem guten CS eine höhere Wahrscheinlichkeit für ein gutes funktionelles Überleben und eine Verschlechterung des NIHSS während des Krankenhausaufenthaltes. Das gute funktionelle Überleben unterschied sich in einem multivariaten Regressionsmodellen jedoch nicht signifikant in den beiden CS-Kohorten (Goyal et al., 2016).

1.2.8.11.4 Thrombuslast und Thrombuslokalisation

Alemseged und Kollegen entwickelten ein weiteres CTA-Scoring-System, den BATMAN-Score, der neben dem Parameter "Kollateralisierung" den Einfluss von Thrombuslast und Thrombuslokalisation auf das funktionelle Überleben von Patienten mit Basilarisverschluss berücksichtigt (Alemseged et al., 2017).

1.2.8.11.4.1 Pathophysiologische Grundlagen

Ein ungünstiges funktionelles Überleben ist stark mit einer erfolglosen Rekanalisation assoziiert (Berkhemer et al., 2015, Campbell et al., 2015, Goyal et al., 2015, Saver et al., 2015). Komplexe langstreckige thrombotische Verschlüsse können die mechanische Rekanalisation erschweren. Ein schlechtes funktionelles Überleben kann die Folge sein (Brandt et al., 1996, Da Ros et al., 2016).

1.2.8.11.4.2 Basilar Artery on Computed Tomography Angiography (BATMAN) Score

Das Punktesystem basiert auf dem Vorhandensein oder Nichtvorhandensein perfundierter Arterien in der CTA. Die Punkte werden wie folgt vergeben: 1 Punkt für beide Vertebralarterien (die Vertebralarterien werden als ein Segment betrachtet); 1 Punkt für das proximale Segment der Arteria basilaris; 1 Punkt für das mittlere Segment der Arteria basilaris ; 1 Punkt für das distale Segment der Arteria basilaris; 1 Punkt für das Vorhandensein jedes P-1 Segments der PCA und 2 Punkte für die Füllung jeder Pcom. Für die Füllung einer hypoplastischen Pcom wird 1 Punkt anstelle von zwei Punkten vergeben. Eine Punktzahl von 10 deutet auf eine normale arterielle Füllung des hinteren Kreislaufs hin und eine Punktzahl von 0 entspricht keiner Füllung

des hinteren Kreislaufs. In ihrer multivariaten Analyse war, sowohl in der Entwicklungs- als auch in der Validierungskohorte, ein ungünstiger BATMAN-Score (definiert als <7) signifikant mit einem schlechten funktionellen Überleben assoziiert (definiert als ein mRS von 4 oder höher). Ein BATMAN-Score von <7 prognostiziert mit einer Sensitivität von 84 % und einer Spezifität von 76 % ein schlechtes Outcome (Alemseged et al., 2017).

1.2.8.11.5 pASCO

Patienten mit einer guten kollateralen Reserve, trotz eines niedrigen ASPECTS zeigen ein gutes funktionales Überleben nach mechanischer Thrombektomie (van der Hoeven et al., 2016b). Allerdings eignet sich der Collateral Score als alleinstehender Parameter nicht ideal, um das individuelle Risikoprofil eines Patienten vollständig abzubilden. Obwohl bei den meisten Patienten das Vorhandensein von guten Kollateralen gut mit dem Ausmaß der Penumbra korreliert, ist dies nicht immer der Fall. Um die Vorteile von PC-CS und ASPECTS zu kombinieren, entwickelten McDonough et al. den ASCO-Score für Schlaganfälle in der vorderen Zirkulation. Dieses Konzept wurde auch auf Schlaganfälle der hinteren Zirkulation übertragen. Der pcASPECTS und PC-CS wurden unter dem Namen pcASCO-Score zusammengefasst. Der Score zeigt eine bessere prognostische Aussagekraft mit einer guten Diskriminationsfähigkeit, wobei ein klarer Cut-off von 14 Punkten auf der pcASCO-Skala auf ein gutes funktionelles Überleben für die gesamte Patientenkohorte hindeutet (Broocks et al., 2022c).

1.2.8.11.6 Net water uptake (NWU)

Die quantitative Messung der Netto-Wasseraufnahme (engl. "Net Water Uptake, NWU") mittels Computertomografie wurde als wichtiger Surrogat-Biomarker für die Entwicklung eines malignen Ödems bei Schlaganfällen der vorderen Zirkulation beschrieben. Die quantitative Messung der Wasseraufnahme der Läsion kann die Identifizierung von Patienten besonders in einer frühen Phase des Krankheitsverlaufs mit einem hohen Risiko für die Entwicklung eines malignen Infarkts verbessern, welcher mit einer hohen Mortalitätsrate assoziiert ist (Broocks et al., 2021). NWU als prognostischer Marker für ein gutes funktionales Überleben bei posterioren Schlaganfällen ist in der Literatur jedoch noch nicht beschrieben worden.

1.2.8.11.6.1 Pathophysiologische Grundlagen

Während eines Schlaganfalls wird die Blut-Hirn-Schranke infolge des Zusammenbruchs der endothelialen Zell-Zell-Verbindungen im Rahmen der ionischen Ödembildung, gestört, was im weiteren Verlauf zur Bildung eines vasogenen Ödems mit weiterer Nettowasseraufnahme führt. Bei großen raumfordernden Infarkten kann dies zu schweren Komplikationen führen.

1.2.8.11.6.2 Scoring

Es existieren standardisierte Verfahren zur Quantifizierung der aufgenommenen Netto-Wassermenge (Broocks et al., 2018, Minnerup et al., 2016). Die Messung des NWU basiert auf einer densitometrischen Analyse des ischämischen Kerns im Vergleich zur automatisch gespiegelten kontralateralen, gesunden Region des Infarktkerns, die als physiologische Referenz genutzt wird.

1.2.9 Typ-3-PROGRESS-Studien posteriorer Schlaganfall

Der Schlaganfall ist ein heterogenes Krankheitsbild im Hinblick auf Phänotyp und Prognose. Einige Patienten mit Schlaganfall profitieren in Abhängigkeit von ihren klinischen, radiologischen oder biologischen Merkmalen mehr von einer endovaskulären Behandlung als andere. Daher ist die Entwicklung und Etablierung von individuellen Prognosemodellen sehr wichtig, um Rückschlüsse auf Krankheitsmechanismen zu erhalten und eine risikobasierte Stratifizierung von Patienten für Logistik, Forschung und Klinik zu ermöglichen. Insbesondere in einer anspruchsvollen klinischen Situation unter Zeitdruck kann ein Instrument, das den Krankheitsverlauf eines Patienten voraussagt, hilfreich sein, um therapeutische Entscheidungsfindungsprozesse zu optimieren und dem Patienten sowie den Angehörigen wichtige prognostische Informationen bereitzustellen. Daher wurden besonders für Schlaganfälle der vorderen Zirkulation, mehrere Prognosemodelle entwickelt, validiert und in der klinischen Versorgung eingesetzt (Boehme et al., 2014, Ishkanian et al., 2011, Perez de la Ossa et al., 2016, Raza and Rangaraju, 2018) sowie miteinander verglichen (Kremers et al., 2022), um das funktionelle Überleben einzelner Patienten, die potenziell mit EVT behandelt werden können, möglichst genau zu

prognostizieren. Neben klassischen statistischen Methoden zur Modellentwicklung werden zunehmend auch Algorithmen basierend auf maschinellem Lernen (ML) in diesem Zusammenhang genutzt (Zeng et al., 2022). In den letzten Jahren gab es bereits viele „Big-Data-Initiativen“ in der Schlaganfallforschung: Das virtuelle internationale Schlaganfall-Studienarchiv (VISTA) verfügt über klinische Daten mit dokumentierten Basisparametern wie NIHSS, Komorbiditäten oder Laborparametern von 82.000 Patienten (Ali et al., 2007). Bildgebungsdatensätze von Schlaganfallpatienten sind jedoch im Schnitt wesentlich kleiner (z.B. 2800 strukturelle Scans in der MRI-GENIE-Studie, 2950 Scans im Meta-VCI-Kartenkonsortium, 1800 Scans in ENIGMA oder 1333 Scans in einer unizentrischen Studie) (Liew et al., 2022, Schirmer et al., 2019, Weaver et al., 2021), weshalb noch häufiger Prognosemodelle mittels konventioneller ML-Methoden entwickelt werden als mithilfe von „Deep learning (DL)“-Methoden (Zeng et al., 2022). Die Leistung konventioneller ML- und DL-Modelle unterscheidet sich nicht wesentlich voneinander oder von der Leistung bereits klassischer prognostischer Scores. Obwohl einige ML-Modelle eine hohe prognostische Aussagekraft erzielten, wurden die meisten mit kleinen Datensätzen entwickelt und nicht solide extern validiert (Zeng et al., 2022). Bei posteriorem Schlaganfall ist die Datenlage sehr gering. Es existieren keine systematischen Reviews zur Evaluation verschiedener bestehender Prognosemodelle. Viele der existierenden Modelle in diesem Zusammenhang sind nicht intern und extern validiert oder nach PROGRESS-Standard evaluiert.

1.2.9.1 Klassische Prognosemodelle bei posteriorem Schlaganfall

Viele Prognosemodelle sind in reinen BAO-Kollektiven entwickelt worden. Relevante Prädiktoren sind Alter, NIHSS, pcASPECTS, PC-CS und pASCO (Broocks et al., 2022a, Broocks et al., 2022c, Greving et al., 2012). Ein aktuelles, von Broocks et al. entwickeltes Prognosemodell für den thrombektomierten BAO-Schlaganfallsubtyp zur Abschätzung des funktionellen Überlebens nach 90 Tagen (Broocks et al., 2022c) zeigt eine sehr gute Diskriminationsfähigkeit mit einer AUC = 0,91 (Broocks et al., 2022c). Weitere vergleichbare multivariable Regressionsmodelle von Broocks, et al. und Greving et al. zeigen eine ähnliche Diskriminationsfähigkeiten: AUC = 0,93 (95%CI: 0,81–0,99) (Broocks et al., 2022a) beziehungsweise 0.80 (0,77–0,84) (Greving et al., 2012).

Für ein gemischtes posteriores Schlaganfallkollektiv ist ein multivariates Regressionsmodelle von Lin et. al. beschrieben, in dem die Prädiktoren NIHSS, Alter und pcASPECTS zur Prognoseabschätzung des mRS nach 90 Tagen genutzt wurden. Es zeigt eine AUC = 0,77 (95%CI: 0,68–0,86). Weitere Performanceparameter wurden nicht angegeben (Lin et al., 2018).

1.2.9.2 ML-Prognosemodelle bei posteriorem Schlaganfall

ML-basierte Prognosemodelle in gemischten posterioren Schlaganfallkollektiven erzielen ähnliche Diskriminationsleistungen wie konventionell entwickelte Modelle (Feyen et al., 2022, Kniep et al., 2022b).

1.2.10 Typ-4-PROGRESS Studien posteriorer Schlaganfall

Aktuell existieren keine Typ-4-PROGRESS Studien zu echten Prädiktoren des Therapieansprechens bei posteriorem Schlaganfall.

1.3 Material und Methoden

Diese Arbeit wurde entsprechend den nach TRIPOD geforderten Dokumentationskriterien verfasst (Moons et al., 2015). Ziel dieser Arbeit war die Entwicklung eines statistischen Modells zur Prognoseabschätzung des funktionellen Überlebens von Patienten nach 90 Tagen nach ischämischem Schlaganfall der hinteren Zirkulation (unabhängig von einem bestimmten, betroffenen Gefäßast) und erfolgter mechanischer Intervention. Das Modell soll bei der Evaluation zur mechanischen Rekanalisation von Patienten mit Schlaganfällen des hinteren Stromgebietes in der Notfallsituation genutzt werden können und transparent quantitative Aussagen darüber ermöglichen, wie sich ein Patient nach erfolgter mechanischer Thrombektomie funktionell entwickeln wird.

1.3.1 Patienten

In dieser multizentrischen, retrospektiven Beobachtungsstudie wurden Patienten mit akutem Arterienverschluss der zerebralen hinteren Zirkulation und multimodaler Aufnahme-CT analysiert, die zwischen April 2015 und Oktober 2020 in der Akutsituation aufgenommen wurden.

Die *a priori* definierten Einschlusskriterien waren:

- 1) akuter ischämischer Schlaganfall im hinteren Kreislauf diagnostiziert mit multimodaler Computertomografie bei Aufnahme [Nicht-Kontrast-CT (engl. „Non-Contrast Computer Tomography, NCCT“), CTA und CT-Perfusion (CTP)] innerhalb von 12 Stunden nach Auftreten der Symptome
- 2) Verschluss der Arteria basilaris
- 3) Abwesenheit von intrakraniellen Blutungen und vorbestehenden Infarkten im Aufnahme-NCCT
- 4) Abwesenheit von signifikanten Bildgebungsartefakten.

Die erfolgreiche endovaskuläre Rekanalisation wurde anhand der Klassifikation nach Thrombolyse bei zerebralem Infarkt (engl. „Thrombolysis in Cerebral Infarction, TICI“) gemessen und als TICI-Score $\geq 2b$ am Ende des Eingriffs definiert.

Die Studie wurde von der zuständigen Ethikkommission genehmigt. Alle Studienprotokolle und Prozeduren wurden in Übereinstimmung mit der Deklaration von Helsinki durchgeführt.

1.3.2 Bildgebung

Alle Patienten erhielten bei der Aufnahme ein umfassendes Schlaganfall-Bildgebungsprotokoll mit NCCT, CTA und dynamischer zeitaufgelöster CTP, die in gleicher Reihenfolge auf 256- oder 384-Dual-Slice-Scannern (Philips iCT 256, Siemens Force) durchgeführt wurden. NCCT: 120 kV, 280-340 mA, 5,0 mm Schichtrekonstruktion, 1 mm Inkrement; CTA: 120 kV, 260-300 mAs, 5,0-mm-Schichtrekonstruktion, 1-mm-Inkrement, 80 mL hochjodiertes Kontrastmittel und 50 mL NaCl-Spülung mit 4 mL/Sekunde Laufrate; CTP: 80 kV, 200-250 mA, 5 mm Schichtrekonstruktion (max. 10 mm), Schichtabtastrate 1,50 s (min. 1,33 s), Scanzeit 45 s (max. 60 s), biphasische Injektion mit 30 ml (max. 40 ml) hochjodiertem Kontrastmittel mit 350 mg Jod/ml (max. 400 mg/ml) injiziert mit 6 ml/s gefolgt von 30 ml Natriumchlorid-Chaser-Bolus.

1.3.3 Bildanalyse

Zwei erfahrene Board-zertifizierte Neuroradiologen bewerteten jeweils unabhängig voneinander pcASPECTS („pcASPECTS“), BATMAN („BATMAN“), PC-CS („PCCS“) und JAUSS Score („cerebellaredemascore“) in der ersten Computertomographie bei Aufnahme. NWU wurde bei Aufnahme („NWU_in_prozent“), 24 Stunden nach Aufnahme („NWU2_in_prozent“) und im Verlauf („NWU_FU_in_prozent“) bestimmt.

Beide Befunder waren zum Zeitpunkt der Bewertungen der erhobenen Daten verblindet. Befundungsunstimmigkeiten wurden durch gemeinsame Diskussion der beiden Rater geklärt.

Der pcBATAS–Score (pcBATASs) wurde als 20-Punkte-wertiges Graduierungssystem definiert, bestehend aus dem 10-Punkte wertigen „Basilar Artery on Computed Tomography Angiography Prognostic Score (BATMAN)“ dem 10-Punkte-wertigen „posterior circulation Acute Stroke Prognosis Early CT Score (pcASPECTS)“

(Alemseged et al., 2017, Puetz et al., 2008b). Das Scoring für BATMAN und pcASPECTS wurde von zwei erfahrenen, zertifizierten Neuroradiologen durchgeführt, und Diskrepanzen wurden durch gemeinsame Diskussion der beiden Befunder geklärt. Abbildung 2 und 3 veranschaulichen die Berechnung des pcBATASs an einem Fallbeispiel: Für den pcASPECTS ergibt sich bei frühen ischämischen Veränderungen in der linken Occipitallappenregion, im linksseitigen Thalamuskern, in der Pons und im gesamten Kleinhirn ein Wert von 3. Nach Gabe von Kontrastmittel zeigt sich ein Basilarisverschluss im B3 – Segment mit beidseits okkludierten P1-Segmenten (BATMAN = 6). Das Vorhandensein von beidseitigen Pcoms ermöglicht die Perfusion distaler Posteriorsegmente.

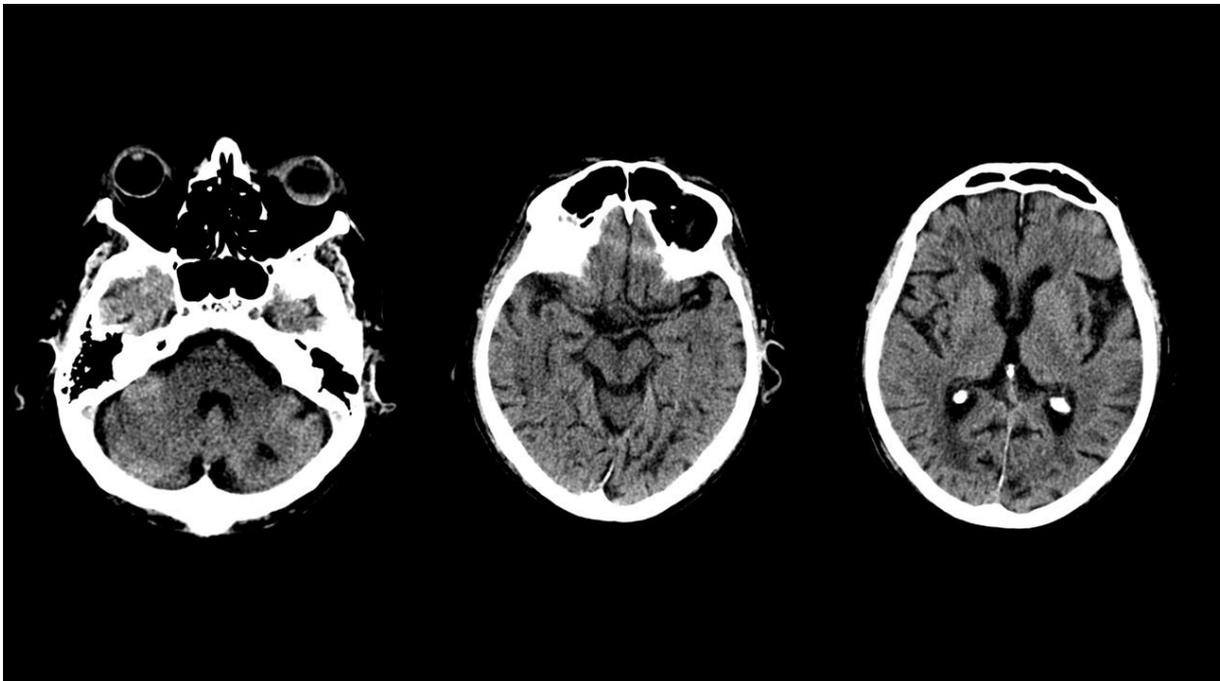


Abbildung 2: Native Computertomografie eines Patienten mit einem niedrigen pcBATASs von 9 (pcASPECTS = 3 + BATMAN = 6) in einem Patienten mit akutem Basilarisverschluss ca. 4 Std. nach Symptombeginn



Abbildung 3: Kontrastmittel gestützte Computertomografie eines Patienten mit einem niedrigen pcBATASs von 9 (pcASPECTS = 3 + BATMAN = 6) in einem Patienten mit Basilariskopfverschluss ca. 4 Std. nach Symptombeginn. Die gut perfundierten P2-Segmente beidseits lassen eine gute kollaterale Reserve durch bilaterale Pcoms (hier im Schnittbild nicht gezeigt) vermuten.

1.3.4 Klinische und andere paraklinische Basisparameter

Folgende klinischen und paraklinischen Parameter wurden zusätzlich aus den Registern manuell extrahiert: Patientenalter bei Aufnahme (Age), Geschlecht (Sex), mRs bei Aufnahme (mRspreStroke), NIHSS bei Aufnahme (NIHSSadmi), venöse Blutglucose bei Aufnahme (Glucoseinitial), Blutdruck bei Aufnahme (RRinitial), Grunderkrankung Bluthochdruck (ComorbidityHTN), Grunderkrankung Diabetes mellitus Typ 2 (ComorbidityDM), Grunderkrankung Hypercholesterinämie beziehungsweise Fettstoffwechselstörung (ComorbidityHLP), Grunderkrankung Vorhofflimmern (ComorbidityVHF), Rauchverhalten (Comorbidity_smoking_war_2), Abciximab i.v (Abciximab_use), einfache Plättchenhemmung (antiplateletTH_one), doppelte Plättchenhemmung (antiplateletTH), Antikoagulation (anticoagulation), maligner Infarkt in der Aufnahme-CT (MALIGNANT), intravenöse Thrombolyse (IVThrombolysis), Zeit Symptombeginn bis erste Bildgebung in Stunden (timeonsetimagingh), Zeit erste Bildgebung bis Leistenpunktion in Stunden (timeImagingAngioh), langstreckiger Verschluss (longOcclusion), Beteiligung der A. basilaris (BAO_involvement), intrakranielle Thrombuslokalisation

(intracran_thrombus_localisation), Pathomechanismus/Pathoätiologie nach CT/DSA Charakteristika (etiology), lokale Thrombolysetherapie (LokaleLyse), Rekanalisationsversuche (No_of_passages), TICI-Score (TICI_cod), TICI 2b/3 (successRecan), NIHSS 24h nach Aufnahme (NIHSS24h) postinterventioneller Aufenthalt auf der Intensivstation (IST) in Tagen (ICU_days), NIHSS bei Entlassung (NIHSSdis), mRS 90 Tage nach Diagnosestellung (mRs90).

1.3.5 Statistische Auswertung

1.3.5.1 Übersicht der statistischen Methodik

Das primäre Ziel der statistischen Auswertung bestand darin, ein prognostisches Modell zu entwickeln, das 1) alle verfügbaren Informationen des Datensatzes nutzt und 2), das optimal in Bezug auf mögliche, wirkende Underfitting oder Overfitting-Mechanismen korrigiert wurde, die potentiell im Rahmen der verschiedenen Modellentwicklungsprozesse auftreten können.

Zunächst wurde der Datensatz quantitativ und qualitativ auf fehlende Werte überprüft. Weiterhin wurden die einzelnen Variablen auf Normalverteilung und Verzerrung untersucht und ein individualisiertes Imputationsmodell entwickelt. Die Entwicklung des Prognosemodells erfolgte anschließend mittels verschiedener Regressionsmodelle. Dabei wurden drei verschiedene Algorithmen miteinander verglichen:

1. Logistische Regression mit Rückwärtselimination und uniformem Shrinkage
2. Ridge- Regression
3. FLAC-Regression

Die entwickelten Modelle wurden mittels ordinärem Bootstrapping intern validiert und in Bezug auf Diskrimination, Kalibrierung und allgemeine Leitungsfähigkeit übergeprüft. Um den klinischen Nutzen zu evaluieren, erfolgte eine Entscheidungskurvenanalyse unter Bestimmung des Net Benefits. Für die statistische Auswertung wurde die Software R (Version R-4.1.2) genutzt.

1.3.5.2 Deskriptive Auswertung des unvollständigen Datensatzes

Der Datensatz wurde zunächst deskriptiv ausgewertet. Dafür wurde die Kohorte dichotomisiert: Ein gutes Ergebnis wurde als gutes funktionelles Überleben 90 Tage nach Aufnahme ($mRS \leq 2$) definiert. Für kategoriale Daten wurden absolute und relative Häufigkeiten angegeben. Die ordinalen und metrischen Variablen wurden durch Mittelwert und Standardabweichung beziehungsweise Median und Interquartilsbereich (IQR) beschrieben. Für alle Variablen wurde der prozentuale Anteil an fehlenden Werten angegeben.

1.3.5.3 Multiple Imputation

Der Datensatz wurde auf fehlende Werte und deren Verteilungsmuster (engl. „missing data pattern“) geprüft. Das Muster des fehlenden Datensatzes wurde mittels `md.pattern()`-Funktion des „MICE“ Packages analysiert (Buuren and Groothuis-Oudshoorn, 2011). Weiterhin wurden mittels Dichtediagramm und Histogramm, sowie QQ-Diagramm die ordinalen und metrischen Variablen grafisch auf Normalverteilung geprüft.

Für metrische Variablen wurde in dieser Arbeit das modifizierte Predictive-Mean-Matching (PMM) – Verfahren „Midastouch“ genutzt (Philipp Gaffert, 2016). Midastouch zeigt bei kleinen Datensätzen geringere Verzerrungen als die von MICE angebotene Methode „PMM“ (Kleinke, 2018). Für die Imputation binär skalierten Parameter wurden binäre logistische Regressionsmodelle verwendet. Geordnete kategoriale Parameter mit hoher Kategorienzahl wurden ebenfalls mittels Midastouch imputiert, da das Verhältnis zwischen der Anzahl der zu interpolierenden Parameter und der Anzahl der Ereignisse klein ist, was mit Nicht-PMM-Verfahren zu Schätzungsproblemen führen kann. Robustere Imputationsmethoden wie PMM, „RandomForest“ und ähnliche sind hier von Vorteil (van Buuren, 2018). Für die Imputation ungeordneter kategorialer Variablen geringer Kategorienzahl wurden multinomiale Modelle verwendet. Der Outcomeparameter `mRS90` fließt zunächst als kontinuierliche Variable in das Imputationsmodell mit ein und wurde erst nach erfolgter Imputation dichotomisiert. (Floden and Bell, 2019). Der `pcBATAS` wurde nach erfolgter Imputation aus `BATMAN` und `pcASPECTS`-Score gebildet. Methodisch wurde zunächst unter Berücksichtigung

bereits erwiesener klinischer Bedeutsamkeit, klinischer Praktikabilität, Wirtschaftlichkeit, sowie pathophysiologischer Zusammenhänge eine Vorauswahl an Kandidatenprädiktoren und potenzieller Hilfsvariablen getroffen (Broocks et al., 2022c, Broocks et al., 2022b, Buuren and Groothuis-Oudshoorn, 2011, Caplan et al., 2004, Chung et al., 2015, Lyden, 2017, Mantatzis, 2022, Miyamoto et al., 2010, O'Donnell et al., 2010, Rangaraju et al., 2016, Sato et al., 2008, Subramanian et al., 2009, Voetsch et al., 2004, Wu et al., 2021, Xun et al., 2021, Zeng et al., 2015). Unvollständige Variablen mit fehlenden Werten $> 50\%$ wurden nicht in das MI-Modell mit aufgenommen. Ebenfalls wurden unvollständige Variablen

- (1) mit fehlenden Werten im Bereich 30-50% und einer Verzerrung > 0.81 bei Fallzahlen im Bereich $N = 100$ oder
- (2) einer groben Verzerrung (skewness $> 1,6$) a priori nicht in das MI – Modell mitaufgenommen, da dies zu stark verzerrten Parameterschätzungen, sowie unterschätzten Standardfehlern bei kleinen Fallzahlen führt und somit die Gefahr zu fehlerhaften, statistischen Schlussfolgerungen besteht (Kleinke, 2017).

NIHSSdis, NWU2_in_Prozent, timeonsetimagingh, timeimagingAngioh erfüllten die oben genannten geforderten Bedingungen nicht. NIHSSdis, NWU2_in_prozent, timeimagingAngioh wurden aufgrund eines Anteils $> 50\%$ an fehlenden Werten aus dem MI-Modell ausgeschlossen. Eine logarithmische Variablentransformation von timeonsetimagingh führte zu einer mildereren Verzerrung von 0,07, sodass dieser Parameter log-transformiert in das Imputationsmodell aufgenommen wurde.

Bei komplexem Verteilungsmuster der fehlenden Daten wird eine „Flux-Analyse“ durchgeführt (van Buuren, 2018). Der "Influx" einer Variablen quantifiziert, wie gut ihre fehlenden Daten mit den beobachteten Daten anderer Variablen übereinstimmen. Der "Outflux" einer Variablen quantifiziert, wie gut die beobachteten Daten mit den fehlenden Daten anderer Variablen zusammenhängen. Im Allgemeinen werden höhere Outflux- und Influx-Werte bevorzugt (van Buuren, 2018). Der Fluxplot kann verwendet werden, um Variablen zu erkennen, die das Imputationsmodell beeinträchtigen. Variablen, die sich eher im oberen Bereich des Koordinatensystems befinden, sind eher vollständig und daher besser für das Imputationsmodell geeignet. Variablen, die sich im unteren linken und rechten Quadranten befinden (insbesondere

im unteren linken Quadranten) und die für die spätere Analyse kausal uninteressant sind, sollten nicht in das Imputationsmodell mit einbezogen werden (van Buuren, 2018). Die Vorauswahl geeigneter Variablen für das Imputationsmodell wurde durch Prüfung auf Kolinearität komplettiert. Kolineare Parameter wurden entfernt. Zur Prüfung der Konvergenz des imputierten Datensatzes wurden Trace-Line-Plots mittels `plot()` – Funktion aus dem „R Package MICE“ genutzt. Zur Prüfung auf Plausibilität der imputierten Werte wurden „Stripplots“ verwendet. In dieser Arbeit wurde die `stripplot()` – Function aus dem „R Package MICE“ genutzt (Buuren and Groothuis-Oudshoorn, 2011). Die einzelnen Schätzer der Regressionsmodelle wurden mittels MPR – Methode gepoolt oder durch Ermittlung des Medians und der IQR ermittelt (van de Wiel et al., 2009, Marshall et al., 2009).

1.3.5.4 Deskriptive Analyse nach Imputation

Der imputierte Datensatz wurde deskriptiv ausgewertet. Die Kohorte wurde durch ein binäres Ergebnis (gut versus schlecht) dichotomisiert: ein schlechtes Ergebnis wurde als gutes funktionelles Überleben 90 Tage nach Aufnahme ($mRs \leq 2$) definiert. Die metrischen Variablen wurden entweder durch Mittelwert und Standardabweichung oder durch Median und Interquartilsbereich (IQR) beschrieben. Für kategoriale Daten wurden absolute und relative Häufigkeiten angegeben.

1.3.5.5 Multikollinearitätsanalyse und Separationsanalyse

Auf Grund der Möglichkeit des Vorliegens von „Sparse Data Bias“ wurden die Modelle im Rahmen der Regressionsanalysen auf Multikollinearität und Separation geprüft.

1.3.5.6 Regressionsanalyse

Im Folgenden werden die drei Algorithmen, mit denen die Regressionsmodelle entwickelt wurden, genauer beschrieben.

1.3.5.6.1 Algorithmus Regressionsanalyse mit Rückwärtselimination und uniformem Shrinkage

Für die multivariable logistische Regressionsanalyse mit Rückwärtselimination wurden Alter (Age), NIHSS bei Aufnahme (NIHSSadmi), Alteplase-Intervention („IVThrombolysis“) und logarithmiert die Zeit vom Beginn bis zur Bildgebung („timeonsetimagingh“) als unabhängige Variablen zur Vorhersage eines guten funktionellen Überlebens definiert als $mRs \leq 2$, in das Modell integriert. Für die Modellentwicklung wurde ein modifizierter Algorithmus nach Heymans et. al. entwickelt (Heymans et al., 2007). Es wurden unter anderem die R-Packages „psfmi“, CalibrationCurves“ und „DescTools“ genutzt (Heymans, 2022, al., 2017, De Cock B, 2016, Van Calster et al., 2016b). Abbildung 4 zeigt schematisch einen Teil des Algorithmus der klassischen Modellentwicklung.

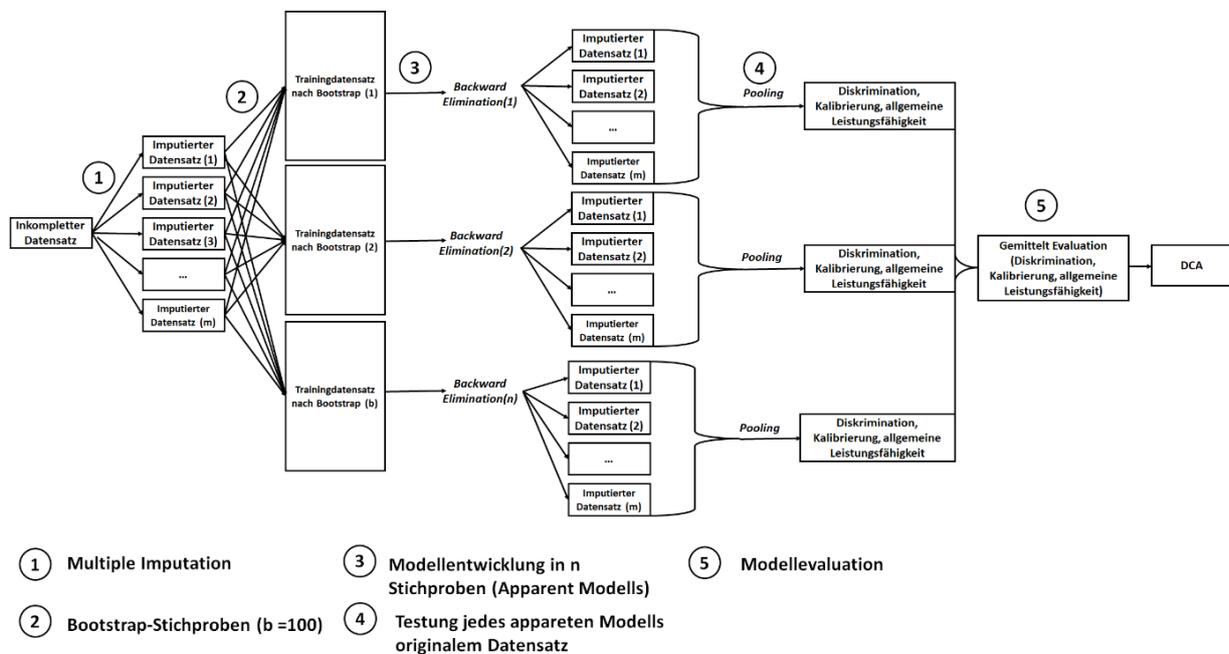


Abbildung 4: Algorithmus für Entwicklung eines logistischen Regressionsmodells, interne Validierung und Evaluation nach multipler Imputation. DCA („decision curve analysis“): Entscheidungskurvenanalyse

Nach Multipler Imputation mit 5 Iterationszyklen ($m = 5$) wurde zunächst ein logistisches Regressionsmodell mit Rückwärtselimination entwickelt (Originalmodell). Lokale, unadjustierte 2-seitige p-Werte $< 0,05$ wurden als statistisch signifikant angesehen und dienten der Variablenselektion.

Die interne Validierung des Modells erfolgte unter Berücksichtigung der geringen Fallzahl ($n = 119$) mittels ordinärem Bootstrapping (Steyerberg, 2018): Aus dem multipel imputierten Datensatz wurde eine Stichprobe gezogen, in der Art, dass aus jedem imputierten Datensatz zufällig fest definierte Fälle gezogen werden. Aus diesem Stichprobendatensatz wurden dann 100 verschiedene Bootstrap-Stichproben generiert (=Trainingsdatensätze). In jedem dieser Bootstrap-Stichproben wird ein gepooltes logistisches Regressionsmodell mit Rückwärtselimination (apparentes Modell) entwickelt. Jedes apparente Modell wird in dem originalen, multipel imputierten Datensatz getestet (Testmodell). Durch die Kombination von Rückwärtselimination und Bootstrapping wurden nach dem Zufallsprinzip mehrere Stichproben aus der beobachteten Stichprobe gezogen und so die Stichprobenvariation in der Grundgesamtheit simuliert. Anschließend wurden schrittweise Regressionsanalysen für jede Bootstrap-Stichprobe durchgeführt. Die Frequenz jedes potenziellen Prädiktors, der iterativ in das Regressionsmodell aufgenommen wurde, gibt Aufschluss über die Stärke der Evidenz, dass der Prädiktor ein echter unabhängiger Prognosefaktor ist. Variablen, die generell einen starken Einfluss auf die abhängige Variable haben, werden häufiger einbezogen als Variablen mit keinem oder einem schwachen Effekt. Das Ziel ist, dass sich das Modell, auf diese Weise dem optimalen Modell in der Population stärker annähert. In dem Regressionsmodell wurden keine nicht-linearen und keine Interaktionsterme verwendet. Zur weiteren Korrektur von Overfitting wurde das entwickelte Modell mittels uniformem Shrinkage optimiert (Pavlou et al., 2015).

1.3.5.6.2 Algorithmus Ridge-Regression

Für die Entwicklung, Validierung und Evaluation des Ridge Regressionsmodells wurde ein modifizierter Algorithmus nach Gunn et. al. entwickelt (Gunn et al., 2022). Es wurden unter anderem die R – Packages „glmnet“, „boot“, „CalibrationCurves“ und „DescTools“ verwendet (Friedman et al., 2010, Simon et al., 2011, al., 2017, De Cock B, 2016, Van Calster et al., 2016b). Abbildung 5 zeigt schematisch den Algorithmus der Modellentwicklung mittels Ridge-Regression.

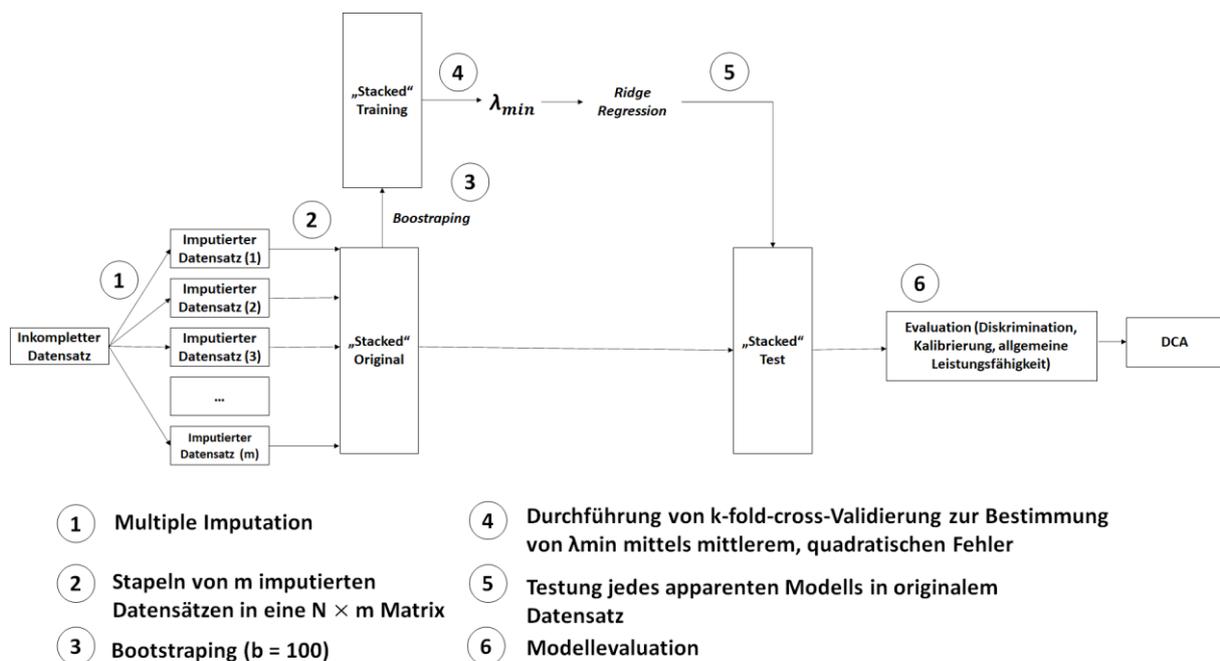


Abbildung 5: Algorithmus für die Entwicklung eines Ridge-Regressionsmodells, interne Validierung und Evaluation nach multipler Imputation. λ_{min} : optimaler Penalty-Parameter „ λ “ mit minimalem MSE; DCA („decision curve analysis“): Entscheidungskurvenanalyse

Nach multipler Imputation mit 5 Iterationszyklen ($m = 5$) wurde zunächst ein Ridge-Regressionsmodell in einem gestapelten (engl. „stacked“) Datensatz, bestehend aus mehreren imputierten Teildatensätzen mit $n \times m$ Reihen für n eingeschlossene Patienten und m imputierten Datensätzen entwickelt (Originalmodell). Die Interne Validierung des Modells erfolgte unter Berücksichtigung der geringen Fallzahl ($n = 119$) mittels ordinärem Bootstrapping (Steyerberg, 2018). Aus dem gestapelten Datensatz wurden 100 Bootstrap-Stichproben ($b=100$) gezogen (=Trainingsdatensätze). In jedem dieser Bootstrap-Stichproben wird ein Ridge-Regressionsmodell (apparentes Modell) entwickelt. Zur Bestimmung des optimalen „Penalty Parameters“ ($=\lambda_{min}$) wurde 10-fold-cross-Validation zur Bestimmung des minimalen mittleren, quadratischen Fehlers (MSE) genutzt. Jedes apparente Modell der 100 Bootstrap-Stichproben wurde in dem originalen, gestapelten, imputierten Datensatz getestet (Testmodell). In dem Regressionsmodell wurden keine nicht-linearen und keine Interaktionsterme verwendet. Auf die Berechnung von Konfidenzintervallen der Regressionskoeffizienten wurde verzichtet.

1.3.5.6.3 Algorithmus FLAC-Regression

Für die Entwicklung, Validierung und Evaluation des FLAC-Regressionsmodells wurde ebenfalls ein modifizierter Algorithmus nach Gunn et. al. entwickelt (Gunn et al., 2022). Es wurden unter anderem die R-Packages „boot“, „CalibrationCurves“, „DescTools“ und „logistf“ verwendet (Simon et al., 2011, al., 2017, De Cock B, 2016, Van Calster et al., 2016b, Heinze and Schemper, 2002, Puhr et al., 2017)

Nach Multipler Imputation mit 5 Iterationszyklen ($m = 5$) wurde zunächst ein FLAC-Regressionsmodell in einem gestapelten (engl. „stacked“) Datensatz, bestehend aus mehreren imputierten Teildatensätzen mit $n \times m$ Reihen für n eingeschlossene Patienten und m imputierten Datensätzen entwickelt (Originalmodell). Die interne Validierung des Modells erfolgte ebenfalls mittels Bootstrapping (Steyerberg, 2018). Aus dem gestapelten Datensatz wurden 100 Stichproben generiert (=Trainingsdatensätze). In jedem dieser Bootstrap-Stichproben wird ein FLIC-Regressionsmodell (apparentes Modell) entwickelt. Jedes apparente Modell der 100 Bootstrap-Stichproben wurde in dem originalen, gestapelten, imputierten Datensatz getestet (Testmodell). In dem Regressionsmodell wurden keine nicht linearen und keine Interaktionsterme verwendet. Auf die Berechnung eines Konfidenzintervalls der Regressionskoeffizienten wurde ebenfalls verzichtet.

1.3.5.7 Modellevaluation

Zunächst wurden die prognostizierten Ereigniswahrscheinlichkeiten der Modelle mittels Dichtegraph deskriptiv ausgewertet. Es wurde für die Prädiktion von $mRS \leq 2$ ROC-AUC, Sensitivität, Spezifität und korrespondierender "Cut-off" der multivariablen, logistischen Regressionsmodelle mittels Youden-Index bestimmt. Die statistische Leistungsfähigkeit jedes Modells wurde in drei Bereichen evaluiert und validiert: Diskriminationsfähigkeit, Kalibrierung und allgemeine Leistungsfähigkeit. Zur Bestimmung der Diskriminationsfähigkeit der Modelle (Originalmodell, apparentes Modell und Testmodell) wurde jeweils die Fläche unter der Kurve (engl. „area under the curve, AUC“) berechnet. Die schwache und moderate Kalibrierung der Modelle wurde mit dem R-Package „calibrationcurves“ berechnet. Zur Bestimmung der Overall Performance und des klinischen Nutzens wurde der standardisierte Brier-Score erhoben.

1.3.5.7.1 Entscheidungskurvenanalyse (engl. „DCA“)

Für die drei Modelle wurden Entscheidungskurvenanalysen durchgeführt. Es wurde das R Package „dcurves“ verwendet (Sjöberg, 2022).

1.4 Ergebnisse

Insgesamt wurden 119 Patienten eingeschlossen, von denen 28 Patienten (24%) ein gutes, funktionelles Überleben nach 90 Tagen aufwiesen.

1.4.1 Native Datenanalyse und Imputationsmodellentwicklung

Tabelle 1 zeigt die deskriptive Auswertung des unvollständigen Datensatzes für die dichotomisierte Kohorte nach mRs ≤ 2 . Einige Parameter („Zeit Symptombeginn bis erste Bildgebung“, „Zeit erste Bildgebung bis Leistenpunktion“, „lokale Thrombolysetherapie ja/nein“) weisen einen hohen Anteil an fehlenden Werten auf.

Tabelle 1: Deskriptive Auswertung des nativen dichotomisierten Datensatzes mit fehlenden Werten.

Variablen	Fehlende Werte mRS 3-6		Fehlende Werte mRS 0-2	
	mRS 3-6	n (%)	mRS 0-2	n (%)
Alter in Jahren (SD)	71,6 (12,1)	1 (1,1)	64,5 (15,6)	0 (0)
Geschlecht weiblich (%)	46 (50,5)	0 (0)	11 (39)	0 (0)
mRs bei Aufnahme, n(%)	2 (0-5)	0 (0)	0 (0-1,5)	0 (0)
NHSS bei Aufnahme, Median(IQR)	30 (14-42)	15 (16,5)	9,5 (0,5-18,5)	2 (7,1)
Glucose, Median mg/dl(IQR)	139 (115-182,5)	17 (18,7)	127 (116-147,25)	2 (7,1)
Blutdruck, Median mmHg(IQR)	155 (129,8-176,5)	43 (47,3)	146 (137,0-169,5)	6 (21,4)
Grunderkrankung Bluthochdruck, n(%)	65 (71)	4 (4)	18 (64)	5 (18)
Grunderkrankung Diabetes, n(%)	25 (27)	4 (4)	3 (11)	6 (21)
Grunderkrankung Hypercholesterinämie, n(%)	17 (19)	4 (4)	2 (7)	5 (18)
Grunderkrankung Vorhofflimmern, n(%)	34 (37)	4 (4)	7 (25)	5 (18)
Raucher- <i>nie</i> , n(%)	70 (77)	-	21 (75)	-
Raucher- <i>aktuell</i> , n(%)	15 (17)	-	2 (7)	-
Raucher- <i>ehemalig</i> , n(%)	2 (2)	-	5 (18)	-
Abciximab <i>Lv.</i> , n(%)	11 (12)	46 (51)	4 (14,3)	11 (39,3)
Einfache Plättchenhemmung, n(%)	16 (17,6)	37 (40,7)	12 (42,9)	7 (25)
Doppelte Plättchenhemmung, n(%)	4 (4)	38 (42)	0 (0)	7 (25)
Antikoagulation, n(%)	7 (7,7)	38 (42)	2 (7)	7 (25)
pcASPECTS, Median(IQR)	9 (7-9)	0 (0)	9,5 (9-10)	0 (0)
Net Water Uptake, Median(IQR)	12,1(1,5-18)	5(5,5)	2,1(0-13,9)	1(3,6)
PCCS, Median(IQR)	5 (4-7)	15 (16,5)	6 (5-6,5)	5 (2,2)
BATMAN, Median(IQR)	6 (5-8)	15 (16,5)	7(5,5-8)	5(2,2)
Jauss Score, Median(IQR)	0 (0-2)	12 (13,2)	0(0)	2 (7)
Maligner Infarkt, n(%)	23 (25)	2 (2)	0(0)	0 (0)
intravenöse Thrombolysie, n(%)	48 (52,7)	3 (3,3)	16 (57)	0 (0)
Zeit Symptombeginn bis erste Bildgebung in Std., Median(IQR)	3,6 (1,7-6,1)	51(56)	2,8 (1,2-4,9)	8 (28,6)
Zeit erste Bildgebung bis Leistenpunktion in Std., Median(IQR)	0,5 (0,3-1,6)	48(52,7)	1,1 (0,3-1,6)	12 (42,9)
langstreckiger Verschluss, n(%)	53 (58)	4(4)	11 (39,3)	5 (18)
Beteiligung der A. basilaris, n(%)	79 (86,8)	4(4)	23 (82)	5 (18)
Thrombuslokalisierung -andere, n(%)	19 (20,9)	-	3 (10,7)	-
Thrombuslokalisierung - B1 und proximal, n(%)	2 (2,2)	-	0 (0)	-
Thrombuslokalisierung - B2/B3, n(%)	4 (4,4)	-	4 (14,3)	-
Thrombuslokalisierung -B3 und distal, n(%)	12 (13,2)	-	8 (28,6)	-
Thrombuslokalisierung gemischt, n(%)	50 (54,9)	-	8 (28,6)	-
Ätiologie - kardioembolisch, n(%)	31 (34,1)	-	13 (46,4)	-
Ätiologie - artherosklerotisch, n(%)	25 (27,5)	-	7 (25)	-
Ätiologie - Dissektion, n(%)	1 (1,1)	-	0 (0)	-
Ätiologie - andere, n(%)	4 (4,4)	-	0 (0)	-
Ätiologie - kardioembolisch artherosklerotisch gemischt, n(%)	21 (23,1)	-	3 (10,7)	-
Lokale Thrombolysetherapie, n(%)	20 (22)	46 (50,5)	4 (14,3)	11 (39,3)
Rekanalisationsversuche, n(%)	1 (1-2,5)	48 (52,7)	1 (1-1)	11 (39,3)
TICI, Median(IQR)	3 (2-4)	6 (7)	4 (3-4)	2 (7)
erfolgreiche Rekanalisation, n(%)	58 (64)	6 (7)	26 (93)	2 (7)
NHSS nach 24h, Median(IQR)	42 (24,25-42)	54 (59,3)	3 (0,5-4)	7 (0,25)
Net Water Uptake nach 24h, Median(IQR)	11 (6-16,5)	63 (69,2)	10,7 (3,7-15,6)	22 (78,6)
NET Water Uptake im Verlauf, Median(IQR)	24,2 (17,5-31,9)	17 (18,7)	18 (4-27,4)	7 (0,25)
postinterventioneller Aufenthalt auf IST in Tagen, Median(IQR)	6 (2-10)	46 (50,5)	1 (1-2)	11 (39,3)
NHSS bei Entlassung, Median(%)	20 (10-27)	54 (59,3)	1 (0-3)	7 (0,25)
mRa90, Median(IQR)	5 (4-6)	0 (0)	10(-1)	0 (0)

1.4.1.1 Md.-Pattern-Graph-Diagnostik

Die Analyse des „md.pattern“- Graphs zeigt ein multivariates, nicht-monotones, allgemeines Verteilungsmuster der fehlenden Datenwerte mit 25 spezifischen Verteilungsmustern (Abbildung 6). Aufgrund des komplexen Verteilungsmusters wurde eine Influx – und Outflux-Diagnostik durchgeführt, um die Variablenselektion für das Imputationsmodell zu optimieren (siehe Abschnitt 1.4.2.1. Flux-Analyse).

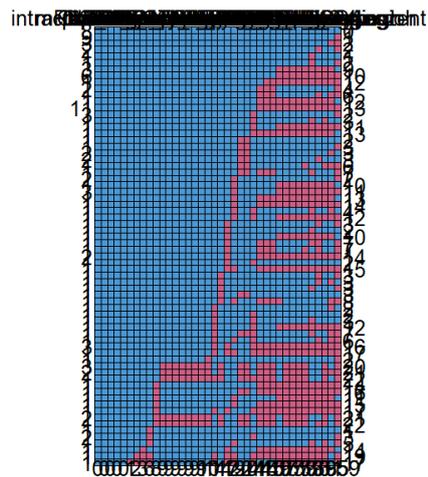


Abbildung 6: md-Pattern- Graph zeigt ein komplexes Verteilungsmuster der fehlenden Daten. Eine Matrix mit $ncol(x)+1$ Spalten, in der jede Zeile einem fehlenden Datenmuster entspricht (blau=beobachtet, rot=fehlend). Die Zeilen und Spalten sind nach der Anzahl der fehlenden Informationen sortiert.

1.4.1.2 Ergebnis der Auswertung der Verteilungseigenschaften der ordinalen und metrischen Variablen

Die Verteilungseigenschaften der potenziellen MI-Modellvariablen wurden grafisch ausgewertet, um bei Nicht-Normalverteilung die Imputationsalgorithmen dementsprechend anzupassen. Im Folgenden sind Histogramm, Dichtediagramm und QQ-Plots für die metrischen Prädiktoren dargestellt (Abbildung 7-9)

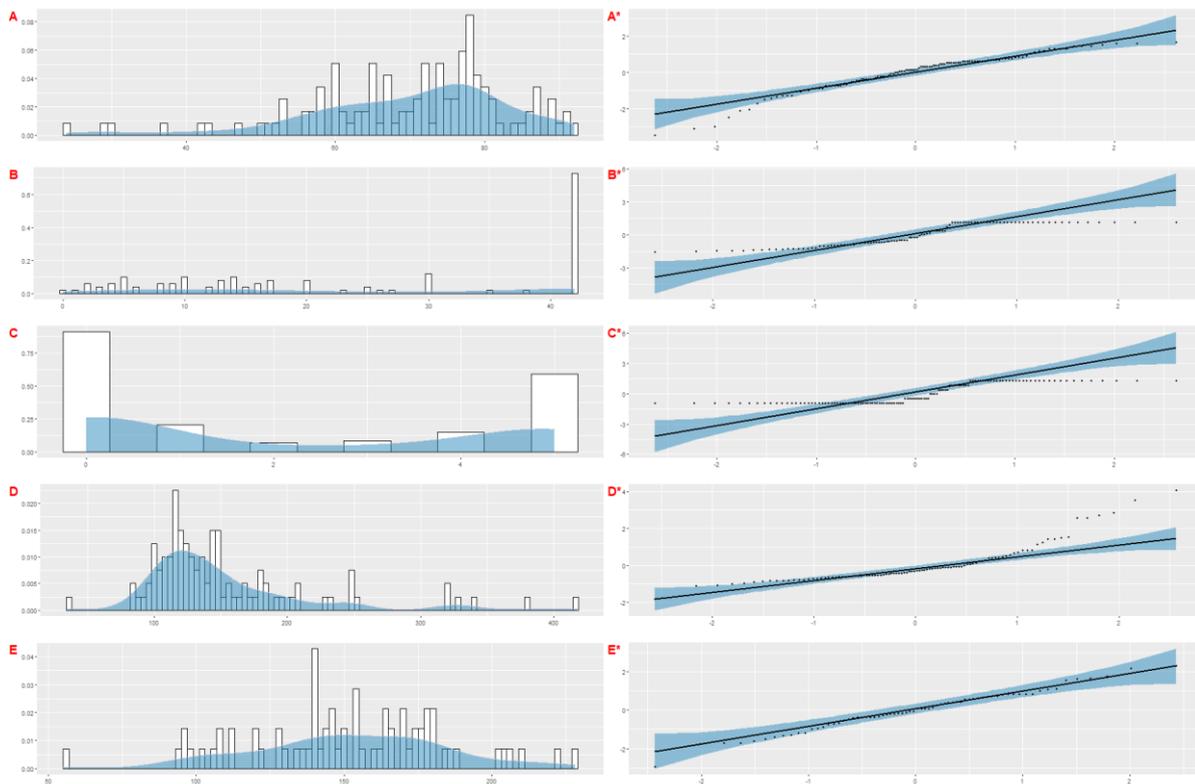


Abbildung 7: Kombiniertes Histogramm und Dichtediagramm, sowie QQ-Plots*1 der metrischen Variablen des MI-Modells. A/A*: Alter bei Aufnahme („Age“), B/B*: NIHSS bei Aufnahme („NIHSSadmi“), C/C*: mRS vor Schlaganfall („mRspreStroke“); D/D*: Blutglukosekonzentration bei Aufnahme („Glucoseinitial“); E/E*: Blutdruck bei Aufnahme („RRinitial“)

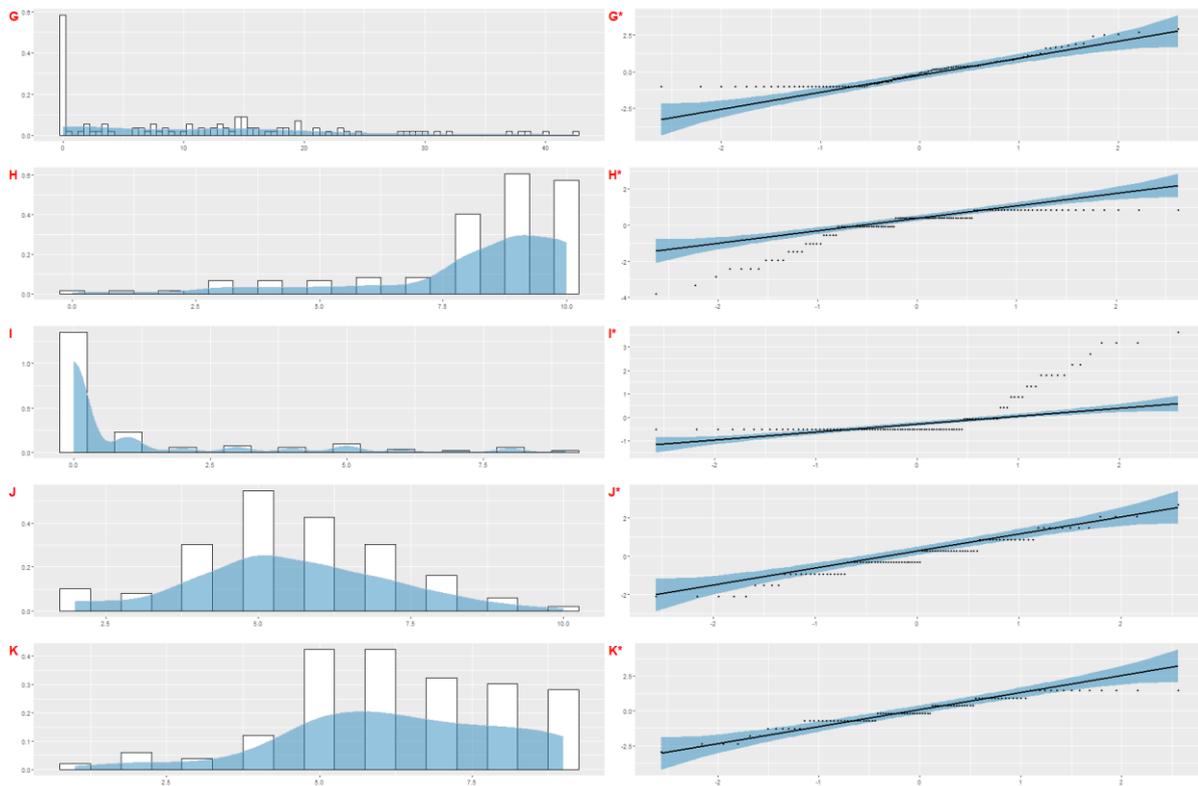


Abbildung 8: Kombiniertes Histogramm und Dichtediagramm, sowie QQ-Plots* weiterer, metrischer Variablen des MI-Modells: G/G*: Net Water Uptake in der Aufnahme-CT („NWU_in_prozent“); H/H*: pcASPECTS in der Aufnahme-CT (pcASPECTS); I/I*: Jauss-Score in der Aufnahme-CT („cerebellaredemascore“); J/J*: PC-CS in der Aufnahme-CT („PCCS“), K/K*: BATMAN in der Aufnahme-CT („BATMAN“)

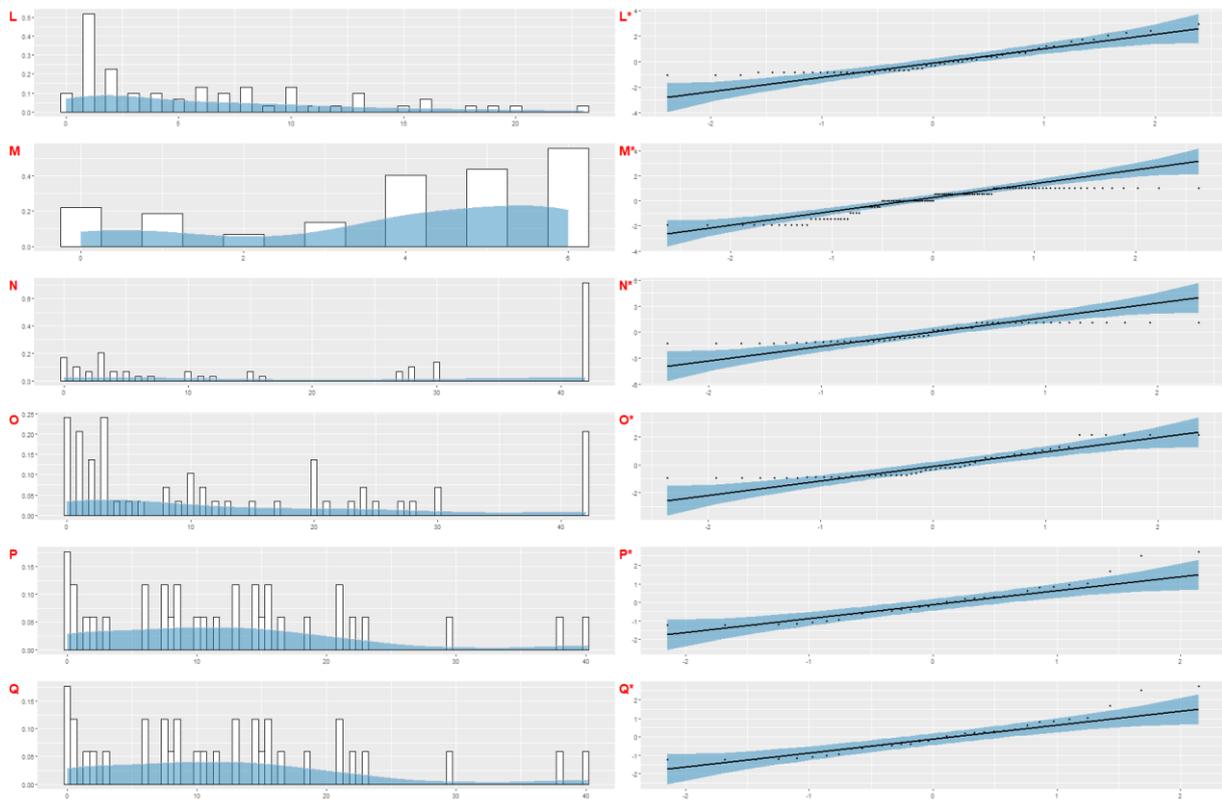


Abbildung 9: Kombiniertes Histogramm und Dichtediagramm, sowie QQ-Plots+ weiterer, metrischer Variablen des MI-Modells. L/L*: Liegezeit auf Intensivstation („ICUdays“); **M/M*:** mRS nach 90 Tagen („mRS90“); **N/N*:** NIHSS nach 24 Stunden nach Schlaganfall („NIHSS24h“); **O/O*:** NIHSS bei Entlassung („NIHSSdis“); **P/P*:** Net Water Uptake in der ersten Verlaufs-CT („NWU2_in_prozent“); **Q/Q*:** Net Water Uptake in der letzten Verlaufs-CT („NWU_FU_in_prozent“)

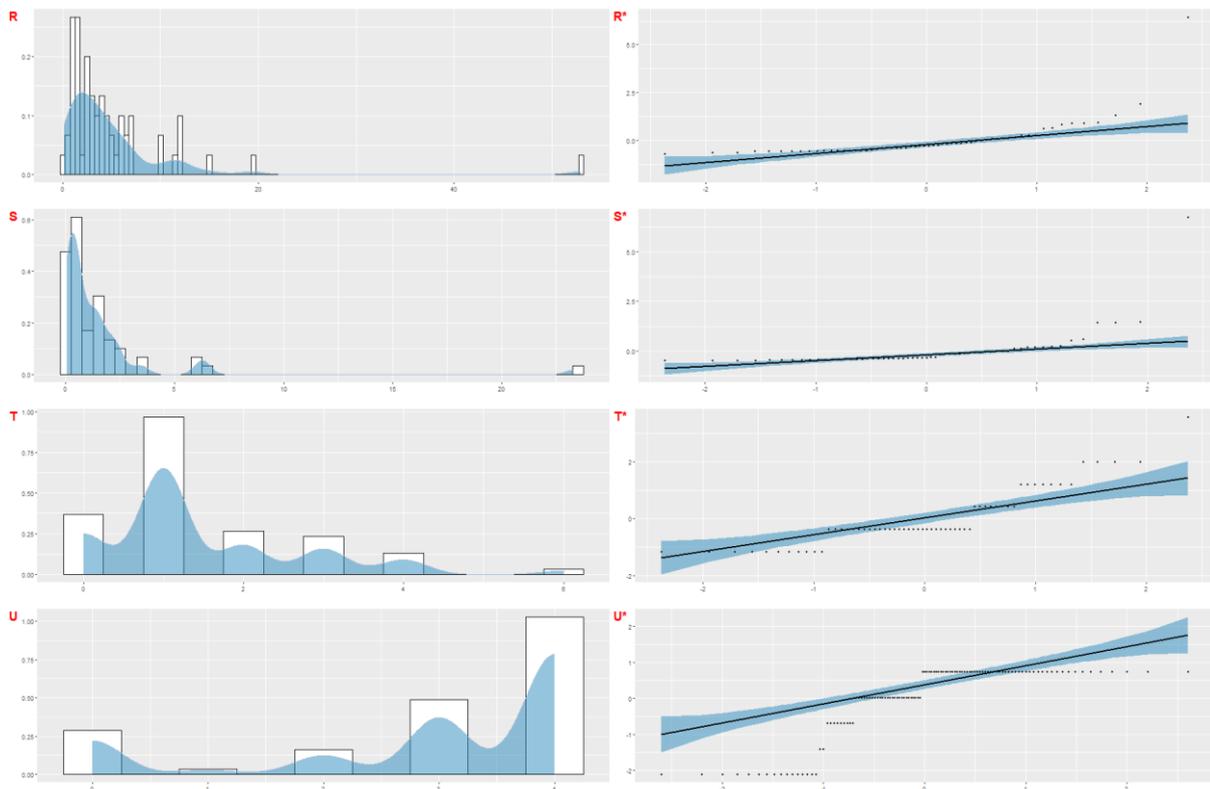


Abbildung 10: Kombiniertes Histogramm und Dichtediagramm sowie QQ-Plots* weiterer, metrischer Variablen des MI-Modells: R/R*: t Zeit Symptombeginn bis erste Bildgebung („timeonsetimagingh“); S/S*: Zeit erste Bildgebung bis Leistenpunkt („timeimagingAngioh“), T/T*: Rekanalisationsversuche („No_of_passages“), U/U*: TICI nach Rekanalisation („TICI_cod“)

Es zeigt sich, dass die meisten Variablen nicht normalverteilt sind. Die Variablen „NIHSSdiss“ und „timeimagingAngioh“ wurden als potenzielle Hilfsvariablen beziehungsweise Prädiktorvariablen aus dem Imputationsmodell ausgeschlossen, da der Anteil an fehlenden Werten > 50% beträgt und somit die Kriterien nach Kleinke, 2017 nicht eingehalten werden können: 1.) absolute Schiefe der Variablenverteilung (engl. „skewness“) < 1,6 2.) absolute Schiefe der Variablenverteilung speziell <0,81 bei einem Anteil fehlender Werte von 30% - 50% mit Fallzahlen im Bereich n = 100. 2.) Fehlende Variablenwerte <50% (Kleinke, 2017). Die Variable „timeonsetimagingh“ zeigte ebenfalls einen hohen Anteil an fehlenden Werten (>30%) und eine Schiefe der Verteilungsfunktion von 4,74. Durch Logarithmieren des Parameters konnte die Schiefe auf <0,81 korrigiert werden, weshalb die logarithmierte Form in die weitere Auswertung integriert wurde.

1.4.1.3 Ergebnis der „Flux-Analyse“

Aufgrund des komplexen Verteilungsmusters, welches im md-Pattern-Graph (siehe Abschnitt 1.4.1.1) detektiert wurde, wurde eine Flux-Analyse durchgeführt, um eine optimale Variablenauswahl für das Imputationsmodell zu gewährleisten. Der Fluxplot kann verwendet werden, um Variablen zu identifizieren, die das Imputationsmodell überladen und keinen weiteren Informationsgewinn liefern. Variablen, die sich in dem unteren Quadranten befinden (insbesondere im linken unteren Quadranten) und für die spätere Analyse uninteressant sind, werden vor der Imputation aus dem Datensatz entfernt. Abbildung 11A zeigt die Fluxanalyse des nativen Datensatzes, in die alle potenziellen Kandidatenprädiktoren und Hilfsvariablen eingeschlossen wurden. Es lassen sich drei Cluster identifizieren. Cluster drei beinhaltet Variablen („No_of_passages“, „Abciximab_use“, „LokaleLyse“, „ICUdays“) mit einem niedrigen Influx $< 0,5$ und einem sehr niedrigen Outflux $< 0,25$. Die Variablen wurden aus dem Datensatz entfernt und die Fluxanalyse wurde wiederholt. Abbildung 11b zeigt einen Informationsverlust des Parameters „bleedmeds“, in dem der Status der medikamentösen Blutgerinnungstherapie (einfache, doppelte Plättchenhemmung oder Antikoagulation) bei Aufnahme zusammengefasst wurde. Nach Entfernung dieses Parameters zeigt sich in einem dritten Zyklus (nicht gezeigt) der Parameter „NIHSS24h“ mit dem geringsten Outflux-Potential. Eine weitere exploratorische Fluxanalyse ohne NIHSS24h führt zu einem Anstieg des Flux-Potentials der Parameter „RRinitial“ und „logtimeonsetimagingh“ (Abbildung 11C). Eine weitere Reduktion der Imputationsmodellvariablen hätte einen massiven Verlust des Outflux-Potentials mehrerer Variablen zufolge (Abbildung 11d), weshalb der Parameter „logtimeonsetimagingh“ im Modell belassen wurde.

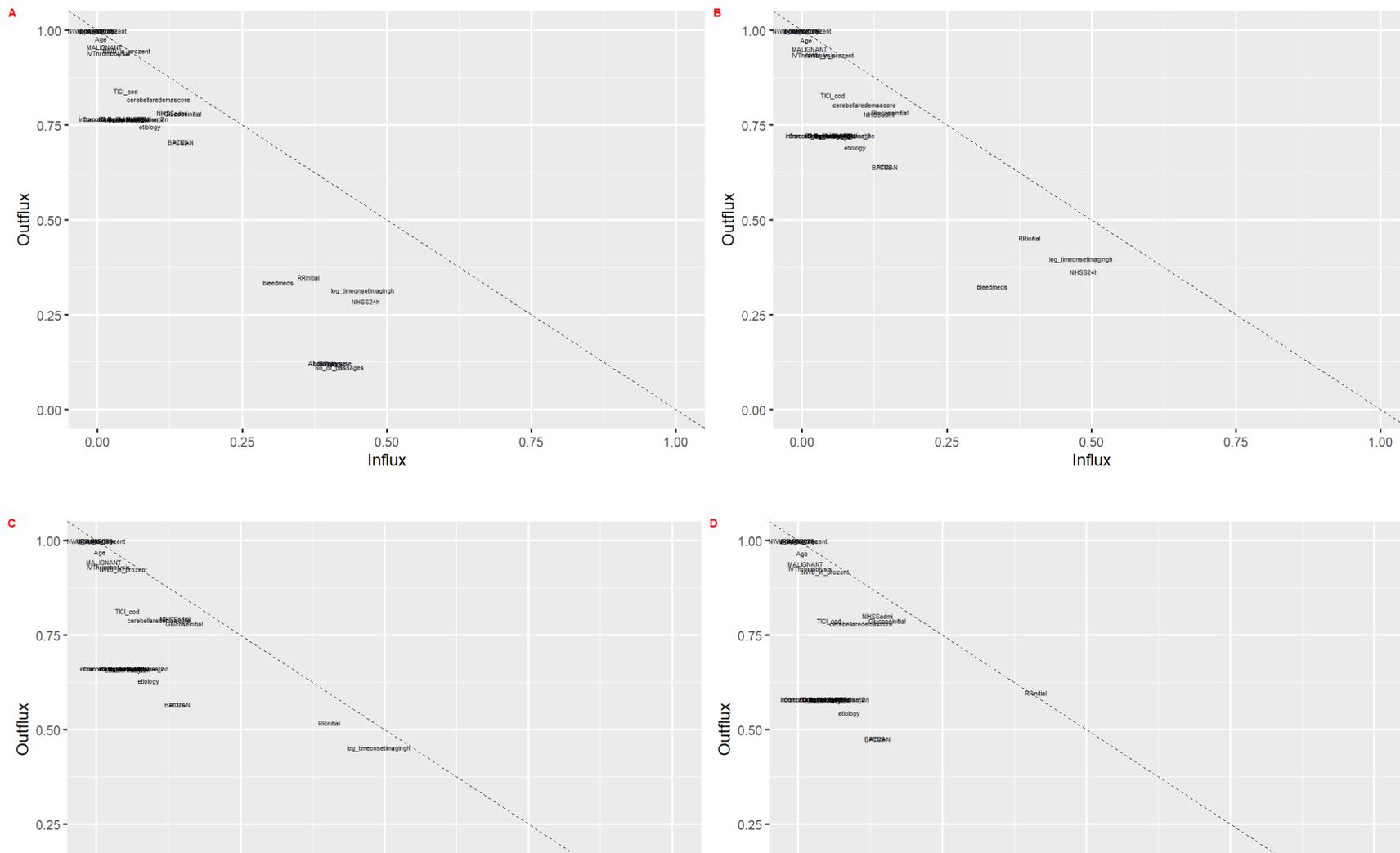


Abbildung 11: Flux-Graphen A-D. Der "Influx" einer Variablen quantifiziert, wie gut ihre fehlenden Daten mit den beobachteten Daten anderer Variablen übereinstimmen. Der "Outflux" einer Variablen quantifiziert, wie gut die beobachteten Daten mit den fehlenden Daten anderer Variablen zusammenhängen. Im Allgemeinen werden höhere Outflux- und Influx-Werte bevorzugt. Für eine detaillierte Erläuterung siehe Text.

1.4.1.4 Finale Konfiguration des Imputationsmodells

Folgende Variablen wurden in das Imputationsmodell eingeschlossen: „Age“, „Sex“, „mRs90“, „cerebellaredemascore“, „Craniectomy“, „NIHSSadmi“, „MALIGNANT“, „pcASPECTS“, „NWU_in_procent“, „NWU_FU_in_procent“, „PCCS“, „BATMAN“, „intracran_thrombus_lokalisation“, „longOcclusion“, „BAO_involvement“, „IVThrombolysis“, „TICI_cod“, „ComorbidityHTN“, „ComorbidityDM“, „ComorbidityHLP“, „ComorbidityVHF“, „Comorbidity_smoking_war_2“, „etiology“, „Glucoseinitial“, „RRinitial“, „log_timeonsetimagingh“, „mRspreStroke“. Die Prädiktormatrix des Imputationsalgorithmus wurde nicht verändert. Auf eine passive Imputation abgeleiteter Variablen (z.B. „MALIGNANT“) wurde verzichtet, da sowohl Werte der Variablen „cerebellaredemascore“ und Werte der Variablen „MALIGNANT“ unabhängig voneinander fehlten. Die Methodenmatrix des MICE-Algorithmus wurde angepasst: Für die Imputation binär skalierten Parameter wurden binäre, logistische Regressionsmodelle verwendet. Geordnete kategoriale Parameter mit höherer Kategorienzahl wurden ebenfalls mittels „Midastouch“ imputiert, da das Verhältnis zwischen der Anzahl der zu interpolierenden Parameter und der Anzahl der Ereignisse klein ist. Für die Imputation ungeordneter Variablen geringer Kategorienzahl wurden multinomiale Modelle verwendet.

1.4.1.5 Ergebnis der „Trace Line– und Stripplots“

Die „Trace-Line-Plots“ für die imputierten metrischen Variablen zeigen eine gute Vermischung der Trendlinien (Abbildung 12-18) und somit eine gute Konvergenz der imputierten Daten.

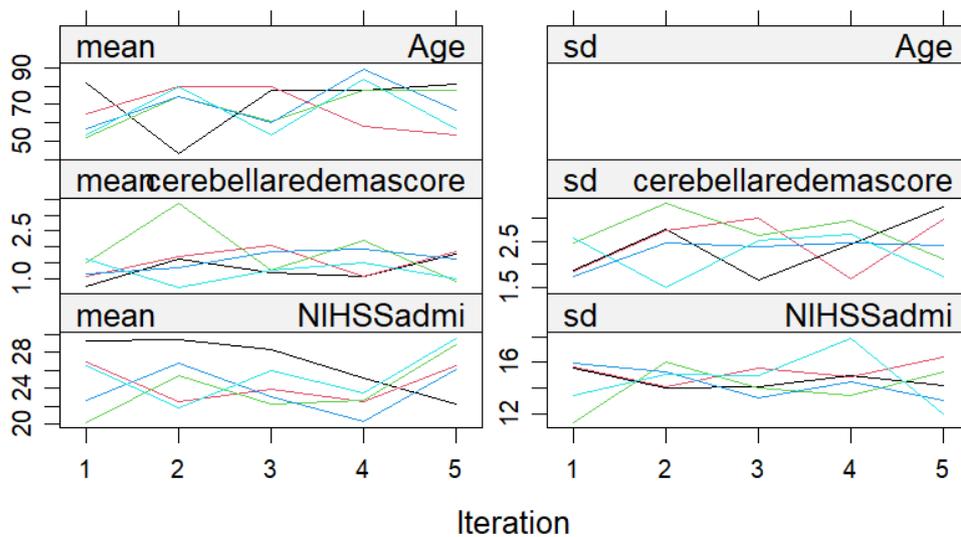


Abbildung 12: Trace-Line-Plots 1 für die metrischen Variablen Alter ("Age"), JAUSS ("cerebellaredemascore") und NIHSS bei Aufnahme ("NIHSSadmi")

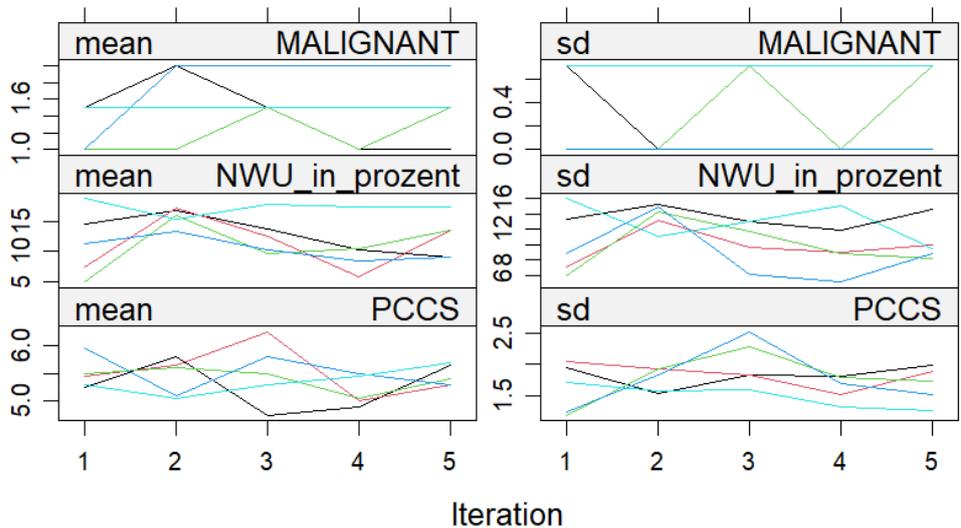


Abbildung 13: Trace-Line-Plots 2 für die binäre Variable Maligner Infarkt in der Aufnahme-CT ("MALIGNANT"), Net Water Uptake in der Aufnahme-CT ("NWU_in_prozent") und PC-CS in der Aufnahme-CT ("PCCS")

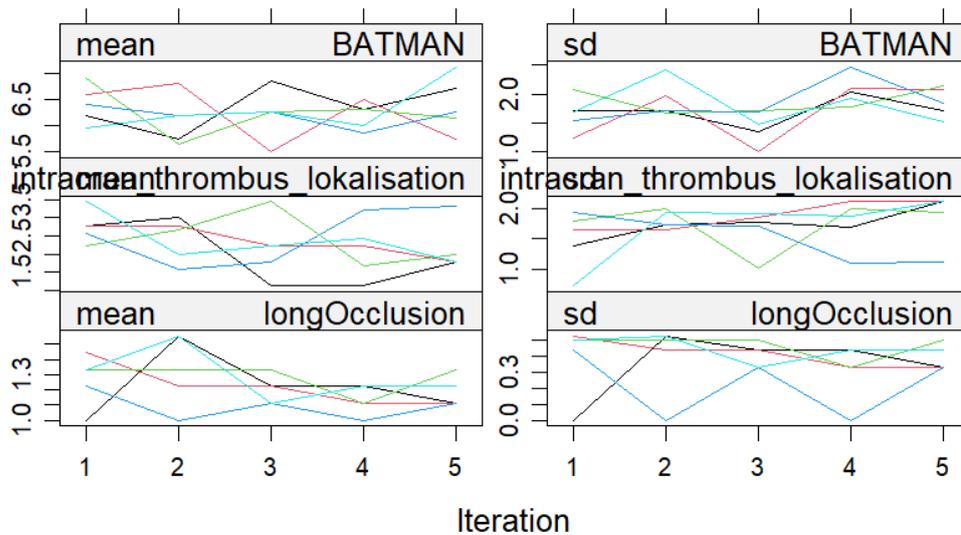


Abbildung 14: Trace-Line-Plots 3 für die Variablen BATMAN in der Aufnahme-CT („BATMAN“), Thrombuslokalisation ("intracran_thrombus_lokalisation") und die binäre Variable Thrombuslänge ("longOcclusion") in der Digitalensubtraktionsangiografie

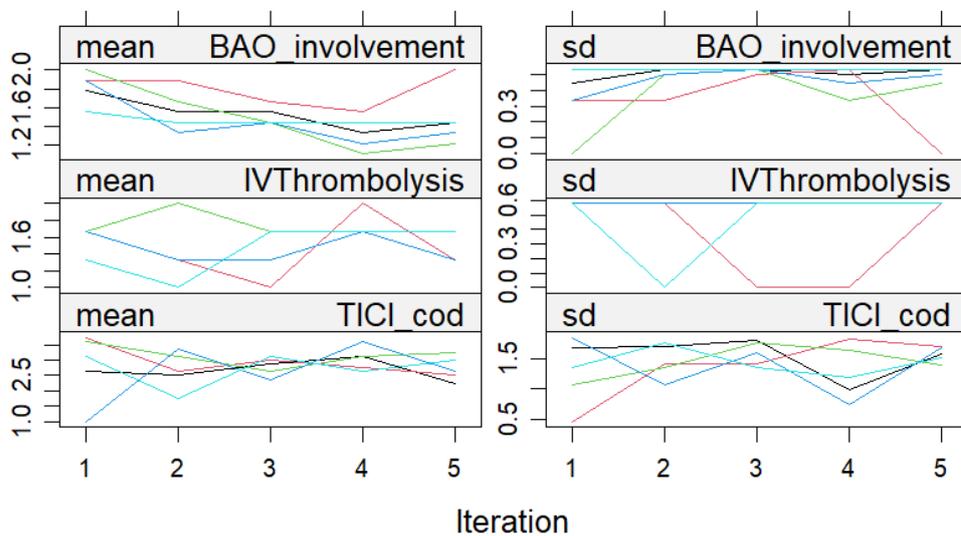


Abbildung 15: Trace-Line-Plots 4 für die binären Variablen „Verschluss/Teilverschluss der A. basilaris“ ("BAO_involvement"), und die kategoriale Variable TICI-Score nach Rekanalisation ("TICI_cod")

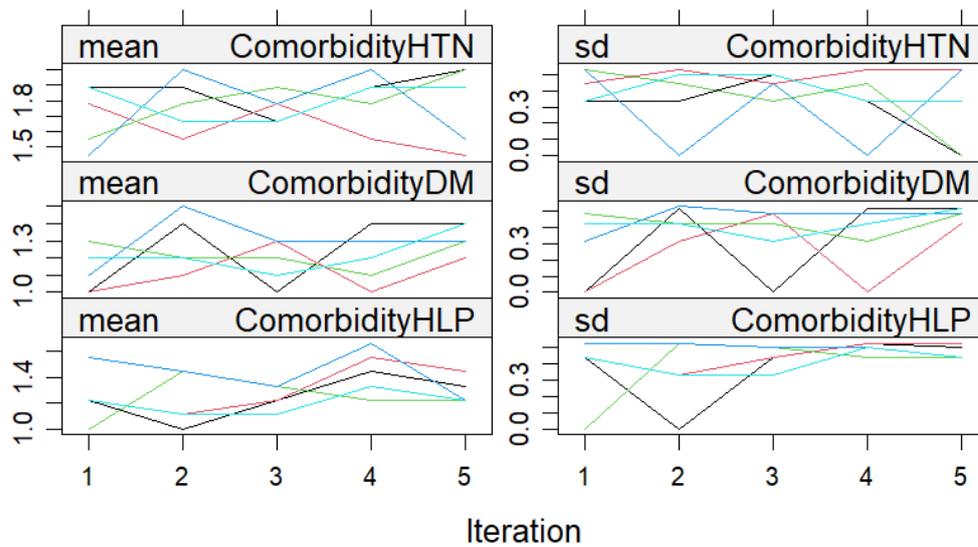


Abbildung 16: Trace-Line-Plots 5 für die binären Variablen Vorerkrankung "arterielle Hypertonie", Vorerkrankung "Diabetes mellitus II", Vorerkrankung "Hyperlipoproteinämie"

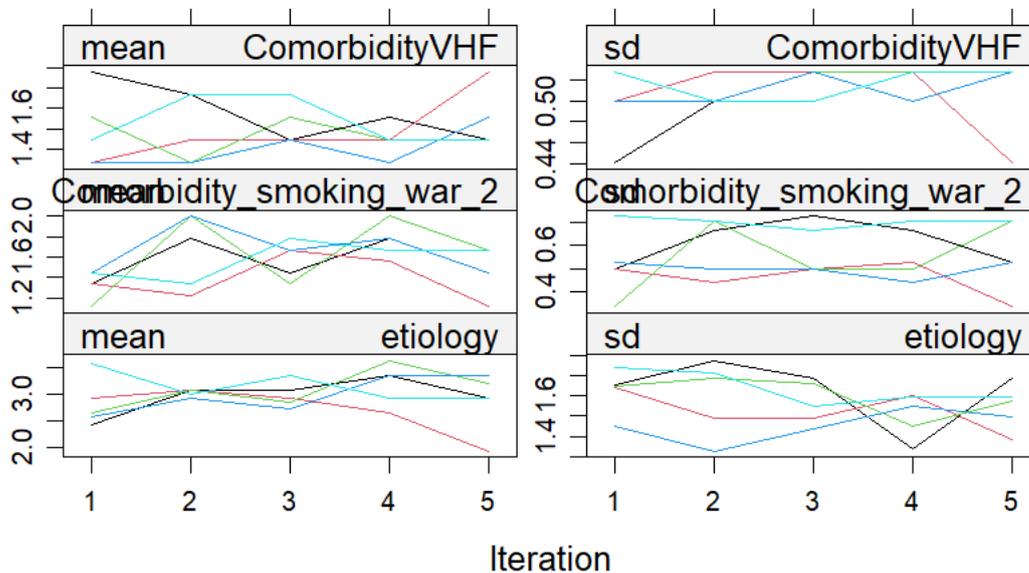


Abbildung 17: Trace-Line-Plots 6 für die binäre Variable Vorerkrankung "Vorhofflimmern", "Nikotinanamnese" ("Comorbidity_smoking_war_2") und Schlaganfallätiologie ("etiology")

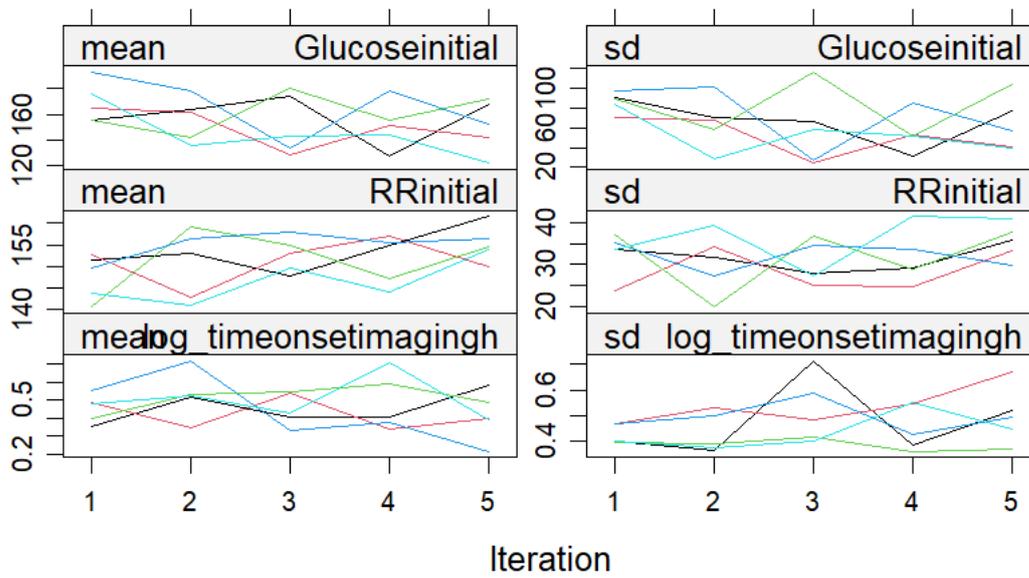


Abbildung 18: Trace-Line-Plots 7 für die Variablen Blutglukosekonzentration bei Aufnahme ("Glucoseinitial"), Blutdruck bei Aufnahme ("RRinitial"), Zeit Symptombeginn bis erste Bildgebung ("timeonsetimagingh")

Die Stripplots zeigen plausibel imputierte Variablenwerte und Bestätigen eine gute Qualität des Imputationsmodells. Die imputierten Datenwerte sind in „rot“ dargestellt, die beobachteten in „blau“:

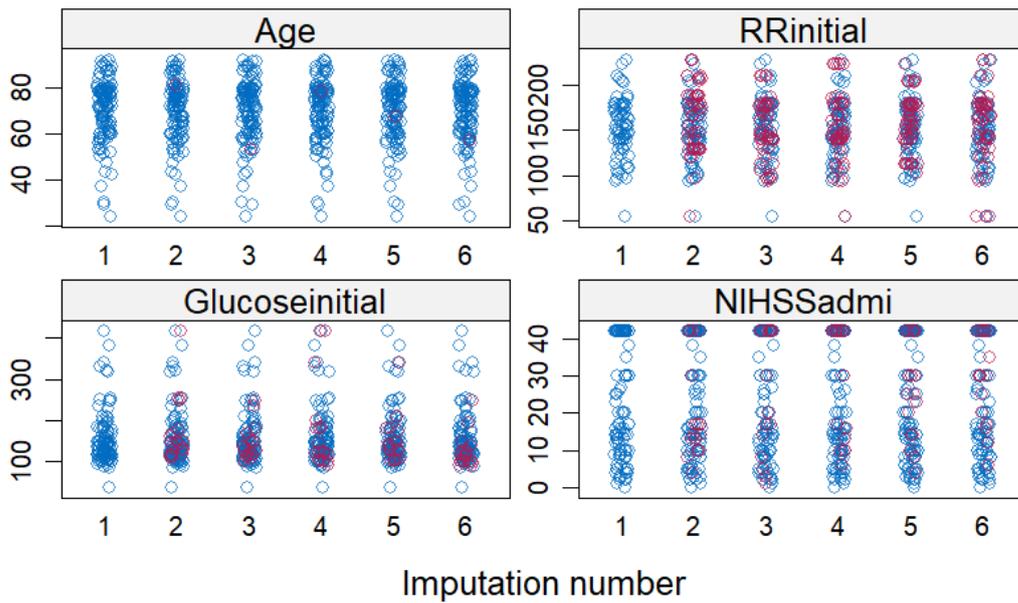


Abbildung 19: Stripplots 1 für die metrischen Variablen Alter ("Age"), Blutglukosekonzentration bei Aufnahme ("Glucoseinitial"), Blutdruck bei Aufnahme ("RRinitial") und die Variable NIHSS bei Aufnahme ("NIHSSadmi")

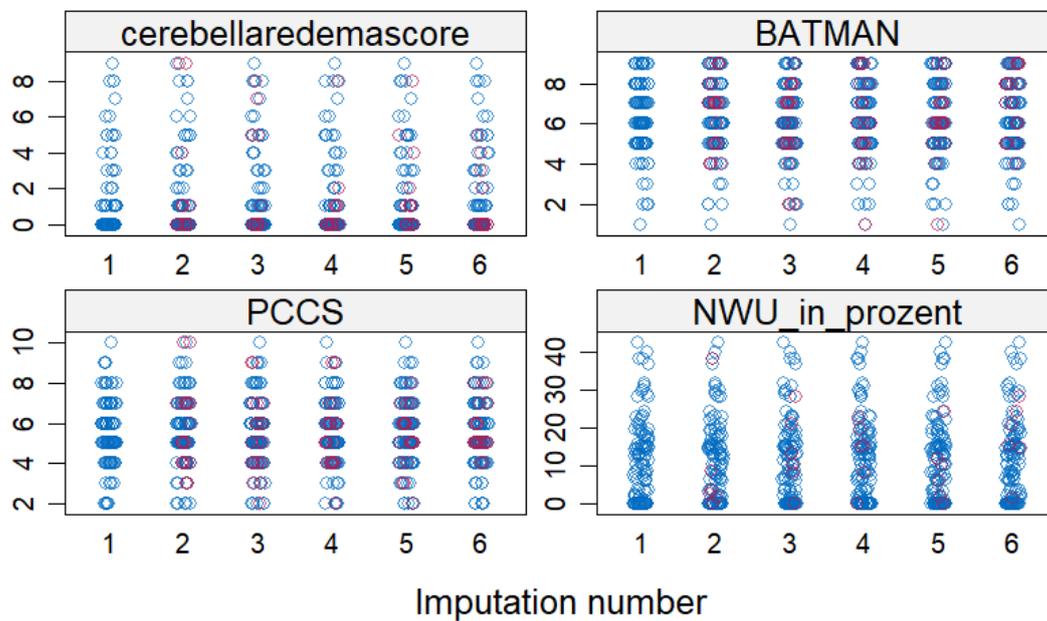


Abbildung 20: Stripplots 2 für die Variablen JAUSS ("cerebellaredemascore"), BATMAN in der Aufnahme-CT („BATMAN“), Net Water Uptake in der Aufnahme-CT ("NWU_in_prozent") und PC-CS in der Aufnahme-CT ("PCCS")

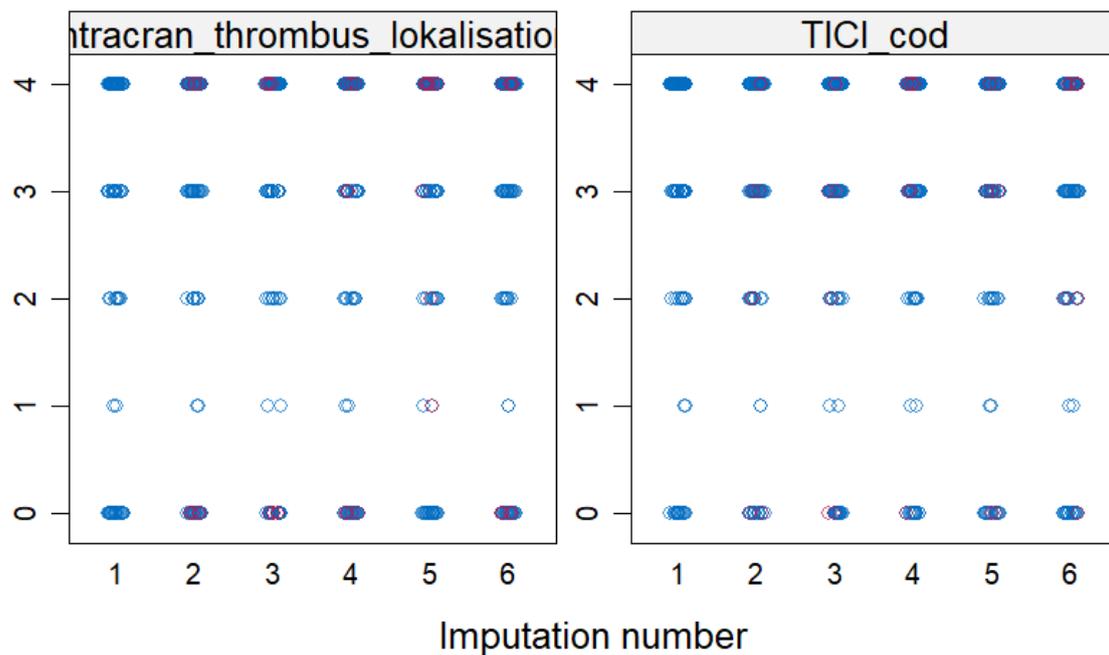


Abbildung 21: Stripplots 3 für die kategoriale Variable Thrombuslokalisation ("intracran_thrombus_localisation") und die kategoriale Variable TICI-Score nach Rekanalisation ("TICI_cod")

1.4.2 Ergebnisse der deskriptiven Analyse des imputierten Datensatzes und Prognosemodellentwicklung

Tabelle 2 zeigt die deskriptive Auswertung des imputierten Datensatzes für die dichotomisierte Kohorte nach $mRs \leq 2$. Die imputierten Werte zeigen eine gute Übereinstimmung mit den nativen Datenwerten.

Tabelle 2: Deskriptive Auswertung des imputierten dichotomisierten Datensatzes

Variablen	mRS (3–6)	mRS (0-2)
Alter in Jahren (SD)	72 (12)	65 (16)
Geschlecht weibl,n (%)	46 (51)	11 (30)
mRs bei Aufnahme, Median(IQR)	2 (0-5)	0 (0-1,5)
NIHSS bei Aufnahme, Median(IQR)	29 (14-42)	11 (4-19)
Glucose, Median mg/dl(IQR)	138 (115-182)	127 (115-146)
Blutdruck, Median mmHg(IQR)	158 (136-178)	148 (138-172)
Grunderkrankung Bluthochdruck, n(%)	68 (75)	21 (75)
Grunderkrankung Diabetes, n(%)	26 (28)	5 (19)
Grunderkrankung Hypercholesterinämie, n(%)	18 (20)	4 (14)
Grunderkrankung Vorhofflimmern, n(%)	36 (40)	10 (34)
Raucher-nie, n(%)	72,2 (79,3)	22 (81,4)
Raucher-akutell, n(%)	16,4 (18)	4,6 (16,4)
Raucher-ehmalig, n(%)	2,4 (2,6)	0,6 (2,1)
pcASPECTS, Median(IQR)	9 (7-9)	10 (9-10)
Net Water Uptake, Median(IQR)	12 (1-18)	2 (0-14)
PCCS, Median(IQR)	5 (4-7)	6 (5-6)
BATMAN, Median(IQR)	6 (5-8)	7 (5-8)
Jauss Score, n(%)	2 (0-2)	0 (0-0)
maligner Infarkt, n(%)	24 (26)	0 (0)
pcBATASs, Median(IQR)	15 (13-16)	16 (15-17)
intravenöse Thrombolyse, Median(IQR)	50 (55)	16 (57)
Zeit Symptombegin bis erste Bildgebung, Std., Median(IQR)	3,6 (1,8-6)	3,1 (1,2-5,8)
langstreckiger Verschluss, n(%)	55 (60)	55 (45)
Beteiligung der A. basilaris, n(%)	81 (90)	25 (90)
Thrombuslokalisierung -andere, n(%)	20 (22)	4 (14,3)
Thrombuslokalisierung - B1 und proximal, n(%)	2 (2,2)	0 (0)
Thrombuslokalisierung - B2/B3, n(%)	4 (4,4)	4,4 (15,7)
Thrombuslokalisierung -B3 und distal, n(%)	13 (14,3)	9 (32,1)
Thrombuslokalisierung gemischt, n(%)	52 (57,1)	10,6 (37,1)
Ätiologie - kardioembolisch, n(%)	32,6 (35,8)	14,4 (51,4)
Ätiologie - artherosklerotisch, n(%)	27,4 (30,1)	8 (28,6)
Ätiologie - Dissektion, n(%)	1,8 (2)	0,6 (2,1)
Ätiologie - andere, n(%)	22,8 (25,1)	3 (10,7)
Ätiologie - kardioembolisch artherosklerotisch gemischt, n(%)	6,4 (7)	2 (7,1)
TICI 0, n(%)	17 (18,7)	0,4 (1,4)
TICI 1, n(%)	2,2 (2,4)	0 (0)
TICI 2a, n(%)	10 (11)	0,4 (1,4)
TICI 2b, n(%)	20,6 (22,6)	8,2 (29,3)
TICI 3, n(%)	41,2 (45,3)	19 (67,9)
erfolgreiche Rekanalisation, n(%)	62 (68)	27 (97)
NET Water Uptake im Verlauf,Median(IQR)	24,5 (17,9-32,5)	20,1 (4,6-27,5)
mRs90, Median(IQR)	5 (4-6)	1 (0-1)

1.4.3 Ergebnisse Multikollinearitätsanalyse und Separationsanalyse

Es lag keine Multikollinearität und keine Separation vor.

1.4.4 Ergebnisse: Modellparameter Regressionsmodelle

Tabelle 3: Binäre logistische Regressionsanalyse für mRS 0-2

Variablen	Beta-Koeffizienten	OR	95%Konfidenzintervall (KI)	p-Values
(Intercept)	-0,875	0,417	0,005-34,328	0,694
Age	-0,042	0,959	0,923-0,996	0,031
NIHSSadmi	-0,062	0,940	0,898-0,984	0,010
pcBATASs	0,256	1,292	1,028-1,624	0,028

Binäre logistische Regressionsanalyse für gutes funktionelles Überleben (mRS 0–2). „Age“ (Alter), „NIHSSadmi“ (National Institutes of Health Stroke Scale bei Aufnahme), „pcBATASs“ (kombinierter Score von BATMAN und pASCO)

In der multivariablen, logistischen Regressionsanalyse mittels Rückwärtselimination zeigt sich, dass pcBATASs (OR: 1,292, 95%KI: 1,028–1,624, **p = 0.028**), Alter (OR: 0,959, 95%CI: 0,923–0,996, **p = 0,031**) und NIHSSadmi (OR: 0,940, 95%CI: 0,898–0,984, **p = 0,010**) signifikant und unabhängig mit einem guten, funktionalen Überleben assoziiert sind (Tabelle 3).

Tabelle 4: Optimierte Beta-Koeffizienten des logistischen Regressionsmodells

Variablen	Optimierte Beta-Koeffizienten
(Intercept)	-0,883
Age	-0,036
NIHSSadmi	-0,052
pcBATASs	0,217

Optimierte Regressionskoeffizienten: „Age“ (Alter), „NIHSSadmi“ (National Institutes of Health Stroke Scale bei Aufnahme), „pcBATASs“ (kombinierter Score von BATMAN und pASCO)

In Tabelle 4 sind die optimierten Regressionskoeffizienten des „Backward -Modells“ nach uniformem Shrinkage gezeigt. Die Beta-Koeffizienten wurden gegen den Wert 0 korrigiert.

Tabelle 5: Ridge-Regressionsanalyse für mRS 0–2

Variablen	Beta-Koeffizienten	OR
(Intercept)	0,581	1,788
Age	-0,007	0,993
NIHSSadmi	-0,009	0,990
IVThrombolysis	0,024	1,019
log_timeonsetimaging	-0,104	0,901
pcBATASs	0,026	1,027

Ridge-Regressionsanalyse für gutes funktionelles Überleben. „Age“ (Alter); „NIHSSadmi“ (National Institutes of Health Stroke Scale bei Aufnahme), log_timeonsetimaging („Zeit Symptombeginn bis erste Bildgebung“); „pcBATASs“ (kombinierter Score von BATMAN und pASCO). Auf die Berechnung von Konfidenzintervallen wurde bei Ridge- und FLAC-Regression verzichtet.

In Tabelle 5 sind die Modellparameter des Ridge-Regressionsmodells dargestellt. Das Ridge-Regressionsmodell zeigt sehr kleine Beta-Koeffizienten der Modellparameter verglichen mit den Regressionskoeffizienten der anderen Modelle. Darüber hinaus zeigt das Intercept des Ridge-Modells ein positives Vorzeichen im Vergleich zu den Modellen des klassischen logistischen Regressionsmodell und dem FLAC-Regressionsmodell.

Tabelle 6: FLAC-Regressionsanalyse für mRS 0–2

Variablen	Beta-Koeffizient	OR
(Intercept)	-0,157	0,854
Age	-0,042	0,958
NIHSSadmi	-0,067	0,935
IVThrombolysis	0,172	1,188
log_timeonsetimaging	-0,899	0,407
pcBATASs	0,238	1,269

FLAC-Regressionsanalyse: für gutes funktionelles Überleben. „Age“ (Alter); „NIHSSadmi“ (National Institutes of Health Stroke Scale bei Aufnahme), log_timeonsetimagingh („Zeit Symptombeginn bis erste Bildgebung“); „pcBATASs“ (kombinierter Score von BATMAN und pASCO). Auf die Berechnung von Konfidenzintervallen wurde bei Ridge- und FLAC-Regression verzichtet.

In Tabelle 6 sind die Modellparameter des FLAC-Regressionsmodells gezeigt. Ein übermäßig starkes „Shrinkage“ der Regressionskoeffizienten, wie im Ridge-Modell findet sich nicht.

1.4.5 Ergebnisse der Ereigniswahrscheinlichkeitsdichtegraphen

In Abb. 22, 23 und 24 und zeigt sich jeweils die Verteilung des prognostizierten, guten funktionellen Überlebens mittels klassisch entwickeltem Regressionsmodell, FLIC- und Ridge-Regressionsmodell für Patienten nach 90 Tagen und speziell dichotomisiert nach beobachtetem, guten funktionellen Überleben ($mRS \leq 2$).

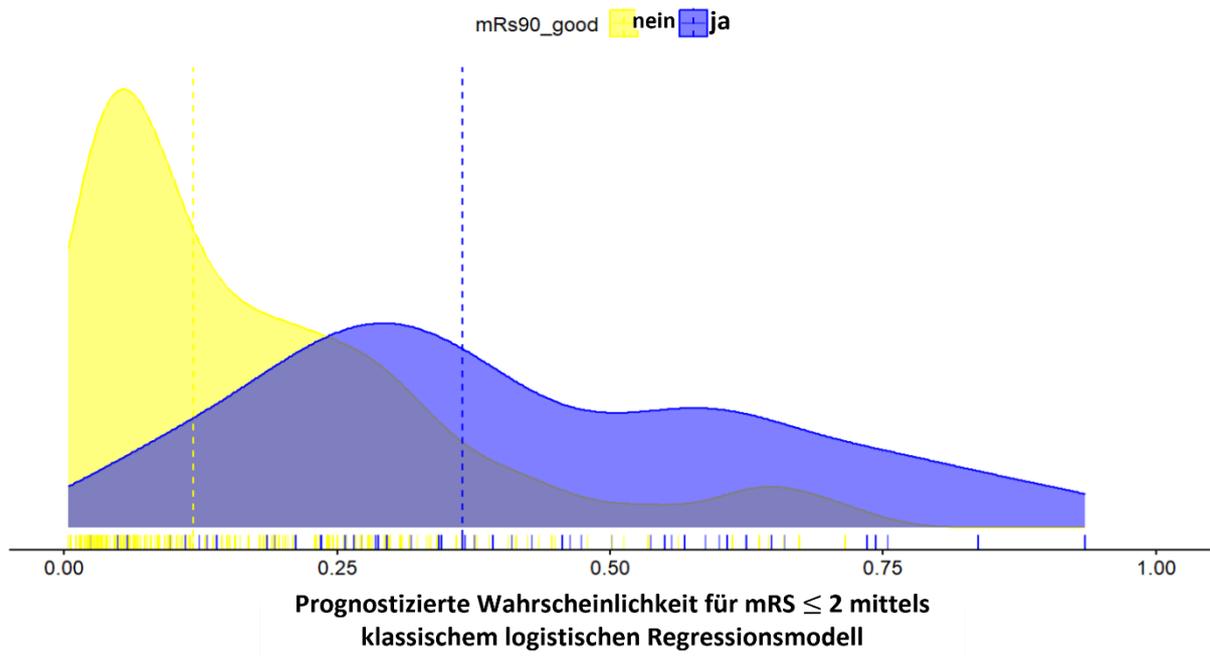


Abbildung 22: Ereigniswahrscheinlichkeitsdichtegraph 1. Ermittelte Ereigniswahrscheinlichkeiten für ein gutes funktionelles Überleben mittels klassischem Regressionsmodell dichotomisiert nach „beobachtetem“ Outcome. Die unterbrochenen Linien zeigen jeweils den Median des geschätzten, guten funktionellen Überlebens in den dichotomisierten Gruppen an.

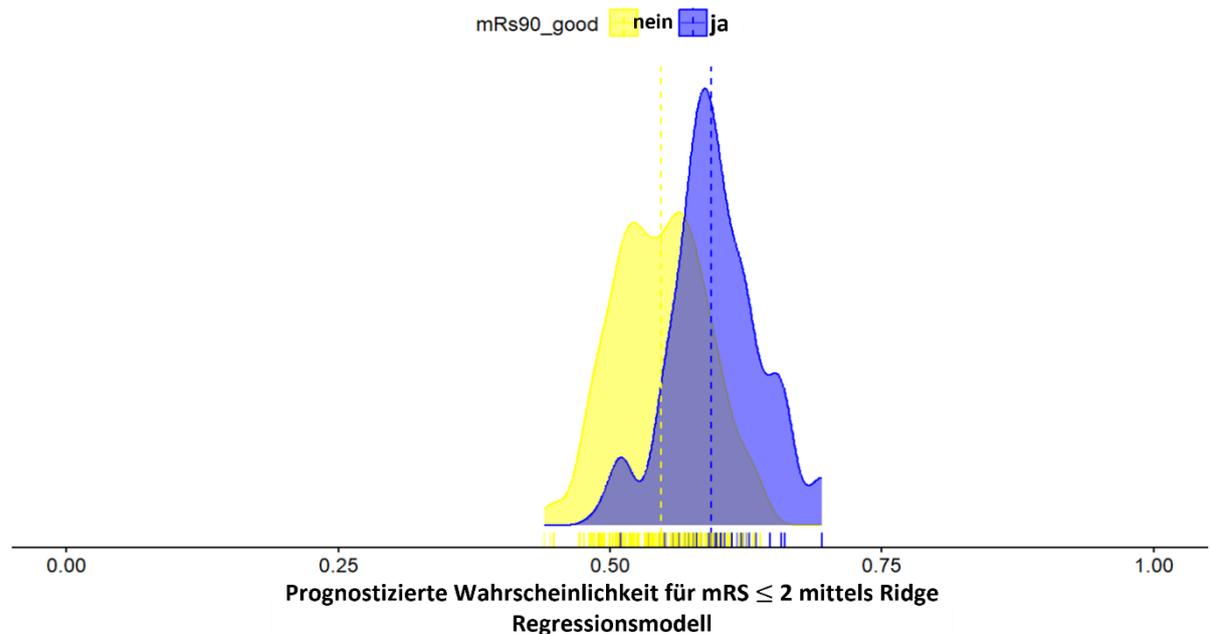


Abbildung 23: Ereigniswahrscheinlichkeitsdichtegraph 2. Ermittelte Ereigniswahrscheinlichkeiten für ein gutes funktionelles Überleben mittels

Ridge-Regressionsmodells dichotomisiert nach „beobachtetem“ Outcome. Die unterbrochenen Linien zeigen jeweils den Median des geschätzten, guten funktionellen Überlebens in den dichotomisierten Gruppen an.

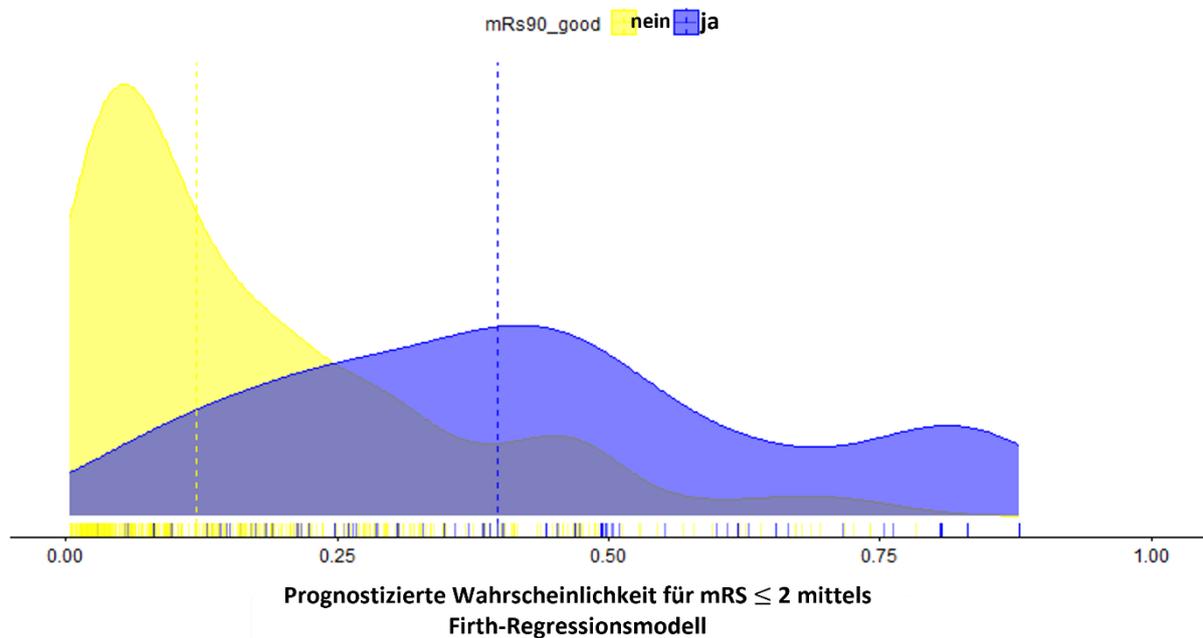


Abbildung 24: Ereigniswahrscheinlichkeitsdichtegraph 3. Ermittelte Ereigniswahrscheinlichkeiten für ein gutes funktionelles Überleben mittels FLAC-Regressionsmodell dichotomisiert nach „beobachtetem“ Outcome. Die unterbrochenen Linien zeigen jeweils den Median des geschätzten, guten funktionellen Überlebens in den dichotomisierten Gruppen an.

Im klassischen Regressionsmodell (Abbildung 22) ist die Überlappung in den Verteilungen der Wahrscheinlichkeiten für das dichotomisierte, beobachtete Überleben groß, wie auch die Optimismus-korrigierte AUC zeigt (siehe Tabelle 3). Die geschätzte mediane Gesamtwahrscheinlichkeit für ein gutes funktionelles Überleben beträgt 19%. Bei den Patienten, die ein gutes funktionelles Überleben zeigten, liegt der entsprechende geschätzte Median bei 36 %, während der Median bei Patienten, die ein schlechtes funktionelles Überleben zeigten, 12 % beträgt. In Abbildung 24 ergibt sich im FLAC-Regressionsmodell eine geschätzte mediane Gesamtwahrscheinlichkeit für ein gutes funktionelles Überleben von 17%. Bei den Patienten, die ein beobachtetes, gutes funktionelles Überleben zeigten, liegt der entsprechende geschätzte Median im Regressionsmodell bei 40 %, während der Median bei Patienten, die ein schlechtes funktionelles Überleben zeigten, 12 %

beträgt. Die niedrigen, geschätzten medianen Wahrscheinlichkeiten für ein gutes funktionelles Überleben in beiden Modellen, sowohl für die Gesamtkohorte als auch die dichotomisierten Kollektive zeigen sich unter anderem auch in dem negativen Intercept in beiden Modellen (siehe Tabelle 3 und 6).

Im Ridge-Regressionsmodell dagegen ist die Verteilungsbreite der prognostizierten Ereigniswahrscheinlichkeiten sehr gering (Abbildung 23). Die Überlappung in den Verteilungen der Wahrscheinlichkeiten für das dichotomisierte, beobachtete Outcome ist ebenfalls groß, wie auch die Optimismus-korrigierte AUC zeigt (Tabelle 6). Die mediane Gesamtwahrscheinlichkeit für ein gutes funktionelles Überleben beträgt 56%. Bei den Patienten, die ein gutes funktionelles Überleben zeigten, liegt der entsprechende Median bei 59 %, während der Median bei Patienten, die ein schlechtes funktionelles Überleben aufwiesen 55 % beträgt. Allgemein lässt der Ereigniswahrscheinlichkeitsdichtegraph eine sehr geringe Verteilungsbreite der geschätzten Modellprognosen und den sehr starken Einfluss des Intercepts in dem Ridge-Regressionsmodell erkennen (siehe auch Tabelle 4).

1.4.6 Ergebnisse der Leistungsfähigkeit des Modells

Für die Prädiktion von $mRS \leq 2$ lagen ROC-AUC, Sensitivität, Spezifität und korrespondierender "Cut-off" (bestimmt mittels Youden-Index) bei 0,81, 83%, 68% und 0,211 (Youden-Index = 1,5) für das klassische Regressionsmodell mit Rückwärtselimination; 0,82, 79% %, 74% % und 0,57 (Youden-Index = 1,5) für das Ridge-Regressionsmodell und 0,82, 82%, 70% % und 0,211 (Youden-Index = 1,5) für das FLAC-Regressionsmodell.

Das klassische logistische Regressionsmodell und das Ridge-Regressionsmodell zeigen eine ähnliche Diskriminationsfähigkeit. Für das logistische Regressionsmodell mittels Rückwärtselimination ergibt sich eine $AUC_{\text{korr}} = 0.779$ (0,803-0,75), für das Ridge-Regressionsmodell eine korrigierte $AUC_{\text{korr}} = 0,806$ (0,823-0,792). Der Brier-Score $_{\text{korr}} = 0,160$ ist im logistischen Regressionsmodell niedriger als im Ridge-Regressionsmodell (Brier Score $_{\text{korr}} = 0.268$). Das logistische Regressionsmodell mit Rückwärtselimination weist eine gute Kalibrierung auf (Slope $_{\text{korr}} = 0,847$ (0,726-0,945). Die korrigierte Steigung der Kalibrierungsfunktion des Ridge-Regressionsmodells beträgt 7,181 (7,762-6,679).

Tabelle 7: Modellevaluation des klassischen Logistischen Regressionsmodells mit Rückwärtselimination

Variablen	Original	Apparent(IQR)	Test(IQR)	Optimismus(IQR)	Korrigiert(IQR)
AUC	0,812	0,822 (0,794 -0,851)	0,790 (0,774-0,81)	0,032 (0,008-0,061)	0,779 (0,803-0,75)
Brier Scaled	0,234	0,252 (0,238-0,351)	0,177 (0,201-0,265)	0,075 (0,006-0,107)	0,160 (0,321-0,219)
Slope	1,000	1,000	0,847 (0,726-0,945)	0,153 (0,274-0,055)	0,847 (0,726-0,945)

AUC: „Area under the curve“; Brier Scaled: „standardisierter Brier-Score“; „Slope: „Steigung des Kalibrierungsgraphen“. Es sind jeweils der Median, 25%-Perzentil sowie 75%-Perzentil angegeben. IQR: Interquartilsabstand

Tabelle 8: Modellevaluation Ridge-Regression

Variablen	Original	Apparent(IQR)	Test(IQR)	Optimismus(IQR)	Korrigiert(IQR)
AUC	0.816	0.825 (0.809-0.837)	0.814 (0.811-0.817)	0.011 (-0.007-0.024)	0.806 (0.823-0.792)
Brier Scaled	0.267	0.266 (0.265-0.268)	0.267 (0.265-0.268)	-0.001 (-0.003-0.002)	0.268 (0.27-0.265)
Slope	7.538	7.680 (7.386-8.024)	7.323 (6.956-7.875)	0.357 (-0.224-0.859)	7.181 (7.762-6.679)

AUC: „Area under the curve“; Brier Scaled: „standardisierter Brier-Score“; „Slope: „Steigung des Kalibrierungsgraphen“. Es sind jeweils der Median, 25%-Perzentil, sowie 75%-Perzentil angegeben. IQR: Interquartilsabstand

Tabelle 9: Modellevaluation FLAC-Regression

Variablen	Original	Apparent(IQR)	Test(IQR)	Optimismus(IQR)	Korrigiert(IQR)
AUC	0,822	0,825 (0,811-0,836)	0,817 (0,816-0,820)	0,008 (-0,007-0,021)	0,809 (0,829-0,801)
Brier Scaled	0,134	0,132 (0,127-0,137)	0,135 (0,134-0,136)	-0,001 (-0,008-0,002)	0,136 (0,142-0,132)
Slope	1,020	1,021 (1,021-1,022)	1,004 (0,93-1,043)	0,357 (-0,022-0,094)	1,004 (1,042-0,926)

AUC: „Area under the curve“; Brier Scaled: „standardisierter Brier-Score“; „Slope: „Steigung des Kalibrierungsgraphen“. Es sind jeweils der Median, 25%-Perzentil, sowie 75%-Perzentil angegeben. IQR: Interquartilsabstand

Abbildung 25, 26 und Abbildung 27 zeigen jeweils die flexiblen Kalibrierungskurven für das klassische Regressionsmodell, das Ridge-Regressionsmodell und das FLAC-Regressionsmodell. Das klassische logistische Regressionsmodell und das FLAC-Regressionsmodell zeigen eine gute, schwache und moderate Kalibrierung. Das Ridge-Regressionsmodell weist einen deutlichen Kalibrierungsfehler auf.

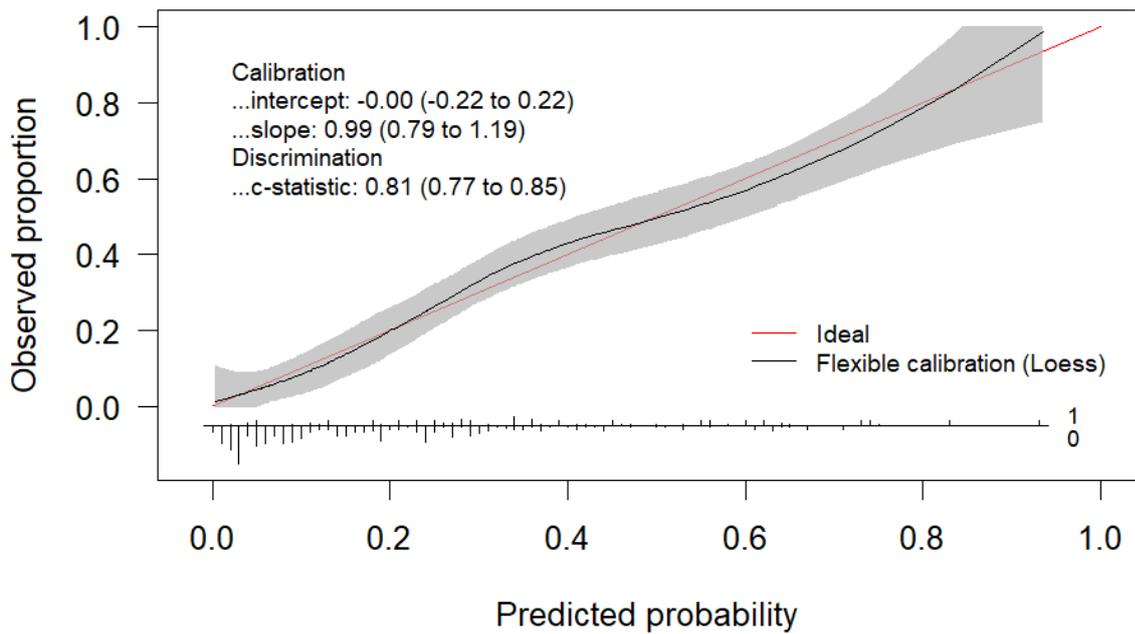


Abbildung 25: Flexible Kalibrierungskurve 1 für das klassische logistische Regressionsmodell. Das Konfidenzintervall ist grau unterlegt.

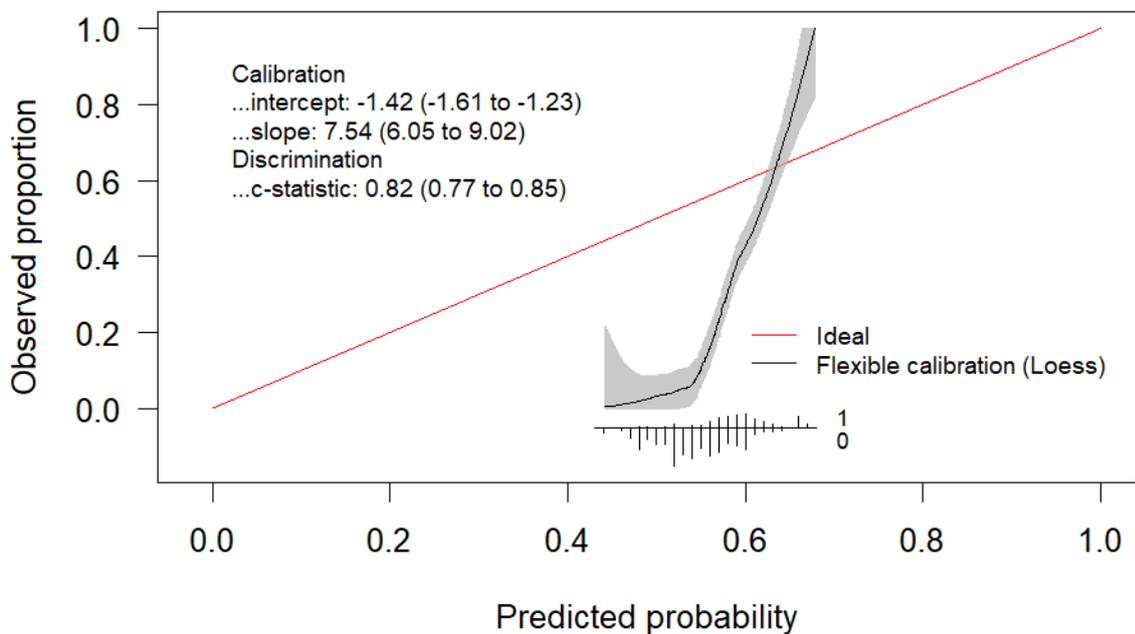


Abbildung 26: Flexible Kalibrierungskurve 2 für das Ridge-Regressionmodell. Das Konfidenzintervall ist grau hinterlegt.

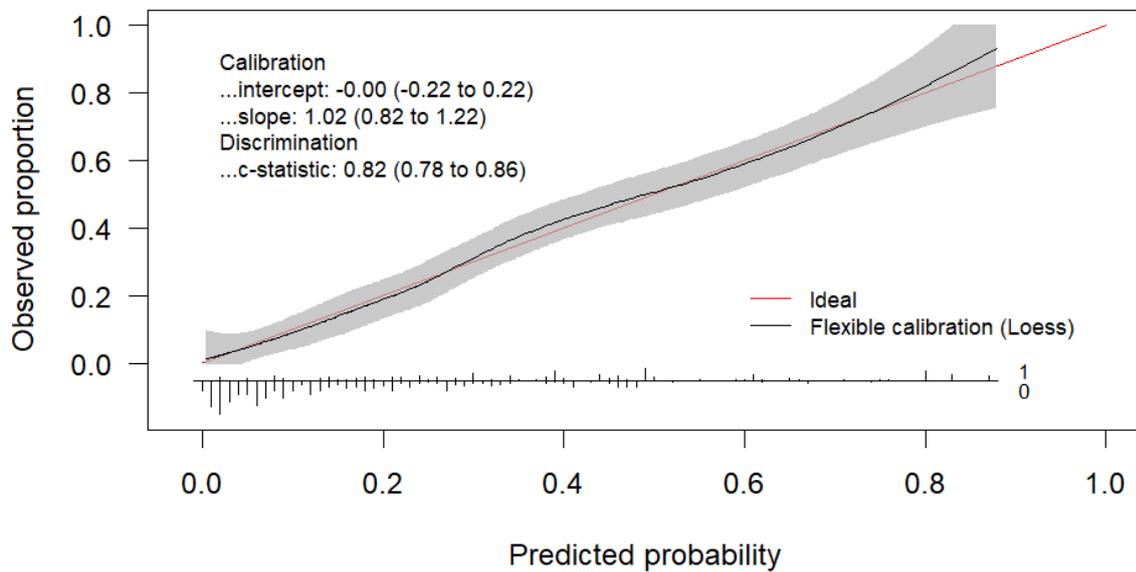


Abbildung 27: Flexible Kalibrierungskurve 3 für das FLAC-Regressionsmodell. Das Konfidenzintervall ist grau hinterlegt.

1.4.7 Ergebnisse der Entscheidungskurvenanalyse

In den Abbildungen 28, 29 und 30 sind die Entscheidungskurven der unterschiedlichen Modelle dargestellt.

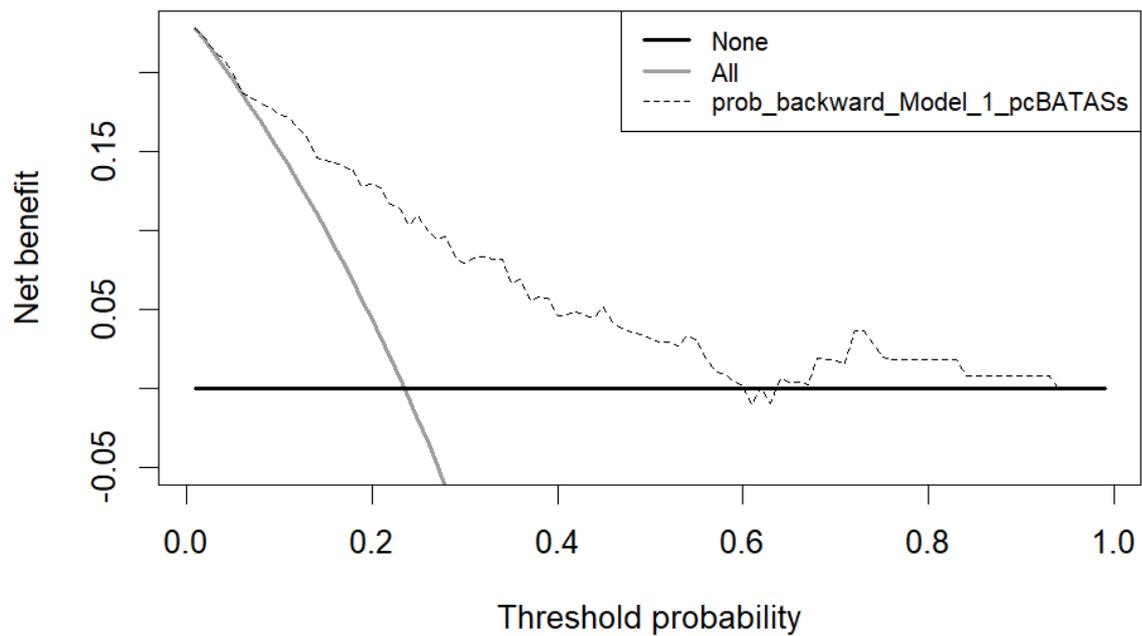


Abbildung 28: Entscheidungskurvenanalyse 1 für das klassische logistische Regressionsmodell. Die gestrichelte Kurve spiegelt den Net Benefit des Modells wider. „None“: Referenzstrategie „Keinen therapieren“; „All“: Referenztherapiestrategie „Alle therapieren“

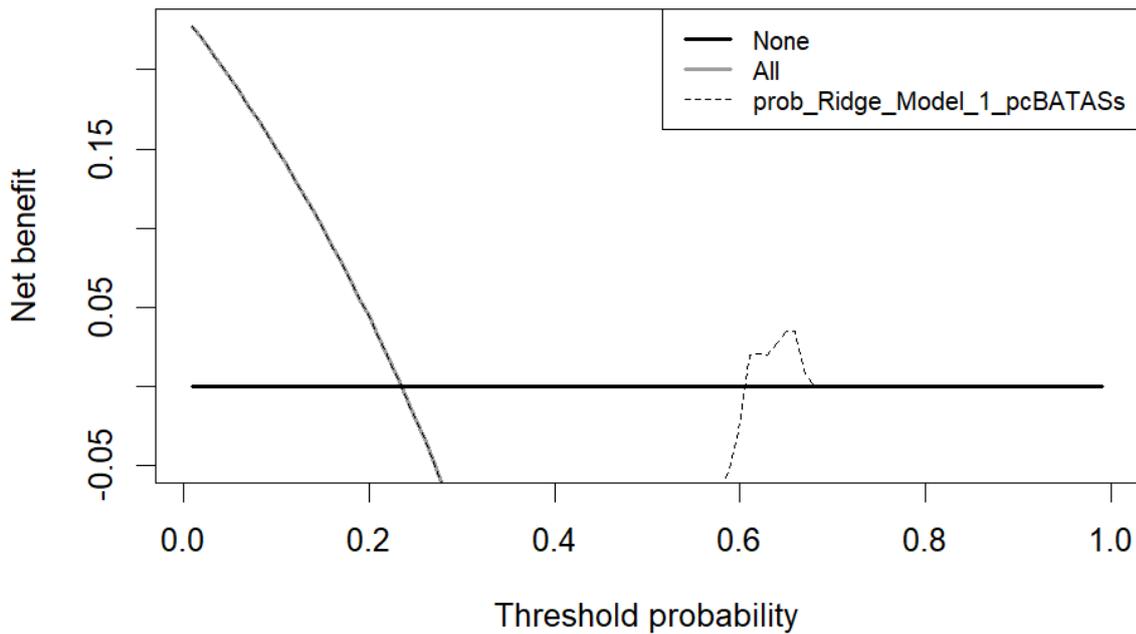


Abbildung 29: Entscheidungskurvenanalyse 2 für das Ridge-Regressionsmodell. Die gestrichelte Kurve spiegelt den Net Benefit des Modells wider. „None“: Referenzstrategie „Keinen therapieren“; „All“: Referenztherapiestrategie „Alle therapieren“

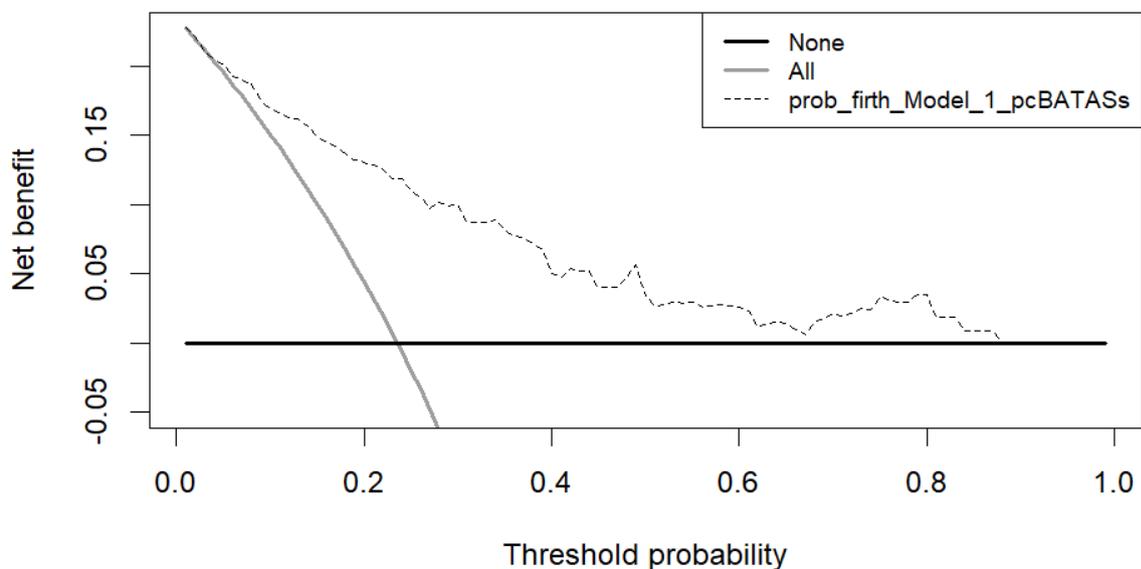


Abbildung 30: Entscheidungskurvenanalyse 3 für das FLAC-Regressionsmodell. Die gestrichelte Kurve spiegelt den Net Benefit des Modells wider. „None“: Referenzstrategie „Keinen therapieren“; „All“: Referenztherapiestrategie „Alle therapieren“

Die Entscheidungskurvenanalysen von klassischem und FLAC-Regressionsmodell zeigen einen größeren Net Benefit beider Modelle über einen weiten Schwellenwertrisikobereich im Vergleich zu den Referenztherapiestrategien (Abbildung 26 und Abbildung 28), wobei das FLAC-Regressionsmodell im Schwellenwertrisikobereich von 60-65% zusätzlich einen besseren Net Benefit als das „Backward-Modell“ aufweist.

Der Net Benefit des Ridge-Regressionsmodells entspricht in einem Schwellenwertrisikobereich von 0-30% dem der Referenztherapiestrategie „alle Patienten therapieren“. Bei höheren Schwellenwertrisikobereichen schneidet die Modell-Entscheidungskurve die Kurve der Referenztherapiestrategie „keinen Patienten therapieren“ und weist in dem Schwellenwertrisikointervall von 60-65% den größten Net Benefit im Vergleich über den gesamten Schwellenwertrisikobereich auf.

1.5 Diskussion

Ziel der Arbeit war es ein Prognosemodell für den posterioren Schlaganfall zu entwickeln, welches den komplexen, klinisch relevanten Wechselwirkungen verschiedener pathomechanistischer Einflussfaktoren dieses Krankheitsbildes Rechnung trägt. Es sollten neben pcBATASs, nur Parameter in das Modell implementiert werden, die standardisiert im akuten Notfall erhoben werden können und bereits als Prognosefaktoren für das untersuchte Krankheitsbild in wissenschaftlichen Publikationen etabliert wurden. Hierbei zeigten sich folgende Hauptergebnisse:

Die multiple Imputation mittels modifizierter Predictive-mean-matching-Methode (Midastouch) ermöglicht auch adäquate Imputationen fehlender Datenwerte bei geringer Fallzahl ($n = 119$) und verzerrtem, nicht normalverteiltem Datensatz durch Implementierung bestimmter „Kleinke-Kriterien“.

Der klassische Regression-Algorithmus (multivariable, logistische Regressionsanalyse mit Rückwärtselektion) zeigt mit den Parametern „pcBATASs“ (OR: 1,292; 95%KI: 1,028–1,624; $p = 0,028$); „Age“ (OR: 0,959; 95%CI: 0,923–0,996; $p = 0,031$) und „NIHSSadmi“ (OR: 0,940, 95%CI: 0,898–0,984, $p = 0,010$) eine gute prognostische Aussagekraft im Hinblick auf ein gutes, funktionales Überleben mit einer medianen $AUC_{\text{korrigiert}}(\text{IQR}) = 0,779$ (0,803-0,750); medianem standardisiertem Brier-Score $_{\text{korrigiert}}(\text{IQR}) = 0,160$ (0,321-0,219) und medianem Slope $_{\text{korrigiert}}(\text{IQR}) = 0,847$ (0,726-0,945).

Die Kombination von pcBATASs (OR: 1,269); „Age“ (OR: 0,958); „NIHSSadmi“ (OR: 0,935); „IVThrombolysis“ (OR = 1,188); „log_timeonsetimaging“ (OR = 0,407) im FLAC-Algorithmus zeigt ebenfalls eine gute prognostische Aussagekraft im Hinblick auf das funktionale Überleben „mRS ≤ 2 “: mediane $AUC_{\text{korrigiert}}(\text{IQR}) = 0,809$ (0,829-0,801); medianer standardisierter Brier-Score $_{\text{korrigiert}}(\text{IQR}) = 0,136$ (0,142-0,132); medianer Slope $_{\text{korrigiert}}(\text{IQR}) = 1,004$ (1,042-0,926).

Ridge-Regression kann, unabhängig vom EPV-Wert, bei Vorliegen von „Sparse Data Bias“ zu einem Overshrinkage der Regressionkoeffizienten beziehungsweise einem „Underfitting“ des Prognosemodells führen

Das pcBATASs-Prognosemodell fungiert neben pcASPECTS, PC-CS, BATMAN, und pASCO als funktionsfähiger (hybrider) neuroradiologischer Prognose-Score zur Prognosebestimmung des guten funktionellen Überlebens nach 90 Tagen von posterioren Schlaganfällen. Er berücksichtigt im Vergleich zu dem pASCO-Prognose-Score zusätzlich die Einflussfaktoren „Thrombuslast“ und „Thrombuslokalisation“.

In der DCA zeigt sich für das FLAC-Regressionsmodell das beste Net-Benefit-Profil über einen weiten Schwellenwertrisikobereich im Vergleich zu den Referenztherapiestrategien. Es sollte jedoch mit Vorsicht im Klinikalltag als Selektionsinstrument zur Thrombektomieevaluation eingesetzt werden, da der in dieser Arbeit verwendete Ansatz zur Modellentwicklung primär als Evaluation eines reinen Prognosemodells in einem behandelten Schlaganfallkollektiv zu verstehen ist und ein echter Therapieeffekt folglich nicht gemessen werden kann.

Im Folgenden werden die Ergebnisse der explorativen Modellentwicklung im Detail interpretiert und bewertet, sowie der klinische Nutzen der entwickelten Modelle nachvollzogen.

1.5.1 Prognosemodellentwicklung

Aufgrund der geringen Fallzahl und dem hohen Anteil fehlender Werte wurde die multiple Imputation (MI) mit Bootstrappingvalidierung (Val) kombiniert, um der Unsicherheit bei den Imputationen und der Unsicherheit bei der Auswahl der Modelle Rechnung zu tragen (MI-Val). Die dafür entwickelten Algorithmen sind dabei aus einem Kompromiss aus möglichst einfacher Umsetzung und dem übergeordneten Ziel, einem Overfitting beziehungsweise einem Underfitting entgegenzuwirken, entstanden. Grundsätzlich wird empfohlen zunächst Bootstrapping-Stichproben aus dem unvollständigen Datensatz zu ziehen und anschließend die multiple Imputation

durchzuführen („Val-MI“) (Musoro et al., 2014, Royle and Cairns, 2021, Wahl et al., 2016). Allerdings ist dadurch die Entwicklung stratifizierter MI-Algorithmen praktisch kaum umsetzbar, weshalb in dieser Arbeit ein gegenteiliger Ansatz verwendet wurde („MI-Val“). Rationale für diesen Ansatz war der Gedanke aufgrund der geringen Datenlage alle verfügbaren Informationen des Datensatzes zu nutzen. Weiterhin ist bekannt, dass die Verwendung von „stacked“ Imputationsstrategien (wie im Ridge- und FLAC-Modell verwendet) aufgrund der artifiziellen Stichprobenvergrößerung bei einem der Ridge-Regression ähnlichen Verfahren, der L1-Regularisierung, zu verzerrten Ergebnissen (z.B. verzerrte Konfidenzintervalle) führt, weshalb grundsätzlich separate Ansätze bevorzugt werden sollten (Gunn et al., 2022). Für Ridge-Modelle ist dieses Problem zwar noch nicht untersucht worden, jedoch nicht unwahrscheinlich. Allerdings sollte ungewollten Effekten, wie „Underfitting“ durch ein niedriges EPV, die sich im Besonderen bei Ridge-Regression ergeben, entgegengewirkt werden. Der „stacked“ Ansatz bietet die Möglichkeit ein niedriges EPV zu optimieren und zeigt dazu technisch deutliche Vorteile bezüglich Datenpräparation und Datenauswertung im Vergleich zu einem separaten Ansatz. Darüber hinaus verwendeten Gunn et. al. in ihrer Arbeit einen weniger optimierten MI-Algorithmus als in dieser Arbeit. Zudem konnten weitere Arbeiten von Thao & Gekus und Wood einen Nachteil in Bezug auf verzerrte Effektschätzungen im Rahmen eines „stacked“ Ansatzes nicht nachweisen (Thao and Gekus, 2019, Wood et al., 2008).

Bei einem Datensatz mit ähnlicher Prävalenz der beiden möglichen Ereignisse der binären, abhängigen Variable und ausreichendem Stichprobenumfang erleichtert die Maximum-Likelihood- Schätzung der Regressionskoeffizienten die Inferenz, d. h. die Interpretierbarkeit der Effektschätzer sowie die Genauigkeit der Vorhersagen unter Berücksichtigung der Kovariaten. Liegt dies nicht vor, wie bei dem in dieser Arbeit beschriebenen Datensatz, kann es zu verzerrten Regressionskoeffizienten und zu allzu optimistischen Schätzungen bei der Modellevaluation und der Modellanpassung kommen, zudem ist das Potenzial, prognoserelevante Variablen durch Variablenselektionsmethoden auszuwählen, begrenzt. Sie können die Gefahr bergen, Prädiktoren mit niedrigem Signal-Rauschverhältnis in das endgültige Modell aufzunehmen, sowie zu instabilen Variablenkonstellationen und dem „Drop Out“ wichtiger Confounder führen (Steyerberg et al., 2001a, Greenland et al., 2016, Pavlou et al., 2016). Durch die Implementierung von multipler Imputation und dem uniformen Shrinkage wird diesem potenziellen Bias effektiv entgegen gewirkt (Steyerberg et al.,

2013). Das Shrinkage der Regressionskoeffizienten nach Modellentwicklung gestaltet sich zudem technisch weniger aufwendig.

Die Rationale zur Verwendung eines gestapelten Datensatzes nach Durchführung der multiplen Imputation war es, insbesondere einem niedrigen EPV entgegenzuwirken, da in Datensätzen mit niedrigen Fallzahlen besonders die Verwendung von Ridge-Regression zu einem Overshrinkage der Regressionskoeffizienten beziehungsweise „Underfitting“, aber auch zu „Overfitting“ führen kann. Diese schwerwiegenden Modellierungsfehler zeigen sich unter anderem durch eine hohe Variabilität im Kalibrierungsgraphen (Pavlou et al., 2016, Van Calster et al., 2020). Firth-Regression beziehungsweise FLAC-Regression kann bei spärlichen Datensätzen hier einen Vorteil bieten. Allerdings kann der Einfluss des durch den Jeffry Priors induzierten Bias nicht genau bestimmt werden, wodurch eine Interpretation bei Effektschätzungen erschwert ist. Im Hinblick auf die allgemeine prognostische Aussagekraft ist das FLAC-Prognosemodell gut interpretierbar und der FLAC-Algorithmus funktioniert optimal bei Anwesenheit von „Sparse Data Bias“. Die Kombination von „pcBATASs“, „Age“, „NIHSSadmi“, „IVThrombolysis“, „log_timeonsetimaging“ zeigt hierbei im Hinblick auf das funktionale Überleben „mRS ≤ 2 “ eine gute prognostische Performance. Das klassische logistische Regressionsmodell weist etwas schwächere Modelleleistungsparameter auf, ist jedoch ebenfalls gut kalibriert. Das Ridge-Regressionsmodell weist einen deutlichen Kalibrierungsfehler auf („Slope $\gg 1$; siehe Abbildung 26).

Die vorhergesagte Wahrscheinlichkeit des Eintritts eines definierten Endpunktes kann systematisch zu hoch oder zu niedrig sein (= Kalibrierungsfehler) und/oder die relative Einstufung des individuellen Risikos eines Patienten der Validierungskohorte kann fehlerhaft sein (= Diskriminationsfehler). Angenommen, man untersucht eine Stichprobe einer Patientenpopulation mit posteriorem Schlaganfall, bei denen die (mediane) Gesamtsterblichkeitsrate nach 3 Monaten 70 % beträgt. Dann wäre ein Modell, das für jeden Patienten eine Mortalitätswahrscheinlichkeit von 70 % nach 3 Monaten vorhersagt, aber nicht zwischen Patienten unterscheiden kann, die noch leben, und solchen, die innerhalb des Intervalls sterben, perfekt kalibriert, jedoch nicht diskriminationsfähig. Umgekehrt würde ein Prognosemodell, das allen überlebenden Patienten eine Sterbewahrscheinlichkeit von 10 % und allen verstorbenen Patienten eine Sterbewahrscheinlichkeit von 11 % zuweist, zwar eine perfekte

Diskriminationsfähigkeit aufweisen, wäre aber schlecht kalibriert. Die relative Bedeutung von Kalibrierung und Diskrimination hängt von der beabsichtigten Anwendung ab. Wenn Prognosemodelle für die Beratung eines Patienten verwendet werden, spielt der Faktor der Kalibrierung eine übergeordnete Rolle. Patienten sind nicht daran interessiert, wie krank sie im Vergleich zu anderen Patienten mit der gleichen Erkrankung sind; stattdessen steht die Frage im Vordergrund, dass ihre Krankheit zum Tod oder zu einem anderen relevanten Endpunkt führt. Der Faktor der Kalibrierung ist auch in der Versorgungsforschung wichtig: Wenn vorhergesagte und beobachtete Sterberaten verglichen werden, um eine unerwartet hohe oder niedrige Mortalität pro Zeit in einzelnen Einrichtungen zu detektieren, können Kalibrierungsfehler zu dem Eindruck führen, dass eine große Anzahl an Krankenhäusern oder Leistungserbringern scheinbar übermäßig hohe oder sehr niedrige Sterblichkeitsraten generieren, obwohl tatsächlich das prognostische Modell nicht optimal kalibriert ist. Wenn dagegen Modelle zur Stratifizierung des Schweregrades erkrankter Patienten genutzt werden sollen, um die Effektivität von Behandlungen innerhalb eines bestimmten Stadiums zu vergleichen, ist der wichtigste Aspekt der Aussagekraft eines Prognosemodells, ob Patienten innerhalb eines Krankheitsstadiums mit gleicher Wahrscheinlichkeit einen definierten Endpunkt erfahren werden und ob die Krankheitsstadien entsprechend der Reihenfolge des Risikos richtig klassifiziert sind (Diskrimination).

Ein optimal kalibriertes Modell hat eine Steigung von 1 und ein Intercept von 0 im „Kalibrierungsplot“ (siehe hierzu auch Abbildungen 28, 29, 30). Steigungen $>>1$ deuten im Durchschnitt darauf hin, dass die Streuung (engl. „spread“) der vorhergesagten Prognosewahrscheinlichkeiten durch das Modell zu gering ist. Dies bestätigt sich grafisch im Ereigniswahrscheinlichkeitsdichtegraphen für das Ridge-Modell (siehe Abbildung 23). Eine mögliche Ursache für den geringen „Spread“ des Modells ist die zu starke Regularisierung durch den Tuningparameter und ein dadurch bedingtes „Overshrinkage“ der Regressionskoeffizienten beziehungsweise ein zu starker Einfluss des Intercepts in dem Prognosemodell. Ein Intercept < 0 deutet im Durchschnitt darauf hin, dass die vorhergesagten Prognosewahrscheinlichkeiten durch das Modell überschätzt sind (Van Calster et al., 2019). Dies wird durch den Vergleich des Ereigniswahrscheinlichkeitsdichtegraphen des Ridge-Modells mit den zwei Ereigniswahrscheinlichkeitsdichtegraphen der anderen beiden Modelle deutlich.

Eine Möglichkeit, den Ridge-Regression-Algorithmus zu optimieren, wäre der Einsatz von „Priors“ in einem hybriden Modell, im Sinne einer Kombination aus frequentistischer und baysscher Methodik: Zunächst wird z.B. ein informativer Prior mittels „Data Augmentation“ in den Datensatz (nach erfolgter multipler Imputation) implementiert, um einer Fehlkalibrierung vorzubeugen. Anschließend erfolgt die Ridge-Regressionsmodell-Entwicklung (Šinkovec et al., 2021). Alternativ könnte der Einsatz von Ridge-Regression mit „Perzentil Methode“ nach Roberts ebenfalls zu einer Optimierung des Modells führen, ebenso wie „Ereignismodellierung“ (z.B. „Propensity Scoring“) (Greenland et al., 2016, Pavlou et al., 2016, Roberts and Nowak, 2014).

Auf die Berechnung der Konfidenzintervalle der Regressionskoeffizienten des Ridge- und des FLAC-Regressionsmodells wurde aus den folgenden Gründen verzichtet: Konfidenzintervalle von Regressionskoeffizienten oder Konfidenzintervalle von Vorhersagewahrscheinlichkeiten, die im Rahmen von penalisierten Regressionsmodellen ermittelt werden, sind problematisch: Regularisierte Regressionsmodelle führen zu einer gezielten Verzerrung z.B. gegen den Wert 0, um den mittleren quadratischen Fehler zu minimieren (Goeman, 2009, Kyung et al., 2010). Eine Interpretation im Sinne von Effektschätzung wird dadurch schwierig. Zudem war das primäre Ziel dieser Arbeit, ein geeignetes Prognosemodell zu entwickeln.

Ein weiteres primäres Ziel in dieser Arbeit war es, Risikofaktoren für Overfitting und Underfitting, wie geringer Stichprobenumfang der Modellentwicklungspopulation, geringe Anzahl der Events oder zu hohe Anzahl der in Frage kommenden Prädiktoren, vorzubeugen. Die Einflussfaktoren Eventanzahl und Prädiktorenanzahl werden im EPV (Ereignisse pro Variable) abgebildet. Bei der Entwicklung eines Prognosemodells wird häufig als Faustregel zur Bestimmung des Stichprobenumfangs das Verhältnis 10:1 von Ereignissen pro Variable (EPV) genutzt (Peduzzi et al., 1995). Die entwickelten Modelle wiesen vor multipler Imputation ein $EPV < 10$ auf, wobei das EPV des Ridge- und FLAC-Regressionsmodells bei höherer Variablenzahl suboptimaler ausfällt, als das des klassischen logistischen Regressionsmodells mit Rückwärtselimination. Der entgegengesetzt wirkende Effekt durch den überhöhten Stichprobenumfang bedingt durch den gestapelten („stacked“) Ansatz im Rahmen der nicht-posthoc Penalisierungen mildert möglicherweise den Einfluss durch das niedrige EPV ab. Der Stichprobenumfang steht in folgender Beziehung zum EPV-Wert: Mit zunehmender Stichprobenanzahl, liegt wahrscheinlich grundsätzlich auch eine höhere

Eventrate vor. Zur Optimierung der Modellentwicklung und zur Reduktion des Phänomens "Overfitting", aber auch Underfitting ist deshalb eine geeignete Stichprobengröße n (Anzahl der Probanden) im Verhältnis zur Anzahl der Prädiktoren Parameter (p) erforderlich. Riley et. al. definierten die folgenden 4 Kriterien für die untere Grenze von n (=Fallzahl) im Rahmen von logistischen Regressionsmodellen mit prognostischer Zielfunktion:

- 1) geringer Optimismus bei den Prädiktorkoeffizienten des Prognosemodells, definiert durch einen globalen Schrumpfungsfaktor von $\geq 0,9$
- 2) eine absolute Differenz von $\leq 0,05$ zwischen dem scheinbaren und dem bereinigten R^2
- 3) präzise Schätzung (Fehler ≤ 10 % des wahren Wertes) der Standardabweichung der Residuen des Modells;
- 4) eine präzise Schätzung des basalen Risikos für den vorhergesagten Endpunkt (Modellintercept) (Riley et al., 2019b)

Zur Kalkulation der minimalen Stichprobengröße nach den Riley-Kriterien für die Entwicklung eines geeigneten Prognosemodells bei einer geschätzten Prävalenz des guten funktionellen Überlebens nach 90 Tagen thrombektomierter, posteriorer Schlaganfälle von 0,23 (190/829) (Zi et al., 2020) mit einer guten Diskriminationsfähigkeit wie die des Referenzmodells von Broocks, et al. ($AUC_{NIHSS + pASCO}=0,91$) ergibt sich für ein binäres, multivariablen Modell mit 5 Prädiktoren ein $n = 273$ mit 63 Events und einem optimalen EPV = 12,56 (Kalkulation mittels R-Package „pmsampsize“), um theoretisch ein Overfitting bzw. Underfitting zu minimieren und eine präzise Schätzung der Schlüsselparameter des Vorhersagemodells zu gewährleisten. Es ist grundlegend wichtig, die Modellierungsstrategie im Voraus festzulegen und sicherzustellen, dass die Stichprobengröße für die Anzahl der betrachteten Prädiktoren ausreicht (Riley et al., 2019a, van Smeden et al., 2019). Dies ist praktisch jedoch nicht immer umsetzbar, besonders wenn es sich um seltene Krankheitsbilder oder seltenere Entitäten häufiger Krankheitsbilder handelt, wie bei posterioren zerebralen Verschlüssen im Rahmen eines ischämischen Apoplex (Bogousslavsky et al., 1988, Caplan et al., 2004, Moulin et al., 1997). Daher sollten insbesondere in kleineren Datensätzen mit niedriger Eventrate Modellentwicklungsstrategien wie Regularisierungen (Ridge- oder Lasso-Regression) (Moons et al., 2004) oder einfache Modelle verwendet werden. Die Verwendung

simpler Modelle bedeuten weniger Prädiktoren in die Prognosekalkulation zu integrieren, nicht-lineare oder Interaktionsterme wegzulassen oder flexiblere Algorithmen zu verwenden (logistische Regressions- anstatt von Random Forest-Modellen oder im Vorfeld die Anzahl versteckter Neurone in einem neuronalen Netzwerk limitieren) (Van Calster et al., 2019).

Allerdings bietet Regularisierung keine Universallösung per se, um Unsicherheiten in kleinen Datensätzen vorzubeugen (Riley et al., 2021). Es konnte gezeigt werden, dass mittels Ridge-Logistik-Regression entwickelte Prognosemodelle in verschiedenen Szenarien (geringer Stichprobenumfang, unausgeglichene Ereignisraten, schwache Parametereffekte, geringe Anzahl echter Prädiktoren und/oder hohes Rauschen) eine höhere Steigungsvariabilität ihrer Kalibrierungskurve aufweisen (Šinkovec et al., 2021). Auch in diesem Datensatz konnte gezeigt werden, dass eine einfache Reduzierung des Problems auf das Verhältnis „Ereignisrate zu Anzahl der Kovariaten (EPV-Wert)“ nicht ausreichend ist.

1.5.2 Klinischer Nutzen des pcBATASs Prognosemodells

Dies ist eine weitere Arbeit, die ein modernes, aber simples bildgebendes Verfahren in einem neuen multivariablen Prognosemodell integriert, welches zur Beurteilung eines posterioren Schlaganfalls genutzt werden kann, ohne auf die Verwendung der CT-Perfusion oder auf die Notwendigkeit komplexer weiterer CT-Bildnachbearbeitung angewiesen zu sein. Das pcBATASs-Prognosemodell kombiniert die prognostischen Einflüsse von frühen ischämischen Veränderungen, intrakraniell Kollateralenstatus des hinteren Kreislaufs, Thrombuslast und spezifischer Thrombuslokalisierung mit weiteren klinischen Einflussfaktoren (Alter, NIHSS bei Aufnahme) und berücksichtigt damit im Vergleich zu dem pASCO-Score die wichtigen Einflussfaktoren: Thrombuslast und Thrombuslokalisierung. Generell können komplexe, langstreckige thrombotische Verschlüsse die mechanische Rekanalisation erschweren. Ein schlechtes funktionelles Überleben kann die Folge sein (Brandt et al., 1996, Da Ros et al., 2016). Darüber hinaus ist ein möglicher, modulierender Effekt der Thrombuslokalisierung auf die Kollateralenrekruierungsdynamik zu diskutieren: Die Thrombuslokalisierung gibt indirekt einen Hinweis darauf, welche Kollateralkreisläufe potentiell zur Verfügung stehen könnten.

Der praktische Unterschied zwischen pcBATASs und pASCO-Score wird an folgendem Beispiel deutlich: Ein proximaler Verschluss der Basilararterie (B1-Segment) und ein langstreckiger Verschluss des B1- und B2-Segments mit erhaltener Perfusion der PICA, SCA und PCA über beidseitige Pcoms werden im PCCS bzw. pASCO mit einem gleichen Scoringergebnis abgebildet. Für den BATMAN bzw. das pcBATASs Prognosemodell ergeben sich jedoch unterschiedliche Werte. langstreckige B1/B2-Verschlüsse weisen einen niedrigeren Score auf, als kurzstreckige Verschlüsse. Somit wird die generelle Thrombuslast des vertebrobasilären Systems im pcBATAS Prognosemodell deutlicher dargestellt als im pASCO. Die Thrombuslast hat grundsätzlich einen prognostischen Einfluss auf das Überleben von Schlaganfallpatienten (Puetz et al., 2008a).

Kürzlich wurde beschrieben, dass Patienten mit BAO und pcASPECTS ≥ 5 von der mechanischen Thrombektomie profitieren, dabei wurde jedoch beobachtet, dass Patienten mit einem höheren pcASPECTS immer noch signifikante Unterschiede im funktionellen Überleben aufwiesen und in dieser Subpopulation nur bei 11,2 % eine funktionelle Unabhängigkeit nach 90 Tagen vorlag (Sang et al., 2021) Die kombinierte Bewertung der frühen ischämischen Gewebehypoattenuation (d.h. pcASPECTS), der posterioren kollateralen Zirkulation und der Thrombuslast und Thrombuslokalisierung könnte die Vorhersage des funktionellen Überlebens noch weiter verbessern und könnte als Hilfsmittel für die Therapieplanung von posterioren Schlaganfällen verwendet werden, insbesondere bei Fällen mit unklarer Indikation zur mechanischen Thrombektomie (z.B. erweitertes Zeitfenster, niedrige pcASPECTS).

Vergleicht man die Modelle im Hinblick auf Sensitivität und Spezifität, dann zeigen das klassische Regressionsmodell und das FLAC-Regressionsmodell ähnliche Eigenschaften: Beide Modelle sind besser dafür geeignet, Patienten zu identifizieren, die nach Thrombektomie ein gutes funktionelles Überleben zeigen, im Vergleich zum Ridge-Regressionsmodell. Allerdings werden auch mehr Patienten auf Grund der niedrigeren Spezifität als falsch positiv klassifiziert, wodurch die Gefahr besteht, dass mehr potenziell verzichtbare Eingriffe durchgeführt werden, wenn nach Modellalgorithmus entschieden wird.

Die Betrachtung der Gesamtleistung der Modelle (standardisierter Brier Score) deutet jedoch auf den besten klinischen Nutzen durch das FLAC-Regressionsmodell hin. Der Brier Score bevorzugt grundsätzlich kontinuierliche Prognosemodelle über binäre

Testverfahren (Assel et al., 2017). Daher wurde zum besseren potenziellen Vergleich mit externen Prognosemodellen ergänzend eine Entscheidungskurvenanalyse durchgeführt

Das klassische logistische Regressionsmodell mit Rückwärtselimination zeigt nach Auswertung der Entscheidungskurvenanalyse (Abbildung 28) einen besseren klinischen Nutzen als Instrument zur Thrombektomieselektion im Vergleich zu den Referenztherapiestrategien [„alle Patienten therapieren (All) und keinen Patienten therapieren (None)“]. Lediglich in einem schmalen Schwellenwertrisikobereich (60-62%) hat es einen geringeren Net Benefit als die Referenztherapiestrategie „keinen Patienten therapieren“. Der Net Benefit des Ridge-Regressionsmodells entspricht in einem Schwellenwertrisikobereich von 0-30% dem der Referenztherapiestrategie „alle Patienten therapieren“. Bei höheren Schwellenwertrisikobereichen schneidet die Modell-Entscheidungskurve die Kurve der Referenztherapiestrategie „keinen Patienten therapieren“ und weist in einem schmalen Intervall (60-65%) einen höheren Net Benefit auf. Grundsätzlich führen Kalibrierungsfehler mit zunehmender Steigung >1 bei konstantem Intercept im „Calibration-Plot“ in niedrigen Schwellenwertrisikobereichen zur Angleichung der Model-Entscheidungskurve an die Entscheidungskurve der Referenztherapiestrategie „alle Patienten therapieren“ und bei hohem Schwellenwertrisiko zur Angleichung an die Entscheidungskurve der Referenztherapiestrategie „keinen Patienten therapieren“. Ein Kalibrierungsfehler mit einem negativen Intercept bei konstanter optimaler Steigung ($=1$) führt oberhalb der Ereignisrate (hier 24%) zu einem verminderten Net Benefit im Vergleich zur Referenztherapiestrategie „keinen Patienten therapieren“ und zu einer Verminderung des Net Benefits im Vergleich zur Referenztherapiestrategie „alle Patienten therapieren“ unterhalb der Ereignisrate (Van Calster and Vickers, 2015). Die Entscheidungskurve des Ridge-Regressionsmodells sollte aufgrund des schweren Kalibrierungsfehlers nicht verwendet werden. Das FLAC-Regressionsmodell zeigt über einen weiten Schwellenwertrisikobereich im Vergleich zu den Referenztherapiestrategien das beste Net-Benefit-Profil.

Einen weiteren Vorteil, den das pcBATASs-Prognosemodell bietet, liegt in der Modellentwicklungsmethodik begründet: Grundsätzlich variieren Patientenmerkmale und Krankheitsinzidenz- oder Prävalenzraten stark zwischen Gesundheitszentren,

Regionen und Ländern (Bray et al., 2018). Die Entwicklung von Prognosemodellen in „repräsentativen“ Kohorten ist praktisch unmöglich. Die Modellkalibrierung verschlechtert sich mit der Zeit als Reaktion auf die dynamische Natur des klinischen Umfelds. Dies wird als sogenannter „Calibration drift“ bezeichnet (Minne et al., 2012, Paul et al., 2012). Es ist jedoch enorm wichtig, dass Prognosemodelle ihre Vorhersagegenauigkeit im Laufe der Zeit beibehalten, um kontinuierlich eine gute und sichere Prognose zu gewährleisten. Um Kalibrierungsinhomogenitäten und Kalibrierungsverschiebungen („Calibration drift“) effektiv entgegenwirken zu können, bedarf es transparenter und einfacher Werkzeuge, die von lokalen Zentren unabhängig und nachvollziehbar eingesetzt werden können. Die Kombination von logistischen Regressionsmodellen und einfachen Modellaktualisierungsmethoden (z.B. uniformes Shrinkage), wie in dieser Arbeit verwendet, bietet in diesem Zusammenhang diverse Vorteile gegenüber anderen Algorithmen, wie „decision trees“ oder „support vector machine (SVM’s“): 1) Die Implementierung und Aktualisierung ist simpel (Van Calster et al., 2019) 2) Es müssen weniger Annahmen über die Verteilung der abhängigen Variablen berücksichtigt werden. 3) Es müssen keine linearen Beziehungen zwischen den Prädiktoren und der Zielvariablen angenommen werden. D.h. die Modelle können auf diverse Datensätze angewendet werden (Singh et al., 2016), was die unterschiedliche räumliche und zeitliche Implementierung vereinfacht. 4.) Darüber hinaus können einzelne Parameter logistischer Regressionsmodelle nachvollziehbar interpretiert werden. Dadurch kann zum Beispiel der Einfluss von Parametern eines Prognosemodells räumlich und zeitlich verglichen werden, so können möglicherweise pathoätiologische und pathomechanistische Unterschiede in verschiedenen Populationen detektiert werden oder das generelle Verständnis von Krankheiten verbessert werden.

Auf dem Gebiet der Prognoseforschung posteriorer Schlaganfälle ist es jedoch grundsätzlich notwendig, dass Daten zentrumsübergreifend besser harmonisiert werden sowie, dass bereits existierende Scores für Verschlüsse der hinteren Zirkulation in reinen BAO-Kollektiven und gemischten, posterioren Schlaganfallkohorten generell extern validiert werden (Salerno et al., 2022). Dafür ist eine transparente Dokumentation bei der Veröffentlichung von neuen Prognosemodellen, z.B. nach TRIPOD, essenziell.

1.5.3 Limitation der Arbeit

Grundlegende Limitationen dieser Arbeit sind zum einen der spärliche Datensatz mit geringer, nicht repräsentativer Fallzahl ($n = 119$), die unausgeglichene Ereignisraten, sowie der Anteil an fehlenden Datenwerten bestimmter Modellvariablen. Eine weitere Validierung in einem externen Datensatz ist enorm wichtig, um vor allem ein mögliches „Overfitting“ besser einschätzen zu können. Grundsätzlich besteht bei der Entwicklung von prognostischen Modellen in Beobachtungsstudien mit klassischem Studiendesign das Problem, dass kein Vergleich mit Kontrollgruppen möglich ist und ein potenzieller behandlungsmodifizierender Effekt auf den Therapieerfolg durch einen wahren Prädiktor oder ein prädiktives Modell nicht direkt gemessen werden kann. Bei dieser Art von Studiendesign muss die Annahme möglich sein, dass der Prädiktor oder das prädiktive Modell den untersuchten Endpunkt in Abwesenheit des Behandlungseffekts zu keinem Zeitpunkt beeinflusst, um die Definition eines echten Therapieeffektmodulators zu erfüllen. Wenn der Faktor mit dem Endpunkt bei behandelten und auch bei unbehandelten Personen assoziiert ist, dann kann er auch als therapieunabhängiger Prognosefaktor verstanden werden und nicht als reiner Prädiktor für das Ansprechen auf die Behandlung. Das Modell sollte daher mit Vorsicht im Klinikalltag als Selektionsinstrument zur Thrombektomieevaluation eingesetzt werden, da der in dieser Arbeit verwendete Ansatz zur Modellentwicklung primär als Evaluation eines reinen Prognosemodells in einem behandelten Schlaganfallkollektiv zu verstehen ist und ein echter Therapieeffekt folglich nicht gemessen werden kann.

Darüber hinaus ist ein Vergleich von Effektschätzungen durch das klassische Regressionsmodell und das FLAC-Regressionsmodell schwierig, da verschiedene Poolingmethoden („non-stacked“ vs. „stacked“) verwendet wurden.

Zudem erfolgte die Entwicklung und Validierung der Prognosemodelle in einem posterioren Schlaganfallkollektiv mit einem sehr hohen Anteil an BAO-Subtypen, weshalb die (klinische) Anwendung des Prognosemodells in Nicht-BAO-Kollektiven sehr bedacht genutzt werden sollte. Zudem ist das BATMAN-Klassifizierungssystem, welches auch indirekt Teil des pcBATASs – Prognosemodells ist, bislang nur für BAO – Schlaganfälle validiert worden.

1.6 Zusammenfassung

Die Auswahl von geeigneten Patienten für die Schlaganfallbehandlung ist entscheidend, da verheerende Komplikationen den Nutzen der Behandlung beeinträchtigen können. Komplexe, etablierte Prognosemodelle für isolierte Basilarisverschlüsse und für Kohorten mit unspezifischem posterioren Schlaganfall sind bislang nicht ausreichend oder nur teilweise evaluiert und validiert oder bilden nur unvollständig wichtige Einflussfaktoren auf den klinischen Verlauf erkrankter Patienten ab. Ziel dieser Arbeit war die Entwicklung und Optimierung eines neuen Prognosemodells zur Abschätzung des funktionellen Überlebens von Patienten nach Hirnarterienrekanalisation des hinteren Kreislaufes mithilfe von verschiedenen Modellregularisierungsalgorithmen, sowie interner Validierung unter Nutzung des Bootstrapping-Verfahrens in einem zunächst multipel imputierten Datensatz unter Gewährleistung PROGRESS-geforderter Qualitätsstandards. Ergänzend wurde eine Entscheidungskurvenanalyse durchgeführt, um den klinischen Nutzen des Typ-3-Modells nach dem PROGRESS-Klassifikationsschema zu quantifizieren. Mit dem pcBATASs-Prognosemodell werden die Wechselbeziehungen klinischer (Alter, NIHSS etc.) und neuroradiologischer Prognosefaktoren (BATMAN und pcASPECT Score) in einem multivariablen Prognosemodell abgebildet, welche den komplexen Zusammenhängen verschiedener, pathomechanistischer Einflussfaktoren auf den Krankheitsverlauf des posterioren Schlaganfalls Rechnung tragen. Als erweitertes Prognosemodell des pASCO-Scores berücksichtigt das pcBATASs-Modell zusätzlich den Einfluss der Thrombuslast und der Thrombuslokalisierung bei der Prognosevorhersage des funktionellen Überlebens (mRS) thrombektomierter Patienten nach 90 Tagen. Die multiple Imputation mittels modifizierter „Predictive-mean-matching“-Methode („Midastouch“) und definierten „Kleinke-Kriterien“ ermöglicht auch Imputationen fehlender Datenwerte bei niedriger Fallzahl ($n=119$) und verzerrtem, nicht-normalverteilten Datensatz. Die Prognosemodellentwicklung mittels FLAC-Algorithmus zeigt in dem spärlichen Datensatz die beste prognostische Aussagekraft für ein gutes funktionales Überleben („mRS ≤ 2 “) nach 90 Tagen: mediane $AUC_{\text{korrigiert}}$ (IQR) = 0,809 (0,829-0,801); medianer standardisierter Brier-Score_{korrigiert} (IQR) = 0,136 (0,142-0,132), medianer Slope_{korrigiert} (IQR) = 1,004 (1,042-0,926). Das Ridge-Regressionsmodell zeigt eine mediane $AUC_{\text{korrigiert}}$ (IQR) = 0,806 (0,823-0,792), einen medianen median Brier Score_{korrigiert} (IQR) = 0,268 (0,270-0,265),

einen medianen $\text{Slope}_{\text{korrigiert}}$ (IQR)= 7,181(7,762-6,679) bei schwerem Underfitting, trotz moderatem EPV. Es weist einen schweren Kalibrierungsfehler auf und sollte für die Patientenselektion zur Thrombektomie nicht verwendet werden. In der DCA zeigt sich das beste Net-Risiko-Profil für das FLAC-Regressionsmodell über einen weiten Schwellenwertrisikobereich im Vergleich zu den Referenztherapiestrategien. Es stellt ein mögliches Selektionsinstrument im Klinikalltag zur stratifizierten Thrombektomieevaluation dar.

1.7 Abstract

Appropriate selection of stroke patients for mechanical thrombectomy is essential to reduce complications and achieve maximal therapeutic benefit. Established prognostic models predicting the outcome of posterior stroke have not been sufficiently or partially evaluated, validated, or incompletely represent important factors influencing the clinical course of diseased patients. The aim of this work was the development and optimization of a new prognostic model for the estimation of the functional outcome of patients after mechanical recanalisation of the posterior circulation stroke using different model regularization algorithms, including internal validation using bootstrapping in an imputed data set following PROGRESS reporting standards. Further a decision curve analysis was performed to quantify the clinical utility of the type 3 model according to the PROGRESS classification scheme. With the pcBATASs - prognostic model, the interrelationships of clinical (age, NIHSS, etc.) and neuroradiological prognostic factors (BATMAN and pcASPECTs) in a multivariable prognostic model, which takes into account the complex interrelationships of different pathomechanistic factors influencing the disease course. The pcBATASs - prognostic model additionally considers the influence of the thrombus burden in the prognostic prediction of functional outcome (mRs after 90 days) of thrombectomized patients. Multiple imputation using modified predictive-mean-matching method ("midastouch") also allows imputation of missing data values when sample size is small ($n = 119$) and the data set is non-normally distributed.

Prognostic model development using FLAC algorithm shows the best prognostic performance in the sparse data setting for the dichotomised functional outcome ("mRS ≤ 2 ") at 90 days: median $AUC_{corrected}$ (IQR) = 0,809 (0,829-0,801); median standardized Brier-Score_{corrected} (IQR) = 0,136 (0,142-0,132), median Slope_{corrected} (IQR) = 1,004 (1,042-0,926). The ridge regression model shows median $AUC_{corrected}$ (IQR) = 0,806 (0,823-0,792); median Brier score_{corrected} (IQR) = 0,268 (0,270-0,265), median slope_{corrected} (IQR) = 7,181 (7,762-6,679) as a result of severe underfitting, despite moderate EPV. The Ridge Regressionmodell has a severe calibration error and should not be used for patient selection for thrombectomy or prognostic evaluation. DCA shows the best net benefit profil for the FLAC-Regressionmodell over a wide threshold risk range compared with reference therapy strategies. It could be used as selection tool for stratified thrombectomy evaluation.

1.8 Glossar

A.	Arteria
Abciximab_use	kodierte Variable für das binäre Ereignis „Abciximabtherapie i.v. in der Akuttherapie“
AICA	Anterior Inferior Cerebellar Artery
ASPECTS	Alberta Stroke Program Early CT Score
AUC	Area Under The Curve
Age	kodierte Variable für Patientenalter bei Aufnahme
anticoagulation	kodierte Variable für das binäre Ereignis „Antikoagulation in der Grundmedikation“
antiplateletTH_one	kodierte Variable für das binäre Ereignis „einfache Plättchenhemmung in der Grundmedikation“
antiplateletTH	kodierte Variable für das binäre Ereignis „doppelte Plättchenhemmung in der Grundmedikation“
BA	A. Basilaris
BAO	Basilar Artery Occlusion
BAO_involvement	kodierte Variable für das binäre Ereignis „Beteiligung der A. basilaris“
BATMAN	Basilar ArTery on computed toMography ANgiography
cerebellaredemascore	kodierte Variable für „JAUSS-Score“
ComorbidityDM	Kodierte Variable für „Grunderkrankung Diabetes mellitus Typ 2 bei Aufnahme“
ComorbidityHLP	kodierte Variable für Grunderkrankung „Hypercholesterinämie bei Aufnahme“
ComorbidityHTN	kodierte Variable für „Grunderkrankung Bluthochdruck bei Aufnahme“
Comorbidity_smoking_war_2	kodierte Variable „Rauchverhalten bei Aufnahme“
ComorbidityVHF	kodierte Variable für „Grunderkrankung Vorhofflimmern bei Aufnahme“
CTA	Computertomographische Angiografie
DSA	Digitale Subtraktionsangiografie
ECVA	extrakranielle Vertebralarterie
EPV	Events Pro Variable
etiology	kodierte Variable für den primären Pathomechanismus/Pathoätiologie nach CT/DSA Charakteristika
EVT	endovaskuläre Therapie
FLAC	Firth's Logistic regression with Added Covariate
FLIC	Firth's Logistic regression with Intercept-Correction
Glucoseinitial	kodierte Variable für „venöse Blutglucose bei Aufnahme“

intracran_thrombus_localisation	kodierte Variable für das jeweilige thrombotisch verschlossene Gefäß
ICVA	intrakranielle Vertebralarterie
ICU_days	kodierte Variable für „postinterventionellen Aufenthalt auf der Intensivstation (IST) in Tagen“
IVThrombolysis	kodierte Variable für das binäre Ereignis „i.v. Thrombolyse in der akuten Schlaganfalltherapie“
LASSO	Least Absolute Shrinkage and Selection Operator
LokaleLyse	kodierte Variable für das binäre Ereignis „intraarterielle medikamentöse Lyse in der Akuttherapie“
longOcclusion	kodierte Variable für das binäre Ereignis „langstreckiger thrombotischer Verschluss“
MALIGNANT	kodierte Variable für das binäre Ereignis „Maligner Infarkt im Aufnahme-CT“
MAR	Missing At Random
MCAR	Missing Completely At Random
MICE	Multivariate Imputation by Chained Equations
MCA	Middle Cerebral Artery
MNAR	Missing Not At Random
md.pattern	missing data pattern
mRS	modified Rankin Scale
mRspreStroke	kodierte Variable für mRS vor Schlaganfallereignis
mRs90	kodierte Variable für mRS 90 Tage nach Diagnosestellung
MVNI	MultiVariate Normal Imputation
NCCT	Non-Contrast Computer Tomography
NIHSS	National Institutes of Health Stroke Score
NIHSSadmi	kodierte Variable für NIHSS bei Aufnahme
NIHSS24h	kodierte Variable für den NIHSS 24h nach Aufnahme
NIHSSdis	kodierte Variable für den NIHSS bei Entlassung
No_of_passages	kodierte Variable für die Anzahl der Rekanalisationsversuche
NWU	Net Water Uptake
NWU_in_prozent	kodierte Variable für NWU bei Aufnahme
NWU2_in_prozent	kodierte Variable für NWU 24 Stunden nach Aufnahme
NWU_FU_in_prozent	kodierte Variable für NWU im Verlauf des ersten Krankenhausaufenthalts nach Schlaganfallereignis

pASCO	Hybrider Score aus pcASPECTS und PC-CS
PCA	Posteriore Cerebral Artery
pcASPECTS	Posterior Circulation Acute Stroke Prognosis Early CT Score
pcBATASs	Hybrider Score aus „Basilar Artery on Computed Tomography Angiography Prognostic Score (BATMAN)“ und Acute Stroke Prognosis Early CT Score (pcASPECTS)“
PC-CS	Posterior Circulation Collateral Score
PCCS	kodierte Variable für den PC-CS
Pcom	Posterior communicating artery
PICA	Posterior Inferior Cerebral Artery
PMM	Predictive Mean Matching
RCT	Randomized-Controlled-Trial
RRinitial	kodierte Variable für den Blutdruck bei Aufnahme
SCA	Superior Cerebral Artery
Sex	kodiert Variable für Patientengeschlecht bei Aufnahme
SMT	medizinischer Standardtherapie
SMCV	Superficial Middle Cerebral Vene
successRecan	kodierte Variable für das binäre Ereignis „TICI 2b/3“
TRIPOD	„Transparent Reporting Of A Multivariable Prediction Model For Individual Prognosis Or Diagnosis“
TICI	Thrombolysis in Cerebral Infarction
TICI_cod	kodierte Variable für den TICI-Score
timeonsetimagingh	kodierte Variable für Zeit Symptombeginn bis erste Bildgebung in Stunden nach Schlaganfallereignis
timeImagingAngioh	kodierte Variable für Zeit erste Bildgebung bis Leistenpunktion in Stunden bei Schlaganfallereignis
VISTA	Virtuelles Internationales Schlaganfall-Studienarchiv

1.9 Literaturverzeichnis

- AFANADOR, N. L., SMOLINSKA, A., TRAN, T. N. & BLANCHET, L. 2016. Unsupervised random forest: a tutorial with case studies. *Journal of Chemometrics*, 30, 232-241.
- AHMED, R. A., DMYTRIW, A. A., PATEL, A. B., STAPLETON, C. J., VRANIC, J. E., RABINOV, J. D., LESLIE-MAZWI, T. M., ROST, N. S., HIRSCH, J. A. & REGENHARDT, R. W. 2022. Basilar artery occlusion: A review of clinicoradiologic features, treatment selection, and endovascular techniques. *Interv Neuroradiol*, 15910199221106049.
- AL., A. E. M. 2017. *DescTools: Tools for Descriptive Statistics* [Online]. Available: <https://cran.r-project.org/package=DescTools> [Accessed 18.01.2023 2023].
- ALEMSEGED, F., SHAH, D. G., DIOMEDI, M., SALLUSTIO, F., BIVARD, A., SHARMA, G., MITCHELL, P. J., DOWLING, R. J., BUSH, S., YAN, B., CALTAGIRONE, C., FLORIS, R., PARSONS, M. W., LEVI, C. R., DAVIS, S. M. & CAMPBELL, B. C. V. 2017. The Basilar Artery on Computed Tomography Angiography Prognostic Score for Basilar Artery Occlusion. *Stroke*, 48, 631-637.
- ALI, M., BATH, P. M. W., CURRAM, J., DAVIS, S. M., DIENER, H.-C., DONNAN, G. A., FISHER, M., GREGSON, B. A., GROTTA, J., HACKE, W., HENNERICI, M. G., HOMMEL, M., KASTE, M., MARLER, J. R., SACCO, R. L., TEAL, P., WAHLGREN, N.-G., WARACH, S., WEIR, C. J. & LEES, K. R. 2007. The Virtual International Stroke Trials Archive. *Stroke*, 38, 1905-1910.
- ALTMAN, D. G., LAUSEN, B., SAUERBREI, W. & SCHUMACHER, M. 1994. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*, 86, 829-35.
- ALTMAN, D. G. & ROYSTON, P. 2000. What do we mean by validating a prognostic model? *Stat Med*, 19, 453-73.
- AMBLER, G., SEAMAN, S. & OMAR, R. Z. 2012. An evaluation of penalised survival methods for developing prognostic models with rare events. *Statistics in Medicine*, 31, 1150-1161.
- ARBABSHIRANI, M. R., PLIS, S., SUI, J. & CALHOUN, V. D. 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145, 137-165.
- ARNOLD, A. M. & KRONMAL, R. A. 2003. Multiple imputation of baseline data in the cardiovascular health study. *Am J Epidemiol*, 157, 74-84.
- ARNOLD, M., NEDELTCHEV, K., SCHROTH, G., BAUMGARTNER, R. W., REMONDA, L., LOHER, T. J., STEPPER, F., STURZENEGGER, M., SCHUKNECHT, B. & MATTLE, H. P. 2004. Clinical and radiological predictors of recanalisation and outcome of 40 patients with acute basilar artery occlusion treated with intra-arterial thrombolysis. *Journal of neurology, neurosurgery, and psychiatry*, 75, 857-862.
- ASSEL, M., SJOBERG, D. D. & VICKERS, A. J. 2017. The Brier score does not evaluate the clinical utility of diagnostic tests or prediction models. *Diagnostic and Prognostic Research*, 1, 19.
- AUSTIN, P. C. & STEYERBERG, E. W. 2014. Bootstrap confidence intervals for loess-based calibration curves. *Statistics in Medicine*, 33, 2699-2700.
- BARBER, P. A., DEMCHUK, A. M., ZHANG, J. & BUCHAN, A. M. 2000. Validity and reliability of a quantitative computed tomography score in predicting outcome of

- hyperacute stroke before thrombolytic therapy. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. *Lancet*, 355, 1670-4.
- BARLINN, K., BECKER, U., PUETZ, V., DZIALOWSKI, I., KUNZ, A., KEPPLINGER, J., VON KUMMER, R. & GAHN, G. 2012. Combined treatment with intravenous abciximab and intraarterial tPA yields high recanalization rate in patients with acute basilar artery occlusion. *J Neuroimaging*, 22, 167-71.
- BEJOT, Y., BAILLY, H., DURIER, J. & GIROUD, M. 2016. Epidemiology of stroke in Europe and trends for the 21st century. *Presse Med*, 45, e391-e398.
- BERKHEMER, O. A., FRANSEN, P. S., BEUMER, D., VAN DEN BERG, L. A., LINGSMA, H. F., YOO, A. J., SCHONEWILLE, W. J., VOS, J. A., NEDERKOORN, P. J., WERMER, M. J., VAN WALDERVEEN, M. A., STAALS, J., HOFMEIJER, J., VAN OOSTAYEN, J. A., LYCKLAMA À NIJEHOLT, G. J., BOITEN, J., BROUWER, P. A., EMMER, B. J., DE BRUIJN, S. F., VAN DIJK, L. C., KAPPELLE, L. J., LO, R. H., VAN DIJK, E. J., DE VRIES, J., DE KORT, P. L., VAN ROOIJ, W. J., VAN DEN BERG, J. S., VAN HASSELT, B. A., AERDEN, L. A., DALLINGA, R. J., VISSER, M. C., BOT, J. C., VROOMEN, P. C., ESHGHI, O., SCHREUDER, T. H., HEIJBOER, R. J., KEIZER, K., TIELBEEK, A. V., DEN HERTOOG, H. M., GERRITS, D. G., VAN DEN BERG-VOS, R. M., KARAS, G. B., STEYERBERG, E. W., FLACH, H. Z., MARQUERING, H. A., SPRENGERS, M. E., JENNISKENS, S. F., BEENEN, L. F., VAN DEN BERG, R., KOUDSTAAL, P. J., VAN ZWAM, W. H., ROOS, Y. B., VAN DER LUGT, A., VAN OOSTENBRUGGE, R. J., MAJOIE, C. B. & DIPPEL, D. W. 2015. A randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med*, 372, 11-20.
- BOEHME, A. K., RAWAL, P. V., LYERLY, M. J., ALBRIGHT, K. C., BAVARSAD SHAHRIPOUR, R., PALAZZO, P., KAPOOR, N., ALVI, M., HOUSTON, J. T., HARRIGAN, M. R., CAVA, L., SISSON, A., ALEXANDROV, A. W. & ALEXANDROV, A. V. 2014. Investigating the Utility of Previously Developed Prediction Scores in Acute Ischemic Stroke Patients in the Stroke Belt. *Journal of Stroke and Cerebrovascular Diseases*, 23, 2001-2006.
- BOGOUSSLAVSKY, J., VAN MELLE, G. & REGLI, F. 1988. The Lausanne Stroke Registry: analysis of 1,000 consecutive patients with first stroke. *Stroke*, 19, 1083-92.
- BRACARD, S., DUCROCQ, X., MAS, J. L., SOUDANT, M., OPPENHEIM, C., MOULIN, T. & GUILLEMIN, F. 2016. Mechanical thrombectomy after intravenous alteplase versus alteplase alone after stroke (THRACE): a randomised controlled trial. *Lancet Neurol*, 15, 1138-47.
- BRANDT, T., KNAUTH, M., WILDERMUTH, S., WINTER, R., KUMMER, R. V., SARTOR, K. & HACKE, W. 1999. CT Angiography and Doppler Sonography for Emergency Assessment in Acute Basilar Artery Ischemia. *Stroke*, 30, 606-612.
- BRANDT, T., PESSIN, M. S., KWAN, E. S. & CAPLAN, L. R. 1995. Survival with Basilar Artery Occlusion. *Cerebrovascular Diseases*, 5, 182-187.
- BRANDT, T., VON KUMMER, R., MÜLLER-KÜPPERS, M. & HACKE, W. 1996. Thrombolytic therapy of acute basilar artery occlusion. Variables affecting recanalization and outcome. *Stroke*, 27, 875-81.
- BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68, 394-424.
- BROOCKS, G., ELSAYED, S., KNIEP, H., KEMMLING, A., FLOTTMANN, F., BECHSTEIN, M., FAIZY, T. D., MEYER, L., LINDNER, T., SPORNS, P.,

- RUSCHE, T., SCHÖN, G., MADER, M. M., NAWABI, J., FIEHLER, J. & HANNING, U. 2021. Early Prediction of Malignant Cerebellar Edema in Posterior Circulation Stroke Using Quantitative Lesion Water Uptake. *Neurosurgery*, 88, 531-537.
- BROOCKS, G., FAIZY, T. D., MEYER, L., GROFFMANN, M., ELSAYED, S., KNIEP, H., FLOTTMANN, F., BECHSTEIN, M., RUSCHE, T., SCHÖN, G., NAWABI, J., SPORNS, P., FIEHLER, J., KEMMLING, A. & HANNING, U. 2022a. Posterior circulation collateral flow modifies the effect of thrombectomy on outcome in acute basilar artery occlusion. *Int J Stroke*, 17, 761-769.
- BROOCKS, G., FLOTTMANN, F., ERNST, M., FAIZY, T. D., MINNERUP, J., SIEMONSEN, S., FIEHLER, J. & KEMMLING, A. 2018. Computed Tomography-Based Imaging of Voxel-Wise Lesion Water Uptake in Ischemic Brain: Relationship Between Density and Direct Volumetry. *Invest Radiol*, 53, 207-213.
- BROOCKS, G., GROFFMANN, M., MEYER, L., ELSAYED, S., KNIEP, H., KEMMLING, A., VAN HORN, N., MCDONOUGH, R., FAIZY, T. D., BECHSTEIN, M., SPORNS, P., RUSCHE, T., SCHÖN, G., NAWABI, J., FIEHLER, J. & HANNING, U. 2022b. Higher baseline blood glucose is associated with reduced likelihood for successful recanalization in patients with basilar artery occlusion. *J Neurol*, 269, 3286-3294.
- BROOCKS, G., MEYER, L., FAIZY, T. D., ELSAYED, S., KNIEP, H., KEMMLING, A., HAUPT, W., MCDONOUGH, R., BECHSTEIN, M., SPORNS, P., RUSCHE, T., SCHÖN, G., NAWABI, J., FIEHLER, J. & HANNING, U. 2022c. New imaging score for outcome prediction in basilar artery occlusion stroke. *European Radiology*, 32, 4491-4499.
- BURGES, C. J. C. 1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121-167.
- BUUREN, S. V. & GROOTHUIS-OUDSHOORN, K. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45.
- CAMPBELL, B. C., MITCHELL, P. J., KLEINIG, T. J., DEWEY, H. M., CHURILOV, L., YASSI, N., YAN, B., DOWLING, R. J., PARSONS, M. W., OXLEY, T. J., WU, T. Y., BROOKS, M., SIMPSON, M. A., MITEFF, F., LEVI, C. R., KRAUSE, M., HARRINGTON, T. J., FAULDER, K. C., STEINFORT, B. S., PRIGLINGER, M., ANG, T., SCROOP, R., BARBER, P. A., MCGUINNESS, B., WIJERATNE, T., PHAN, T. G., CHONG, W., CHANDRA, R. V., BLADIN, C. F., BADVE, M., RICE, H., DE VILLIERS, L., MA, H., DESMOND, P. M., DONNAN, G. A. & DAVIS, S. M. 2015. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med*, 372, 1009-18.
- CAPLAN, L. R., WITYK, R. J., GLASS, T. A., TAPIA, J., PAZDERA, L., CHANG, H. M., TEAL, P., DASHE, J. F., CHAVES, C. J., BREEN, J. C., VEMMOS, K., AMARENCO, P., TETTENBORN, B., LEARY, M., ESTOL, C., DEWITT, L. D. & PESSIN, M. S. 2004. New England Medical Center Posterior Circulation registry. *Ann Neurol*, 56, 389-98.
- CHANDRA, R. V., LAW, C. P., YAN, B., DOWLING, R. J. & MITCHELL, P. J. 2011. Glasgow coma scale does not predict outcome post-intra-arterial treatment for basilar artery thrombosis. *AJNR Am J Neuroradiol*, 32, 576-80.
- CHARLSON, M. E., POMPEI, P., ALES, K. L. & MACKENZIE, C. R. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*, 40, 373-83.
- CHUNG, C.-P., YONG, C.-S., CHANG, F.-C., SHENG, W.-Y., HUANG, H.-C., TSAI, J.-Y., HSU, H.-Y. & HU, H.-H. 2015. Stroke etiology is associated with outcome

- in posterior circulation stroke. *Annals of Clinical and Translational Neurology*, 2, 510-517.
- CLARK, G. M. 2008. Prognostic factors versus predictive factors: Examples from a clinical trial of erlotinib. *Mol Oncol*, 1, 406-12.
- COLLINS, G. S., OGUNDIMU, E. O. & ALTMAN, D. G. 2016. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Statistics in Medicine*, 35, 214-226.
- COLLINS, L. M., SCHAFER, J. L. & KAM, C. M. 2001. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*, 6, 330-51.
- COLLINS, R. 2012. What makes UK Biobank special? *Lancet*, 379, 1173-4.
- COPAS, J. 1983. Regression, Prediction and Shrinkage. *Journal of the royal statistical society series b-methodological*, 45, 311-335.
- COPAS, J. B. 1997. Using regression models for prediction: shrinkage and regression to the mean. *Stat Methods Med Res*, 6, 167-83.
- COSTALAT, V., MACHI, P., LOBOTESIS, K., MALDONADO, I., VENDRELL, J. F., RIQUELME, C., MOURAND, I., MILHAUD, D., HÉROUM, C., PERRIGAULT, P.-F., ARQUIZAN, C. & BONAFÉ, A. 2011. Rescue, Combined, and Stand-Alone Thrombectomy in the Management of Large Vessel Occlusion Stroke Using the Solitaire Device: A Prospective 50-Patient Single-Center Study. *Stroke*, 42, 1929-1935.
- COYLE, P. & HEISTAD, D. D. 1991. Development of collaterals in the cerebral circulation. *Blood Vessels*, 28, 183-9.
- CRAMER, J. 2002. The Origins of Logistic Regression. Tinbergen Institute.
- CROSS, D. T., 3RD, MORAN, C. J., AKINS, P. T., ANGTUACO, E. E., DERDEYN, C. P. & DIRINGER, M. N. 1998. Collateral circulation and outcome after basilar artery thrombolysis. *AJNR Am J Neuroradiol*, 19, 1557-63.
- DA ROS, V., MESCHINI, A., GANDINI, R., DEL GIUDICE, C., GARACI, F., STANZIONE, P., RIZZATO, B., DIOMEDI, M., SIMONETTI, G., FLORIS, R. & SALLUSTIO, F. 2016. Proposal for a Vascular Computed Tomography-Based Grading System in Posterior Circulation Stroke: A Single-Center Experience. *Journal of Stroke and Cerebrovascular Diseases*, 25, 368-377.
- DAVIS, S. E., LASKO, T. A., CHEN, G., SIEW, E. D. & MATHENY, M. E. 2017. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc*, 24, 1052-1061.
- DE COCK B, N. D., VAN CALSTER B, STEYERBERG E, VERGOUWE Y. 2016. *The CalibrationCurves package: validating predicted probabilities against binary events* [Online]. Available: <https://github.com/BavoDC/CalibrationCurves> [Accessed 18.01.2023 2023].
- DONDERS, A. R. T., VAN DER HEIJDEN, G. J. M. G., STIJNEN, T. & MOONS, K. G. M. 2006. Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59, 1087-1091.
- DONG, Y. & PENG, C.-Y. J. 2013. Principled missing data methods for researchers. *SpringerPlus*, 2, 222.
- DUZAN, H. & MOHAMAD SHARIFF, N. S. 2015. Ridge Regression for Solving the Multicollinearity Problem: Review of Methods and Models. *Journal of Applied Sciences*, 15, 392-404.
- ECKERT, B., KOCH, C., THOMALLA, G., KUCINSKI, T., GRZYSKA, U., ROETHER, J., ALFKE, K., JANSEN, O. & ZEUMER, H. 2005. Aggressive Therapy With Intravenous Abciximab and Intra-Arterial rtPA and Additional PTA/Stenting

- Improves Clinical Outcome in Acute Vertebrobasilar Occlusion. *Stroke*, 36, 1160-1165.
- EFRON, B. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, 1-26, 26.
- FABRITIUS, M. P., THIERFELDER, K. M., MEINEL, F. G., OTHMAN, A. E., DORN, F., SABEL, B. O., SCHEFFLER, P., ERTL-WAGNER, B., SOMMER, W. H. & KUNZ, W. G. 2017. Early Imaging Prediction of Malignant Cerebellar Edema Development in Acute Ischemic Stroke. *Stroke*, 48, 2597-2600.
- FEYEN, L., SCHOTT, P., OCHMANN, H., KATOH, M., HAAGE, P. & FREYHARDT, P. 2022. Value of machine learning to predict functional outcome of endovascular treatment for acute ischaemic stroke of the posterior circulation. *Neuroradiol J*, 35, 363-369.
- FIRTH, D. 1993. Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80, 27-38.
- FLODEN, L. & BELL, M. L. 2019. Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study. *BMC Medical Research Methodology*, 19, 161.
- FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33, 1-22.
- GLASS, T. A., HENNESSEY, P. M., PAZDERA, L., CHANG, H. M., WITYK, R. J., DEWITT, L. D., PESSIN, M. S. & CAPLAN, L. R. 2002. Outcome at 30 days in the New England Medical Center Posterior Circulation Registry. *Arch Neurol*, 59, 369-76.
- GOEMAN, J. 2009. L1 and L2 Penalized Regression Models.
- GOULD, S. J. 1985. The Median isn't the Message. *Discover Magazine*.
- GOYAL, M., DEMCHUK, A. M., MENON, B. K., EESA, M., REMPEL, J. L., THORNTON, J., ROY, D., JOVIN, T. G., WILLINSKY, R. A., SAPKOTA, B. L., DOWLATSHAHI, D., FREI, D. F., KAMAL, N. R., MONTANERA, W. J., POPPE, A. Y., RYCKBORST, K. J., SILVER, F. L., SHUAIB, A., TAMPIERI, D., WILLIAMS, D., BANG, O. Y., BAXTER, B. W., BURNS, P. A., CHOE, H., HEO, J. H., HOLMSTEDT, C. A., JANKOWITZ, B., KELLY, M., LINARES, G., MANDZIA, J. L., SHANKAR, J., SOHN, S. I., SWARTZ, R. H., BARBER, P. A., COUTTS, S. B., SMITH, E. E., MORRISH, W. F., WEILL, A., SUBRAMANIAM, S., MITHA, A. P., WONG, J. H., LOWERISON, M. W., SAJOBI, T. T. & HILL, M. D. 2015. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*, 372, 1019-30.
- GOYAL, N., TSIVGOULIS, G., NICKELE, C., DOSS, V. T., HOIT, D., ALEXANDROV, A. V., ARTHUR, A. & ELIJOVICH, L. 2016. Posterior circulation CT angiography collaterals predict outcome of endovascular acute ischemic stroke therapy for basilar artery occlusion. *J Neurointerv Surg*, 8, 783-6.
- GREENLAND, S., MANSOURNIA, M. A. & ALTMAN, D. G. 2016. Sparse data bias: a problem hiding in plain sight. *BMJ*, 352, i1981.
- GREVING, J. P., SCHONEWILLE, W. J., WIJMAN, C. A., MICHEL, P., KAPPELLE, L. J. & ALGRA, A. 2012. Predicting outcome after acute basilar artery occlusion based on admission characteristics. *Neurology*, 78, 1058-63.
- GUNN, H. J., HAYATI REZVAN, P., FERNÁNDEZ, M. I. & COMULADA, W. S. 2022. How to apply variable selection machine learning algorithms with multiply imputed data: A missing discussion. *Psychol Methods*.
- HAMBURG, M. A. & COLLINS, F. S. 2010. The path to personalized medicine. *N Engl J Med*, 363, 301-4.

- HANS LIPPERT, R. P. 1985. *Arterial Variations in Man : Classification and Frequency*, Munich, Germany:, JF Bergmann Verlag.
- HARRELL, F. E., JR., LEE, K. L. & MARK, D. B. 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 15, 361-87.
- HAYATI REZVAN, P., LEE, K. J. & SIMPSON, J. A. 2015. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Medical Research Methodology*, 15, 30.
- HE, Y. 2010. Missing data analysis using multiple imputation: getting to the heart of the matter. *Circ Cardiovasc Qual Outcomes*, 3, 98-105.
- HEINZE, G. & SCHEMPER, M. 2002. A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21, 2409-2419.
- HEMINGWAY, H. 2006. Prognosis research: why is Dr. Lydgate still waiting? *J Clin Epidemiol*, 59, 1229-38.
- HEMINGWAY, H., CROFT, P., PEREL, P., HAYDEN, J. A., ABRAMS, K., TIMMIS, A., BRIGGS, A., UDUMYAN, R., MOONS, K. G., STEYERBERG, E. W., ROBERTS, I., SCHROTER, S., ALTMAN, D. G., RILEY, R. D. & GROUP, P. 2013. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*, 346, e5595.
- HERNÁNDEZ, A. V., STEYERBERG, E. W. & HABBEMA, J. D. F. 2004. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *Journal of Clinical Epidemiology*, 57, 454-460.
- HEYMANS, M. 2022. *psfmi: Prediction Model Pooling, Selection and Performance Evaluation Across Multiply Imputed Datasets*. R package version 1.1.0, [Online]. Available: <https://mweymans.github.io/psfmi/> [Accessed 18.01.2023 2023].
- HEYMANS, M. W., VAN BUUREN, S., KNOL, D. L., VAN MECHELEN, W. & DE VET, H. C. 2007. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Med Res Methodol*, 7, 33.
- HILDEN, J., HABBEMA, J. D. & BJERREGAARD, B. 1978. The measurement of performance in probabilistic diagnosis. II. Trustworthiness of the exact values of the diagnostic probabilities. *Methods Inf Med*, 17, 227-37.
- HINGORANI, A. D., WINDT, D. A. V. D., RILEY, R. D., ABRAMS, K., MOONS, K. G. M., STEYERBERG, E. W., SCHROTER, S., SAUERBREI, W., ALTMAN, D. G. & HEMINGWAY, H. 2013. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal*, 346, e5793.
- HINTON, G., DENG, L., YU, D., DAHL, G. E., MOHAMED, A. R., JAITLY, N., SENIOR, A., VANHOUCHE, V., NGUYEN, P., SAINATH, T. N. & KINGSBURY, B. 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, 29, 82-97.
- HOERL, A. E. & KENNARD, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55-67.
- HOLLÄNDER, N. & SAUERBREI, W. 2007. On Statistical Approaches for the Multivariable Analysis of Prognostic Marker Studies. In: AUGET, J.-L., BALAKRISHNAN, N., MESBAH, M. & MOLENBERGHS, G. (eds.) *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis*. Boston, MA: Birkhäuser Boston.

- HONG, J. M., CHOI, J. Y., LEE, J. H., YONG, S. W., BANG, O. Y., JOO, I. S. & HUH, K. 2009. Impact of posterior communicating artery on basilar artery stenotic occlusive disease. *J Neurol Neurosurg Psychiatry*, 80, 1390-3.
- HUDDA, M. T., FEWTRELL, M. S., HAROUN, D., LUM, S., WILLIAMS, J. E., WELLS, J. C. K., RILEY, R. D., OWEN, C. G., COOK, D. G., RUDNICKA, A. R., WHINCUP, P. H. & NIGHTINGALE, C. M. 2019. Development and validation of a prediction model for fat mass in children and adolescents: meta-analysis using individual participant data. *BMJ*, 366, l4293.
- INOA, V., ARON, A. W., STAFF, I., FORTUNATO, G. & SANSING, L. H. 2014. Lower NIH Stroke Scale Scores Are Required to Accurately Predict a Good Prognosis in Posterior Circulation Stroke. *Cerebrovascular Diseases*, 37, 251-255.
- ISHKANIAN, A. A., MCCULLOUGH-HICKS, M. E., APPELBOOM, G., PIAZZA, M. A., HWANG, B. Y., BRUCE, S. S., HANNAN, L. M., CONNOLLY, E. S., LAVINE, S. D. & MEYERS, P. M. 2011. Improving patient selection for endovascular treatment of acute cerebral ischemia: a review of the literature and an external validation of the Houston IAT and THRIVE predictive scoring systems. *Neurosurgical Focus FOC*, 30, E7.
- JAIN, A. K., JIANCHANG, M. & MOHIUDDIN, K. M. 1996. Artificial neural networks: a tutorial. *Computer*, 29, 31-44.
- JAKOBSEN, J. C., GLUUD, C., WETTERSLEV, J. & WINKEL, P. 2017a. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*, 17, 162.
- JAKOBSEN, J. C., GLUUD, C., WETTERSLEV, J. & WINKEL, P. 2017b. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Medical Research Methodology*, 17, 162.
- JARMAN, B., PIETER, D., VAN DER VEEN, A. A., KOOL, R. B., AYLIN, P., BOTTLE, A., WESTERT, G. P. & JONES, S. 2010. The hospital standardised mortality ratio: a powerful tool for Dutch hospitals to assess their quality of care? *Quality and Safety in Health Care*, 19, 9.
- JONES, T. H., MORAWETZ, R. B., CROWELL, R. M., MARCOUX, F. W., FITZGIBBON, S. J., DEGIROLAMI, U. & OJEMANN, R. G. 1981. Thresholds of focal cerebral ischemia in awake monkeys. *J Neurosurg*, 54, 773-82.
- JOVIN, T. G., CHAMORRO, A., COBO, E., DE MIQUEL, M. A., MOLINA, C. A., ROVIRA, A., SAN ROMÁN, L., SERENA, J., ABILLEIRA, S., RIBÓ, M., MILLÁN, M., URRÁ, X., CARDONA, P., LÓPEZ-CANCIO, E., TOMASELLO, A., CASTAÑO, C., BLASCO, J., AJA, L., DORADO, L., QUESADA, H., RUBIERA, M., HERNANDEZ-PÉREZ, M., GOYAL, M., DEMCHUK, A. M., VON KUMMER, R., GALLOFRÉ, M. & DÁVALOS, A. 2015. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med*, 372, 2296-306.
- JOVIN, T. G., LI, C., WU, L., WU, C., CHEN, J., JIANG, C., SHI, Z., GAO, Z., SONG, C., CHEN, W., PENG, Y., YAO, C., WEI, M., LI, T., WEI, L., XIAO, G., YANG, H., REN, M., DUAN, J., LIU, X., YANG, Q., LIU, Y., ZHU, Q., SHI, W., ZHU, Q., LI, X., GUO, Z., YANG, Q., HOU, C., ZHAO, W., MA, Q., ZHANG, Y., JIAO, L., ZHANG, H., LIEBESKIND, D. S., LIANG, H., JADHAV, A. P., WEN, C., BROWN, S., ZHU, L., YE, H., RIBO, M., CHANG, M., SONG, H., CHEN, J. & JI, X. 2022. Trial of Thrombectomy 6 to 24 Hours after Stroke Due to Basilar-Artery Occlusion. *New England Journal of Medicine*, 387, 1373-1384.
- JUNG, S., MONO, M. L., FISCHER, U., GALIMANIS, A., FINDLING, O., DE MARCHIS, G. M., WECK, A., NEDELTCHEV, K., COLUCCI, G., MORDASINI, P.,

- BREKENFELD, C., EL-KOUSSY, M., GRALLA, J., SCHROTH, G., MATTLE, H. P. & ARNOLD, M. 2011. Three-month and long-term outcomes and their predictors in acute basilar artery occlusion treated with intra-arterial thrombolysis. *Stroke*, 42, 1946-51.
- JUSTICE, A. C., COVINSKY, K. E. & BERLIN, J. A. 1999. Assessing the generalizability of prognostic information. *Ann Intern Med*, 130, 515-24.
- KANG, D. H., JUNG, C., YOON, W., KIM, S. K., BAEK, B. H., KIM, J. T., PARK, M. S., KIM, Y. W., HWANG, Y. H., KIM, Y. S., KIM, B. J., HAN, M. K. & BAE, H. J. 2018. Endovascular Thrombectomy for Acute Basilar Artery Occlusion: A Multicenter Retrospective Observational Study. *J Am Heart Assoc*, 7.
- KASHIWAGI, J., KIYOSUE, H., HORI, Y., OKAHARA, M., TANOUE, S., SAGARA, Y., ABE, T. & MORI, H. 2010. Endovascular recanalization of acute intracranial vertebrobasilar artery occlusion using local fibrinolysis and additional balloon angioplasty. *Neuroradiology*, 52, 361-70.
- KENT, P., CANCELLIERE, C., BOYLE, E., CASSIDY, J. D. & KONGSTED, A. 2020. A conceptual framework for prognostic research. *BMC Medical Research Methodology*, 20, 172.
- KHOURY, N. N., DARSAUT, T. E., GHOSTINE, J., DESCHAINTE, Y., DANEULT, N., DUROCHER, A., LANTHIER, S., POPPE, A. Y., ODIER, C., LEBRUN, L. H., GUILBERT, F., GENTRIC, J. C., BATISTA, A., WEILL, A., ROY, D., BRACARD, S. & RAYMOND, J. 2017. Endovascular thrombectomy and medical therapy versus medical therapy alone in acute stroke: A randomized care trial. *J Neuroradiol*, 44, 198-202.
- KLEINKE, K. 2017. Multiple Imputation Under Violated Distributional Assumptions: A Systematic Evaluation of the Assumed Robustness of Predictive Mean Matching. *Journal of Educational and Behavioral Statistics*, 42, 371-404.
- KLEINKE, K. 2018. Multiple imputation by predictive mean matching when sample size is small. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 14, 3-15.
- KNIEP, H., BECHSTEIN, M., BROOCKS, G., BREKENFELD, C., FLOTTMANN, F., VAN HORN, N., GEEST, V., FAIZY, T. D., DEB-CHATTERJI, M., ALEGIANI, A., THOMALLA, G., GELLIßEN, S., FIEHLER, J., HANNING, U. & MEYER, L. 2022a. Early surrogates of outcome after thrombectomy in posterior circulation stroke. *Eur J Neurol*, 29, 3296-3306.
- KNIEP, H. C., ELSAYED, S., NAWABI, J., BROOCKS, G., MEYER, L., BECHSTEIN, M., VAN HORN, N., PSYCHOGIOS, M., THOMALLA, G., FLOTTMANN, F., KEMMLING, A., GELLIßEN, S., FIEHLER, J., SPORNS, P. B. & HANNING, U. 2022b. Imaging-based outcome prediction in posterior circulation stroke. *Journal of Neurology*, 269, 3800-3809.
- KOHAVI, R. A. P., F. 1998. Glossary of terms. Machine Learning—Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Machine Learning*, 30, 271-274.
- KREMERS, F., VENEMA, E., DUVEKOT, M., YO, L., BOKKERS, R., NIJEHOLT, G. L. À., ES, A. V., LUGT, A. V. D., MAJOIE, C., BURKE, J., ROOZENBEEK, B., LINGSMA, H. & DIPPEL, D. 2022. Outcome Prediction Models for Endovascular Treatment of Ischemic Stroke: Systematic Review and External Validation. *Stroke*, 53, 825-836.
- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. 2012. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 84 - 90.

- KUBIK, C. S. & ADAMS, R. D. 1946. OCCLUSION OF THE BASILAR ARTERY—A CLINICAL AND PATHOLOGICAL STUDY. *Brain*, 69, 73-121.
- KYUNG, M., GILL, J., GHOSH, M. & CASELLA, G. 2010. Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, 5, 369-412.
- KYZAS, P. A., LOIZOU, K. T. & IOANNIDIS, J. P. 2005. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst*, 97, 1043-55.
- LANGEZAAL, L. C. M., VAN DER HOEVEN, E. J. R. J., MONT'ALVERNE, F. J. A., DE CARVALHO, J. J. F., LIMA, F. O., DIPPEL, D. W. J., VAN DER LUGT, A., LO, R. T. H., BOITEN, J., LYCKLAMA À NIJEHOLT, G. J., STAALS, J., VAN ZWAM, W. H., NEDERKOORN, P. J., MAJOIE, C. B. L. M., GERBER, J. C., MAZIGHI, M., PIOTIN, M., ZINI, A., VALLONE, S., HOFMEIJER, J., MARTINS, S. O., NOLTE, C. H., SZABO, K., DIAS, F. A., ABUD, D. G., WERMER, M. J. H., REMMERS, M. J. M., SCHNEIDER, H., RUECKERT, C. M., DE LAAT, K. F., YOO, A. J., VAN DOORMAAL, P.-J., VAN ES, A. C. G. M., EMMER, B. J., MICHEL, P., PUETZ, V., AUDEBERT, H. J., PONTES-NETO, O. M., VOS, J.-A., KAPPELLE, L. J., ALGRA, A. & SCHONEWILLE, W. J. 2021. Endovascular Therapy for Stroke Due to Basilar-Artery Occlusion. *New England Journal of Medicine*, 384, 1910-1920.
- LE CESSIE, S. & VAN HOUWELINGEN, J. C. 1992. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41, 191-201.
- LEE, B. I., NAM, H. S., HEO, J. H. & KIM, D. I. 2001. Yonsei Stroke Registry. Analysis of 1,000 patients with acute cerebral infarctions. *Cerebrovasc Dis*, 12, 145-51.
- LEE, J. H., HAN, S. J., YUN, Y. H., CHOI, H. C., JUNG, S., CHO, S. J., YU, K. H., LEE, S. M., HWANG, S. H., SONG, H. K., KWON, K. H. & LEE, B. C. 2006. Posterior circulation ischemic stroke in Korean population. *Eur J Neurol*, 13, 742-8.
- LEE, K. J. & CARLIN, J. B. 2010. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol*, 171, 624-32.
- LEE, K. J. & CARLIN, J. B. 2017. Multiple imputation in the presence of non-normal data. *Stat Med*, 36, 606-617.
- LEE, K. J., GALATI, J. C., SIMPSON, J. A. & CARLIN, J. B. 2012. Comparison of methods for imputing ordinal data using multivariate normal imputation: a case study of non-linear effects in a large cohort study. *Stat Med*, 31, 4164-74.
- LEE, Y. Y., YOON, W., KIM, S. K., BAEK, B. H., KIM, G. S., KIM, J. T. & PARK, M. S. 2017. Acute Basilar Artery Occlusion: Differences in Characteristics and Outcomes after Endovascular Therapy between Patients with and without Underlying Severe Atherosclerotic Stenosis. *AJNR Am J Neuroradiol*, 38, 1600-1604.
- LEGRAND, L., NAGGARA, O., TURC, G., MELLERIO, C., ROCA, P., CALVET, D., LABEYRIE, M. A., BARON, J. C., MAS, J. L., MEDER, J. F., TOUZÉ, E. & OPPENHEIM, C. 2013. Clot burden score on admission T2*-MRI predicts recanalization in acute stroke. *Stroke*, 44, 1878-84.
- LIEBESKIND, D. S. 2003. Collateral circulation. *Stroke*, 34, 2279-84.
- LIEW, S. L., ZAVALIANGOS-PETROPULU, A., JAHANSHAD, N., LANG, C. E., HAYWARD, K. S., LOHSE, K. R., JULIANO, J. M., ASSOGNA, F., BAUGH, L. A., BHATTACHARYA, A. K., BIGJAHAN, B., BORICH, M. R., BOYD, L. A., BRODTMANN, A., BUETEFISCH, C. M., BYBLOW, W. D., CASSIDY, J. M., CONFORTO, A. B., CRADDOCK, R. C., DIMYAN, M. A., DULA, A. N., ERMER, E., ETHERTON, M. R., FERCHO, K. A., GREGORY, C. M., HADIDCHI, S., HOLGUIN, J. A., HWANG, D. H., JUNG, S., KAUTZ, S. A., KHLIF, M. S.,

- KHOSHAB, N., KIM, B., KIM, H., KUCEYESKI, A., LOTZE, M., MACINTOSH, B. J., MARGETIS, J. L., MOHAMED, F. B., PIRAS, F., RAMOS-MURGUIALDAY, A., RICHARD, G., ROBERTS, P., ROBERTSON, A. D., RONDINA, J. M., ROST, N. S., SANOSSIAN, N., SCHWEIGHOFER, N., SEO, N. J., SHIROISHI, M. S., SOEKADAR, S. R., SPALLETTA, G., STINEAR, C. M., SURI, A., TANG, W. K. W., THIELMAN, G. T., VECCHIO, D., VILLRINGER, A., WARD, N. S., WERDEN, E., WESTLYE, L. T., WINSTEIN, C., WITTENBERG, G. F., WONG, K. A., YU, C., CRAMER, S. C. & THOMPSON, P. M. 2022. The ENIGMA Stroke Recovery Working Group: Big data neuroimaging to study brain-behavior relationships after stroke. *Hum Brain Mapp*, 43, 129-148.
- LIN, S. F., CHEN, C. I., HU, H. H. & BAI, C. H. 2018. Predicting functional outcomes of posterior circulation acute ischemic stroke in first 36 h of stroke onset. *J Neurol*, 265, 926-932.
- LIN, Y. H., CHEN, K. W., TANG, S. C. & LEE, C. W. 2020. Endovascular Treatment Outcome and CT Angiography Findings in Acute Basilar Artery Occlusion with and without Underlying Intracranial Atherosclerotic Stenosis. *J Vasc Interv Radiol*, 31, 747-753.
- LINDSBERG, P. J. & MATTLE, H. P. 2006. Therapy of basilar artery occlusion: a systematic analysis comparing intra-arterial and intravenous thrombolysis. *Stroke*, 37, 922-8.
- LINDSBERG, P. J., PEKKOLA, J., STRBIAN, D., SAIRANEN, T., MATTLE, H. P. & SCHROTH, G. 2015. Time window for recanalization in basilar artery occlusion: Speculative synthesis. *Neurology*, 85, 1806-15.
- LINGSMA, H. F., ROOZENBEEK, B., LI, B., LU, J., WEIR, J., BUTCHER, I., MARMAROU, A., MURRAY, G. D., MAAS, A. I. & STEYERBERG, E. W. 2011. Large between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study. *Neurosurgery*, 68, 601-7; discussion 607-8.
- LIU, X., DAI, Q., YE, R., ZI, W., LIU, Y., WANG, H., ZHU, W., MA, M., YIN, Q., LI, M., FAN, X., SUN, W., HAN, Y., LV, Q., LIU, R., YANG, D., SHI, Z., ZHENG, D., DENG, X., WAN, Y., WANG, Z., GENG, Y., CHEN, X., ZHOU, Z., LIAO, G., JIN, P., LIU, Y., LIU, X., ZHANG, M., ZHOU, F., SHI, H., ZHANG, Y., GUO, F., YIN, C., NIU, G., ZHANG, M., CAI, X., ZHU, Q., CHEN, Z., LIANG, Y., LI, B., LIN, M., WANG, W., XU, H., FU, X., LIU, W., TIAN, X., GONG, Z., SHI, H., WANG, C., LV, P., TAO, Z., ZHU, L., YANG, S., HU, W., JIANG, P., LIEBESKIND, D. S., PEREIRA, V. M., LEUNG, T., YAN, B., DAVIS, S., XU, G. & NOGUEIRA, R. G. 2020. Endovascular treatment versus standard medical treatment for vertebrobasilar artery occlusion (BEST): an open-label, randomised controlled trial. *Lancet Neurol*, 19, 115-122.
- LUTSEP, H. L., RYMER, M. M. & NESBIT, G. M. 2008. Vertebrobasilar revascularization rates and outcomes in the MERCI and multi-MERCI trials. *J Stroke Cerebrovasc Dis*, 17, 55-7.
- LYDEN, P. 2017. Using the National Institutes of Health Stroke Scale. *Stroke*, 48, 513-519.
- MADLEY-DOWD, P., HUGHES, R., TILLING, K. & HERON, J. 2019. The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63-73.
- MAKALIC, E. & SCHMIDT, D. F. Review of Modern Logistic Regression Methods with Application to Small and Medium Sample Size Problems. *In: LI, J., ed. AI 2010:*

- Advances in Artificial Intelligence, 2011// 2011 Berlin, Heidelberg. Springer Berlin Heidelberg, 213-222.
- MANSOURNIA, M. A., GEROLDINGER, A., GREENLAND, S. & HEINZE, G. 2018. Separation in Logistic Regression: Causes, Consequences, and Control. *American Journal of Epidemiology*, 187, 864-870.
- MANTATZIS, M. 2022. Editorial comment: new imaging score for outcome prediction in basilar artery occlusion stroke. *European Radiology*, 32, 4489-4490.
- MARSHALL, A., ALTMAN, D. G., HOLDER, R. L. & ROYSTON, P. 2009. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*, 9, 57.
- MATHERS, C. D. & LONCAR, D. 2006. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*, 3, e442.
- MEINEL, T. R., KAESMACHER, J., CHALOULOS-IAKOVIDIS, P., PANOS, L., MORDASINI, P., MOSIMANN, P. J., MICHEL, P., HAJDU, S., RIBO, M., REQUENA, M., MAEGERLEIN, C., FRIEDRICH, B., COSTALAT, V., BENALI, A., PIEROT, L., GAWLITZA, M., SCHAAFSMA, J., PEREIRA, V. M., GRALLA, J. & FISCHER, U. 2019. Mechanical thrombectomy for basilar artery occlusion: efficacy, outcomes, and futile recanalization in comparison with the anterior circulation. *J Neurointerv Surg*, 11, 1174-1180.
- MERWICK, Á. & WERRING, D. 2014. Posterior circulation ischaemic stroke. *BMJ : British Medical Journal*, 348, g3175.
- MEYER, J. S. & DENNY-BROWN, D. 1957. The cerebral collateral circulation. I. Factors influencing collateral blood flow. *Neurology*, 7, 447-58.
- MINNE, L., ESLAMI, S., DE KEIZER, N., DE JONGE, E., DE ROOIJ, S. E. & ABU-HANNA, A. 2012. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. *Intensive Care Medicine*, 38, 40-46.
- MINNERUP, J., BROOCKS, G., KALKOFFEN, J., LANGNER, S., KNAUTH, M., PSYCHOGIOS, M. N., WERSCHING, H., TEUBER, A., HEINDEL, W., ECKERT, B., WIENDL, H., SCHRAMM, P., FIEHLER, J. & KEMMLING, A. 2016. Computed tomography-based quantification of lesion water uptake identifies patients within 4.5 hours of stroke onset: A multicenter observational study. *Ann Neurol*, 80, 924-934.
- MITEFF, F., FAULDER, K. C., GOH, A. C., STEINFORT, B. S., SUE, C. & HARRINGTON, T. J. 2011. Mechanical thrombectomy with a self-expanding retrievable intracranial stent (Solitaire AB): experience in 26 patients with acute cerebral artery occlusion. *AJNR Am J Neuroradiol*, 32, 1078-81.
- MIYAMOTO, N., TANAKA, Y., UENO, Y., TANAKA, R., HATTORI, N. & URABE, T. 2010. Comparison of clinical backgrounds with anterior versus posterior circulation infarcts. *J Stroke Cerebrovasc Dis*, 19, 393-7.
- MOCCO, J., ZAIDAT, O. O., VON KUMMER, R., YOO, A. J., GUPTA, R., LOPES, D., FREI, D., SHOWNKEEN, H., BUDZIK, R., AJANI, Z. A., GROSSMAN, A., ALTSCHUL, D., MCDUGALL, C., BLAKE, L., FITZSIMMONS, B. F., YAVAGAL, D., TERRY, J., FARKAS, J., LEE, S. K., BAXTER, B., WIESMANN, M., KNAUTH, M., HECK, D., HUSSAIN, S., CHIU, D., ALEXANDER, M. J., MALISCH, T., KIRMANI, J., MISKOLCZI, L. & KHATRI, P. 2016. Aspiration Thrombectomy After Intravenous Alteplase Versus Intravenous Alteplase Alone. *Stroke*, 47, 2331-8.
- MOKIN, M., SONIG, A., SIVAKANTHAN, S., REN, Z., ELIJOVICH, L., ARTHUR, A., GOYAL, N., KAN, P., DUCKWORTH, E., VEZNEDAROGLU, E., BINNING, M. J., LIEBMAN, K. M., RAO, V., TURNER, R. D. T., TURK, A. S., BAXTER, B. W.,

- DABUS, G., LINFANTE, I., SNYDER, K. V., LEVY, E. I. & SIDDIQUI, A. H. 2016. Clinical and Procedural Predictors of Outcomes From the Endovascular Treatment of Posterior Circulation Strokes. *Stroke*, 47, 782-8.
- MOONS, K. G., ALTMAN, D. G., REITSMA, J. B., IOANNIDIS, J. P., MACASKILL, P., STEYERBERG, E. W., VICKERS, A. J., RANSOHOFF, D. F. & COLLINS, G. S. 2015. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*, 162, W1-73.
- MOONS, K. G., DONDERS, A. R., STEYERBERG, E. W. & HARRELL, F. E. 2004. Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: a clinical example. *J Clin Epidemiol*, 57, 1262-70.
- MOONS, K. G. M., ROYSTON, P., VERGOUWE, Y., GROBBEE, D. E. & ALTMAN, D. G. 2009. Prognosis and prognostic research: what, why, and how? *BMJ*, 338, b375.
- MORRIS, T. P., WHITE, I. R. & ROYSTON, P. 2014. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14, 75.
- MOULIN, T., TATU, L., CRÉPIN-LEBLOND, T., CHAVOT, D., BERGÈS, S. & RUMBACH, T. 1997. The Besançon Stroke Registry: an acute stroke registry of 2,500 consecutive patients. *Eur Neurol*, 38, 10-20.
- MUELLER DP, Y. W., CHANDRAN KB 1991. MRI arterial enhancement in acute ischemic stroke: a correlation of angiographic findings and in-vitro experience. *29th annual meeting of the American Society of Neuroradiology*.
- MUIR, K. W., FORD, G. A., MESSOW, C. M., FORD, I., MURRAY, A., CLIFTON, A., BROWN, M. M., MADIGAN, J., LENTHALL, R., ROBERTSON, F., DIXIT, A., CLOUD, G. C., WARDLAW, J., FREEMAN, J. & WHITE, P. 2017. Endovascular therapy for acute ischaemic stroke: the Pragmatic Ischaemic Stroke Thrombectomy Evaluation (PISTE) randomised, controlled trial. *J Neurol Neurosurg Psychiatry*, 88, 38-44.
- MUSORO, J. Z., ZWINDERMAN, A. H., PUHAN, M. A., TER RIET, G. & GESKUS, R. B. 2014. Validation of prediction models based on lasso regression with multiply imputed data. *BMC Medical Research Methodology*, 14, 116.
- NAGEL, S., SCHELLINGER, P. D., HARTMANN, M., JUETTLER, E., HUTTNER, H. B., RINGLEB, P., SCHWAB, S. & KÖHRMANN, M. 2009. Therapy of Acute Basilar Artery Occlusion. *Stroke*, 40, 140-146.
- NATEKIN, A. & KNOLL, A. 2013. Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7.
- O'DONNELL, M. J., XAVIER, D., LIU, L., ZHANG, H., CHIN, S. L., RAO-MELACINI, P., RANGARAJAN, S., ISLAM, S., PAIS, P., MCQUEEN, M. J., MONDO, C., DAMASCENO, A., LOPEZ-JARAMILLO, P., HANKEY, G. J., DANS, A. L., YUSOFF, K., TRUELSEN, T., DIENER, H. C., SACCO, R. L., RYGLEWICZ, D., CZLONKOWSKA, A., WEIMAR, C., WANG, X. & YUSUF, S. 2010. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet*, 376, 112-23.
- OBERMEYER, Z. 2016. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*, 375, 1216-9.
- OOMMEN, T., BAISE, L. G. & VOGEL, R. M. 2011. Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. *Mathematical Geosciences*, 43, 99-120.

- PANKEN, A. M. & HEYMANS, M. W. 2022. A simple pooling method for variable selection in multiply imputed datasets outperformed complex methods. *BMC Med Res Methodol*, 22, 214.
- PAUL, E., BAILEY, M., VAN LINT, A. & PILCHER, D. V. 2012. Performance of Apache III over Time in Australia and New Zealand: A Retrospective Cohort Study. *Anaesthesia and Intensive Care*, 40, 980-994.
- PAVLOU, M., AMBLER, G., SEAMAN, S., DE IORIO, M. & OMAR, R. Z. 2016. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Statistics in Medicine*, 35, 1159-1177.
- PAVLOU, M., AMBLER, G., SEAMAN, S. R., GUTTMANN, O., ELLIOTT, P., KING, M. & OMAR, R. Z. 2015. How to develop a more accurate risk prediction model when there are few events. *BMJ : British Medical Journal*, 351, h3868.
- PEDUZZI, P., CONCATO, J., FEINSTEIN, A. R. & HOLFORD, T. R. 1995. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, 48, 1503-1510.
- PEPE, M. & JANES, H. Methods for Evaluating Prediction Performance of Biomarkers and Tests. In: LEE, M.-L. T., GAIL, M., PFEIFFER, R., SATTEN, G., CAI, T. & GANDY, A., eds. Risk Assessment and Evaluation of Predictions, 2013// 2013 New York, NY. Springer New York, 107-142.
- PEREZ DE LA OSSA, N., RANGARAJU, S., JOVIN, T., DÁVALOS, A. & NULL, N. 2016. Abstract WP26: Validation of Predictive Scales of Functional Outcome After Endovascular Therapy Using the REVASCAT Data. *Stroke*, 47, AWP26-AWP26.
- PEXMAN, J. H., BARBER, P. A., HILL, M. D., SEVICK, R. J., DEMCHUK, A. M., HUDON, M. E., HU, W. Y. & BUCHAN, A. M. 2001. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. *AJNR Am J Neuroradiol*, 22, 1534-42.
- PFEFFERKORN, T., HOLTMANNSPÖTTER, M., SCHMIDT, C., BENDER, A., PFISTER, H. W., STRAUBE, A., MAYER, T. E., BRÜCKMANN, H., DICHGANS, M. & FESL, G. 2010. Drip, ship, and retrieve: cooperative recanalization therapy in acute basilar artery occlusion. *Stroke*, 41, 722-6.
- PHILIPP GAFFERT, F. M. A. V. B. 2016. Towards an MI-proper Predictive Mean Matching.
- PUETZ, V., DZIALOWSKI, I., HILL, M. D., SUBRAMANIAM, S., SYLAJA, P. N., KROL, A., O'REILLY, C., HUDON, M. E., HU, W. Y., COUTTS, S. B., BARBER, P. A., WATSON, T., ROY, J. & DEMCHUK, A. M. 2008a. Intracranial thrombus extent predicts clinical outcome, final infarct size and hemorrhagic transformation in ischemic stroke: the clot burden score. *Int J Stroke*, 3, 230-6.
- PUETZ, V., SYLAJA, P. N., COUTTS, S. B., HILL, M. D., DZIALOWSKI, I., MUELLER, P., BECKER, U., URBAN, G., O'REILLY, C., BARBER, P. A., SHARMA, P., GOYAL, M., GAHN, G., VON KUMMER, R. & DEMCHUK, A. M. 2008b. Extent of Hypoattenuation on CT Angiography Source Images Predicts Functional Outcome in Patients With Basilar Artery Occlusion. *Stroke*, 39, 2485-2490.
- PUHR, R., HEINZE, G., NOLD, M., LUSA, L. & GEROLDINGER, A. 2017. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Statistics in Medicine*, 36, 2302-2317.
- RAGHUNATHAN, T. E., LEPKOWSKI, J. M., HOEWYK, J. V. & SOLENBERGER, P. W. 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27, 85-95.

- RANGARAJU, S., FRANKEL, M. & JOVIN, T. G. 2016. Prognostic Value of the 24-Hour Neurological Examination in Anterior Circulation Ischemic Stroke: A post hoc Analysis of Two Randomized Controlled Stroke Trials. *Interv Neurol*, 4, 120-9.
- RANKIN, J. 1957. Cerebral Vascular Accidents in Patients over the Age of 60: II. Prognosis. *Scottish Medical Journal*, 2, 200-215.
- RAZA, S. A. & RANGARAJU, S. 2018. A Review of Pre-Intervention Prognostic Scores for Early Prognostication and Patient Selection in Endovascular Management of Large Vessel Occlusion Stroke. *Interventional Neurology*, 7, 171-181.
- RENARD, D., LANDRAGIN, N., ROBINSON, A., BRUNEL, H., BONAFE, A., HEROUM, C. & MILHAUD, C. 2008. MRI-based score for acute basilar artery thrombosis. *Cerebrovasc Dis*, 25, 511-6.
- RICH CARUANA, NIKOS KARAMPATZIAKIS & YESSENALINA, A. 2006. An Empirical Evaluation of Supervised Learning in High Dimensions.
- RILEY, R. D., ABRAMS, K. R., SUTTON, A. J., LAMBERT, P. C., JONES, D. R., HENEY, D. & BURCHILL, S. A. 2003. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *British Journal of Cancer*, 88, 1191-1198.
- RILEY, R. D., HAYDEN, J. A., STEYERBERG, E. W., MOONS, K. G., ABRAMS, K., KYZAS, P. A., MALATS, N., BRIGGS, A., SCHROTER, S., ALTMAN, D. G. & HEMINGWAY, H. 2013. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*, 10, e1001380.
- RILEY, R. D., SNELL, K. I., ENSOR, J., BURKE, D. L., HARRELL, F. E., JR., MOONS, K. G. & COLLINS, G. S. 2019a. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med*, 38, 1276-1296.
- RILEY, R. D., SNELL, K. I. E., ENSOR, J., BURKE, D. L., HARRELL JR, F. E., MOONS, K. G. M. & COLLINS, G. S. 2019b. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Statistics in Medicine*, 38, 1276-1296.
- RILEY, R. D., SNELL, K. I. E., MARTIN, G. P., WHITTLE, R., ARCHER, L., SPERRIN, M. & COLLINS, G. S. 2021. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *J Clin Epidemiol*, 132, 88-96.
- RILEY, R. D., VAN DER WINDT, D., CROFT, P. & MOONS, K. G. M. 2019c. *Prognosis Research in Healthcare: Concepts, Methods, and Impact*, Oxford University Press.
- ROBERTS, S. & NOWAK, G. 2014. Stabilizing the lasso against cross-validation variability. *Computational Statistics & Data Analysis*, 70, 198-211.
- ROOZENBEEK, B., MAAS, A. I., LINGSMA, H. F., BUTCHER, I., LU, J., MARMAROU, A., MCHUGH, G. S., WEIR, J., MURRAY, G. D. & STEYERBERG, E. W. 2009. Baseline characteristics and statistical power in randomized controlled trials: selection, prognostic targeting, or covariate adjustment? *Crit Care Med*, 37, 2683-90.
- ROTHWELL, P. M., MEHTA, Z., HOWARD, S. C., GUTNIKOV, S. A. & WARLOW, C. P. 2005. From subgroups to individuals: general principles and the example of carotid endarterectomy. *The Lancet*, 365, 256-265.
- ROYLE, K. L. & CAIRNS, D. A. 2021. The development and validation of prognostic models for overall survival in the presence of missing data in the training dataset: a strategy with a detailed example. *Diagn Progn Res*, 5, 14.
- RUBIN, D. B. 1987. *Multiple imputation for nonresponse in surveys*, New York ;, Wiley.

- RUBIN, D. B. 2004. *Multiple Imputation for Nonresponse in Surveys*, Wiley.
- SAIRANEN, T., STRBIAN, D., SOINNE, L., SILVENNOINEN, H., SALONEN, O., ARTTO, V., KOSKELA, I., HÄPPÖLÄ, O., KASTE, M. & LINDSBERG, P. J. 2011. Intravenous thrombolysis of basilar artery occlusion: predictors of recanalization and outcome. *Stroke*, 42, 2175-9.
- SALERNO, A., STRAMBO, D., NANNONI, S., DUNET, V. & MICHEL, P. 2022. Patterns of ischemic posterior circulation strokes: A clinical, anatomical, and radiological review. *International journal of stroke : official journal of the International Stroke Society*, 17, 714-722.
- SANG, H., LI, F., YUAN, J., LIU, S., LUO, W., WEN, C., ZHU, Q., CHEN, W., LIN, M., QI, L., ZHONG, Y., WANG, Z., LING, W., SHI, Z., CHEN, H., LIU, W., LIU, Z., YAO, X., XIONG, F., ZENG, G., HU, X., DONG, H., MAO, A., YANG, G., HUANG, J., CHEN, L., GONG, Z., TAO, J., LIU, H., WU, D., QIU, Z., YANG, Q., ZI, W. & LI, F. 2021. Values of Baseline Posterior Circulation Acute Stroke Prognosis Early Computed Tomography Score for Treatment Decision of Acute Basilar Artery Occlusion. *Stroke*, 52, 811-820.
- SATO, A., TAKAHASHI, S., SOMA, Y., ISHII, K., KIKUCHI, Y., WATANABE, T. & SAKAMOTO, K. 1991. Cerebral infarction: early detection by means of contrast-enhanced cerebral arteries at MR imaging. *Radiology*, 178, 433-9.
- SATO, S., TOYODA, K., UEHARA, T., TORATANI, N., YOKOTA, C., MORIWAKI, H., NARITOMI, H. & MINEMATSU, K. 2008. Baseline NIH Stroke Scale Score predicting outcome in anterior and posterior circulation strokes. *Neurology*, 70, 2371-7.
- SAUERBREI, W., ROYSTON, P. & BINDER, H. 2007. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, 26, 5512-5528.
- SAVER, J. L., GOYAL, M., BONAFE, A., DIENER, H. C., LEVY, E. I., PEREIRA, V. M., ALBERS, G. W., COGNARD, C., COHEN, D. J., HACKE, W., JANSEN, O., JOVIN, T. G., MATTLE, H. P., NOGUEIRA, R. G., SIDDIQUI, A. H., YAVAGAL, D. R., BAXTER, B. W., DEVLIN, T. G., LOPES, D. K., REDDY, V. K., DU MESNIL DE ROCHEMONT, R., SINGER, O. C. & JAHAN, R. 2015. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med*, 372, 2285-95.
- SCHAFFER, J. L. 1997. *Analysis of Incomplete Multivariate Data*, New York, Chapman and Hall/CRC.
- SCHAFFER, J. L. 1999. Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8, 3-15.
- SCHENKER, N. & TAYLOR, J. M. G. 1996. Partially parametric techniques for multiple imputation. *Computational Statistics & Data Analysis*, 22, 425-446.
- SCHIRMER, M. D., DALCA, A. V., SRIDHARAN, R., GIESE, A. K., DONAHUE, K. L., NARDIN, M. J., MOCKING, S. J. T., MCINTOSH, E. C., FRID, P., WASSELIUS, J., COLE, J. W., HOLMEGAARD, L., JERN, C., JIMENEZ-CONDE, J., LEMMENS, R., LINDGREN, A. G., MESCHIA, J. F., ROQUER, J., RUNDEK, T., SACCO, R. L., SCHMIDT, R., SHARMA, P., SLOWIK, A., THIJS, V., WOO, D., VAGAL, A., XU, H., KITTNER, S. J., MCARDLE, P. F., MITCHELL, B. D., ROSAND, J., WORRALL, B. B., WU, O., GOLLAND, P. & ROST, N. S. 2019. White matter hyperintensity quantification in large-scale clinical acute ischemic stroke cohorts - The MRI-GENIE study. *Neuroimage Clin*, 23, 101884.
- SCHONEWILLE, W. J., ALGRA, A., SERENA, J., MOLINA, C. A. & KAPPELLE, L. J. 2005. Outcome in patients with basilar artery occlusion treated conventionally. *J Neurol Neurosurg Psychiatry*, 76, 1238-41.

- SCHONEWILLE, W. J., WIJMAN, C. A., MICHEL, P., RUECKERT, C. M., WEIMAR, C., MATTLE, H. P., ENGELTER, S. T., TANNE, D., MUIR, K. W., MOLINA, C. A., THIJS, V., AUDEBERT, H., PFEFFERKORN, T., SZABO, K., LINDSBERG, P. J., DE FREITAS, G., KAPPELLE, L. J. & ALGRA, A. 2009. Treatment and outcomes of acute basilar artery occlusion in the Basilar Artery International Cooperation Study (BASICS): a prospective registry study. *Lancet Neurol*, 8, 724-30.
- SEKULA, P., MALLETT, S., ALTMAN, D. G. & SAUERBREI, W. 2017. Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. *PLoS One*, 12, e0178531.
- SILLANPAA, N., SAARINEN, J. T., RUSANEN, H., HAKOMAKI, J., LAHTEELA, A., NUMMINEN, H., ELOVAARA, I., DASTIDAR, P. & SOIMAKALLIO, S. 2012. The clot burden score, the Boston Acute Stroke Imaging Scale, the cerebral blood volume ASPECTS, and two novel imaging parameters in the prediction of clinical outcome of ischemic stroke patients receiving intravenous thrombolytic therapy. *Neuroradiology*, 54, 663-72.
- SIMARD, J. M., KENT, T. A., CHEN, M., TARASOV, K. V. & GERZANICH, V. 2007. Brain oedema in focal ischaemia: molecular pathophysiology and theoretical implications. *Lancet Neurol*, 6, 258-68.
- SIMON, N., FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. 2011. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw*, 39, 1-13.
- SIMON, R. & ALTMAN, D. G. 1994. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*, 69, 979-85.
- SINGER, O. C., BERKEFELD, J., NOLTE, C. H., BOHNER, G., HARING, H. P., TRENKLER, J., GRÖSCHEL, K., MÜLLER-FORELL, W., NIEDERKORN, K., DEUTSCHMANN, H., NEUMANN-HAEFELIN, T., HOHMANN, C., BUSSMEYER, M., MPOTSARIS, A., STOLL, A., BORMANN, A., BRECK, J., SCHLAMANN, M. U., JANDER, S., TUROWSKI, B., PETZOLD, G. C., URBACH, H. & LIEBESKIND, D. S. 2015. Mechanical recanalization in basilar artery occlusion: the ENDOSTROKE study. *Ann Neurol*, 77, 415-24.
- SINGH, A., THAKUR, N. & SHARMA, A. A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 16-18 March 2016 2016. 1310-1315.
- ŠINKOVEC, H., HEINZE, G., BLAGUS, R. & GEROLDINGER, A. 2021. To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. *BMC Medical Research Methodology*, 21, 199.
- SJOBERG, D. 2022. *dcurves: Decision Curve Analysis for Model Evaluation* [Online]. Available: <https://github.com/ddsjoberg/dcurves>, <https://www.danieldsjoberg.com/dcurves/> [Accessed 18.01.2023 2023].
- SOIZE, S., BATISTA, A. L., RODRIGUEZ REGENT, C., TRYSTRAM, D., TISSERAND, M., TURC, G., SERRE, I., BEN HASSEN, W., ZUBER, M., CALVET, D., MAS, J. L., MEDER, J. F., RAYMOND, J., PIEROT, L., OPPENHEIM, C. & NAGGARA, O. 2015. Susceptibility vessel sign on T2* magnetic resonance imaging and recanalization results of mechanical thrombectomy with stent retrievers: a multicentre cohort study. *Eur J Neurol*, 22, 967-72.
- STEYERBERG, E. W. 2018. Validation in prediction research: the waste by data splitting. *Journal of Clinical Epidemiology*, 103, 131-133.

- STEYERBERG, E. W., EIJKEMANS, M. J., HARRELL, F. E., JR. & HABBEMA, J. D. 2001a. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making*, 21, 45-56.
- STEYERBERG, E. W., EIJKEMANS, M. J. C. & HABBEMA, J. D. F. 2001b. Application of Shrinkage Techniques in Logistic Regression Analysis: A Case Study. *Statistica Neerlandica*, 55, 76-88.
- STEYERBERG, E. W., MOONS, K. G., VAN DER WINDT, D. A., HAYDEN, J. A., PEREL, P., SCHROTER, S., RILEY, R. D., HEMINGWAY, H. & ALTMAN, D. G. 2013. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*, 10, e1001381.
- STEYERBERG, E. W., UNO, H., IOANNIDIS, J. P. A. & VAN CALSTER, B. 2018. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol*, 98, 133-143.
- STEYERBERG, E. W., VICKERS, A. J., COOK, N. R., GERDS, T., GONEN, M., OBUCHOWSKI, N., PENCINA, M. J. & KATTAN, M. W. 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*, 21, 128-38.
- SUBRAMANIAN, G., SILVA, J., SILVER, F. L., FANG, J., KAPRAL, M. K., OCZKOWSKI, W., GOULD, L. & O'DONNELL, M. J. 2009. Risk Factors for Posterior Compared to Anterior Ischemic Stroke: An Observational Study of the Registry of the Canadian Stroke Network. *Neuroepidemiology*, 33, 12-16.
- TAN, I. Y., DEMCHUK, A. M., HOPYAN, J., ZHANG, L., GLADSTONE, D., WONG, K., MARTIN, M., SYMONS, S. P., FOX, A. J. & AVIV, R. I. 2009. CT angiography clot burden score and collateral score: correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct. *AJNR Am J Neuroradiol*, 30, 525-31.
- THAO, L. T. P. & GESKUS, R. 2019. A comparison of model selection methods for prediction in the presence of multiply imputed data. *Biometrical Journal*, 61, 343-356.
- TIBSHIRANI, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267-288.
- TILLING, K., WILLIAMSON, E. J., SPRATT, M., STERNE, J. A. C. & CARPENTER, J. R. 2016. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *Journal of Clinical Epidemiology*, 80, 107-115.
- TRUSHEIM, M. R., BERNDT, E. R. & DOUGLAS, F. L. 2007. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov*, 6, 287-93.
- VACH, W. 2013. Calibration of clinical prediction rules does not just assess bias. *Journal of Clinical Epidemiology*, 66, 1296-1301.
- VAN BUUREN, S. 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res*, 16, 219-42.
- VAN BUUREN, S. 2018. *Flexible Imputation of Missing Data, Second Edition*.
- VAN BUUREN, S., BOSHUIZEN, H. C. & KNOOK, D. L. 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*, 18, 681-94.
- VAN CALSTER, B., MCLERNON, D. J., VAN SMEDEN, M., WYNANTS, L., STEYERBERG, E. W., BOSSUYT, P., COLLINS, G. S., MACASKILL, P., MCLERNON, D. J., MOONS, K. G. M., STEYERBERG, E. W., VAN CALSTER, B., VAN SMEDEN, M., VICKERS, ANDREW J., ON BEHALF OF TOPIC GROUP 'EVALUATING DIAGNOSTIC, T. & PREDICTION MODELS' OF THE, S. I. 2019. Calibration: the Achilles heel of predictive analytics. *BMC Medicine*, 17, 230.

- VAN CALSTER, B., NIEBOER, D., VERGOUWE, Y., DE COCK, B., PENCINA, M. J. & STEYERBERG, E. W. 2016a. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*, 74, 167-76.
- VAN CALSTER, B., NIEBOER, D., VERGOUWE, Y., DE COCK, B., PENCINA, M. J. & STEYERBERG, E. W. 2016b. A calibration hierarchy for risk models was defined: from utopia to empirical data. *Journal of Clinical Epidemiology*, 74, 167-176.
- VAN CALSTER, B., VAN SMEDEN, M., DE COCK, B. & STEYERBERG, E. W. 2020. Regression shrinkage methods for clinical prediction models do not guarantee improved performance: Simulation study. *Stat Methods Med Res*, 29, 3166-3178.
- VAN CALSTER, B. & VICKERS, A. J. 2015. Calibration of risk prediction models: impact on decision-analytic performance. *Med Decis Making*, 35, 162-9.
- VAN DE WIEL, M. A., BERKHOF, J. & VAN WIERINGEN, W. N. 2009. Testing the prediction error difference between 2 predictors. *Biostatistics*, 10, 550-60.
- VAN DER HOEVEN, E. J., MCVERRY, F., VOS, J. A., ALGRA, A., PUETZ, V., KAPPELLE, L. J. & SCHONEWILLE, W. J. 2016a. Collateral flow predicts outcome after basilar artery occlusion: The posterior circulation collateral score. *Int J Stroke*, 11, 768-75.
- VAN DER HOEVEN, E. J. R. J., MCVERRY, F., VOS, J. A., ALGRA, A., PUETZ, V., KAPPELLE, L. J. & SCHONEWILLE, W. J. 2016b. Collateral flow predicts outcome after basilar artery occlusion: The posterior circulation collateral score. *International Journal of Stroke*, 11, 768-775.
- VAN HOUWELINGEN, J. C. 2001. Shrinkage and Penalized Likelihood as Methods to Improve Predictive Accuracy. *Statistica Neerlandica*, 55, 17-34.
- VAN SMEDEN, M., MOONS, K. G., DE GROOT, J. A., COLLINS, G. S., ALTMAN, D. G., EIJKEMANS, M. J. & REITSMA, J. B. 2019. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res*, 28, 2455-2474.
- VICKERS, A. J. & CRONIN, A. M. 2010. Everything You Always Wanted to Know About Evaluating Prediction Models (But Were Too Afraid to Ask). *Urology*, 76, 1298-1301.
- VICKERS, A. J. & ELKIN, E. B. 2006. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*, 26, 565-74.
- VICKERS, A. J., VAN CALSTER, B. & STEYERBERG, E. W. 2019. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research*, 3, 18.
- VINK, G., FRANK, L. E., PANNEKOEK, J. & VAN BUUREN, S. 2014. Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, 68, 61-90.
- VOETSCH, B., DEWITT, L. D., PESSIN, M. S. & CAPLAN, L. R. 2004. Basilar artery occlusive disease in the New England Medical Center Posterior Circulation Registry. *Arch Neurol*, 61, 496-504.
- WAHL, S., BOULESTEIX, A.-L., ZIERER, A., THORAND, B. & VAN DE WIEL, M. A. 2016. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. *BMC Medical Research Methodology*, 16, 144.
- WALLISCH, C., BACH, P., HAFERMANN, L., KLEIN, N., SAUERBREI, W., STEYERBERG, E. W., HEINZE, G. & RAUCH, G. 2022. Review of guidance papers on regression modeling in statistical series of medical journals. *PLoS One*, 17, e0262918.

- WANG, F., MUKHERJEE, S., RICHARDSON, S. & HILL, S. M. 2020. High-dimensional regression in practice: an empirical study of finite-sample prediction, variable selection and ranking. *Statistics and Computing*, 30, 697-719.
- WEAVER, N. A., KUIJF, H. J., ABEN, H. P., ABRIGO, J., BAE, H. J., BARBAY, M., BEST, J. G., BORDET, R., CHAPPELL, F. M., CHEN, C., DONDAINE, T., VAN DER GIESSEN, R. S., GODEFROY, O., GYANWALI, B., HAMILTON, O. K. L., HILAL, S., HUENGES WAJER, I. M. C., KANG, Y., KAPPELLE, L. J., KIM, B. J., KÖHLER, S., DE KORT, P. L. M., KOUDSTAAL, P. J., KUCHCINSKI, G., LAM, B. Y. K., LEE, B. C., LEE, K. J., LIM, J. S., LOPES, R., MAKIN, S. D. J., MENDYK, A. M., MOK, V. C. T., OH, M. S., VAN OOSTENBRUGGE, R. J., ROUSSEL, M., SHI, L., STAALS, J., DEL, C. V.-H. M., VENKETASUBRAMANIAN, N., VERHEY, F. R. J., WARDLAW, J. M., WERRING, D. J., XIN, X., YU, K. H., VAN ZANDVOORT, M. J. E., ZHAO, L., BIESBROEK, J. M. & BIESELS, G. J. 2021. Strategic infarct locations for post-stroke cognitive impairment: a pooled analysis of individual patient data from 12 acute ischaemic stroke cohorts. *Lancet Neurol*, 20, 448-459.
- WEI, L., ERINJERI, J. P., ROVAINEN, C. M. & WOOLSEY, T. A. 2001. Collateral Growth and Angiogenesis Around Cortical Stroke. *Stroke*, 32, 2179-2184.
- WESSLER, B. S., YH, L. L., KRAMER, W., CANGELOSI, M., RAMAN, G., LUTZ, J. S. & KENT, D. M. 2015. Clinical Prediction Models for Cardiovascular Disease. *Circulation: Cardiovascular Quality and Outcomes*, 8, 368-375.
- WHITE, I. R., ROYSTON, P. & WOOD, A. M. 2011. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30, 377-399.
- WHITE, I. R. & THOMPSON, S. G. 2005. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med*, 24, 993-1007.
- WHITEPAPER 2010. Equity and excellence: liberating the NHS *In: HEALTH.*, D. O. (ed.). Stationery Office.
- WOOD, A. M., WHITE, I. R. & ROYSTON, P. 2008. How should variable selection be performed with multiply imputed data? *Stat Med*, 27, 3227-46.
- WU, L., RAJAH, G. B., COSKY, E. E., WU, X., LI, C., CHEN, J., ZHAO, W., WU, D., DING, Y. & JI, X. 2021. Outcomes in Endovascular Therapy for Basilar Artery Occlusion: Intracranial Atherosclerotic Disease vs. Embolism. *Aging Dis*, 12, 404-414.
- XIAO-LI, M. 1994. Multiple-Imputation Inferences with Uncongenial Sources of Input. *Statistical Science*, 9, 538-558.
- XU, J., CHEN, X., CHEN, S., CAO, W., ZHAO, H., NI, W., ZHANG, Y., GAO, C., GU, Y., CHENG, X., DONG, Y. & DONG, Q. 2022. Endovascular treatment for basilar artery occlusion: a meta-analysis. *Stroke and Vascular Neurology*, svn-2022-001740.
- XUN, K., MO, J., RUAN, S., DAI, J., ZHANG, W., LV, Y., DU, N., CHEN, S., SHEN, Z. & WU, Y. 2021. A Meta-Analysis of Prognostic Factors in Patients with Posterior Circulation Stroke after Mechanical Thrombectomy. *Cerebrovascular Diseases*, 50, 185-199.
- YAN, S., CHEN, Q., XU, M., SUN, J., LIEBESKIND, D. S. & LOU, M. 2016. Thrombus Length Estimation on Delayed Gadolinium-Enhanced T1. *Stroke*, 47, 756-761.
- YE, J. 1998. On Measuring and Correcting the Effects of Data Mining and Model Selection. *Journal of the American Statistical Association*, 93, 120-131.
- YU, Y. Y., NIU, L., GAO, L., ZHAO, Z. W., DENG, J. P., QU, Y. Z., JIAO, D. R., YANG, J. Q. & GAO, G. D. 2010. Intraarterial thrombolysis and stent placement for acute basilar artery occlusion. *J Vasc Interv Radiol*, 21, 1359-63.

- YUCEL, R. M., HE, Y. & ZASLAVSKY, A. M. 2011. Gaussian-based routines to impute categorical variables in health surveys. *Stat Med*, 30, 3447-60.
- YUH, W. T., UEDA, T., WHITE, M., SCHUSTER, M. E. & TAOKA, T. 1999. The need for objective assessment of the new imaging techniques and understanding the expanding roles of stroke imaging. *AJNR. American journal of neuroradiology*, 20, 1779-1784.
- YUH WTC, C. M. 1992. Magnetic resonance imaging of acute cerebral ischemia. *Neuroimaging ClinNorth Am*, 421–427.
- ZENG, M., OAKDEN-RAYNER, L., BIRD, A., SMITH, L., WU, Z., SCROOP, R., KLEINIG, T., JANNES, J., JENKINSON, M. & PALMER, L. J. 2022. Pre-thrombectomy prognostic prediction of large-vessel ischemic stroke using machine learning: A systematic review and meta-analysis. *Front Neurol*, 13, 945813.
- ZENG, Q., TAO, W., LEI, C., DONG, W. & LIU, M. 2015. Etiology and Risk Factors of Posterior Circulation Infarction Compared with Anterior Circulation Infarction. *J Stroke Cerebrovasc Dis*, 24, 1614-20.
- ZI, W., QIU, Z., WU, D., LI, F., LIU, H., LIU, W., HUANG, W., SHI, Z., BAI, Y., LIU, Z., WANG, L., YANG, S., PU, J., WEN, C., WANG, S., ZHU, Q., CHEN, W., YIN, C., LIN, M., QI, L., ZHONG, Y., WANG, Z., WU, W., CHEN, H., YAO, X., XIONG, F., ZENG, G., ZHOU, Z., WU, Z., WAN, Y., PENG, H., LI, B., HU, X., WEN, H., ZHONG, W., WANG, L., JIN, P., GUO, F., HAN, J., FU, X., AI, Z., TIAN, X., FENG, X., SUN, B., HUANG, Z., LI, W., ZHOU, P., TU, M., SUN, X., LI, H., HE, W., QIU, T., YUAN, Z., YUE, C., YANG, J., LUO, W., GONG, Z., SHUAI, J., NOGUEIRA, R. G. & YANG, Q. 2020. Assessment of Endovascular Treatment for Acute Basilar Artery Occlusion via a Nationwide Prospective Registry. *JAMA Neurol*, 77, 561-573.
- ZOU, H. & HASTIE, T. 2005a. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301-320.
- ZOU, H. & HASTIE, T. 2005b. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67, 301-320.

1.10 Danksagung

Meinen Dank möchte ich insbesondere meinem Doktorarbeitsbetreuer PD Dr. Gabriel Broocks und meiner Doktormutter PD Dr. Uta Hanning aussprechen. Besonders die Vorarbeit, die PD Dr. Hanning und PD Dr. Broocks in bisherigen Studien geleistet haben, ermöglichte es diese Arbeit zu verfassen. Dank gebührt außerdem der Klinik für Neuroradiologische Diagnostik und Intervention des UKE, welche die klinischen und paraklinischen Daten zur Verfügung stellte.

1.11 Lebenslauf

Ausbildung

06/2019 – 12/2022	Promotion in der neuroradiologischen Klinik, UKE. „Optimierte Modellentwicklung zur stratifizierten Prognoseabschätzung bei rekanalisiertem, posteriorem Apoplex in einem „Real World“ Datensatz mittels TRIPOD konformer Algorithmen – das pcBATASs – PROGRESS Typ 3 Prognosemodell.“ (Dissertationsschrift eingereicht)
10/2014 – 12/2021	Studium der Humanmedizin an der Universität Hamburg
10/2013 - 09/2014	Ausbildung zum Physiotherapeuten an der Charité Gesundheitsakademie
10/2011 – 10/2013	B. Sc. Biochemie Georg - August - Universität Göttingen Schwerpunkt Grundlagenstudium (Biologie, Chemie, Physik, höhere Mathematik)
07/2008 - 06/2011	Carl – Friedrich – von – Siemens- Gymnasium Berlin Abschluss: Allgemeine Hochschulreife

Publikationen

01/2022	Broocks G, Groffmann M, Meyer L., Elsayed S., Kniep H., Kemmling A., van Horn N., McDonough R., Faizy T., Bechstein M., Sporns P., Rusche T., Schön G., Nawabi J., Fiehler J., Hanning U. “Higher admission blood glucose reduces likelihood for successful recanalization in patients with acute basilar artery occlusion” in Journal of Neurology
10/2021	Broocks G, Faizy T, Meyer L., Groffmann M, Elsayed S., Kniep H., Kemmling A., van Horn N., McDonough R., Faizy T., Bechstein M., Sporns P., Rusche T., Schön G., Nawabi J., Fiehler J. “Posterior circulation collateral flow modifies the effect of thrombectomy on outcome in acute basilar artery occlusion” in International Journal of Stroke

Kongressbeiträge und wissenschaftliche Projekte

04/2021	Podcast-Projekt „MDM-Training in Rare Diseases“ im Rahmen der Studierendenklinik am MZCSE;
---------	--

finanziert durch den „Claussen-Simons-Fonds“ für
Wissenschaft & Hochschule

10/2019

54. Jahrestagung der Deutschen Gesellschaft für
Neuroradiologie e.V. “Estimation of Lesion Age in
Stroke Patients – Direct Comparison of non –
contrast enhanced computed tomography – to DWI
– FLAIR Mismatch”. Posterbegehung “Stroke
Imaging”. Vorsitz: Prof. Dr. B. Turowski.

Berufliche Laufbahn

Seit 02/2022-05/2022

Impftätigkeit und Erstuntersuchungen für
Geflüchtete für das DRK Bremen

03/2019–01/2022

Studentische Hilfskraft CTC North, Betreuung
klinischer Phase-1-Studien

04/2018 – 01/2022

Wissenschaftlicher Mitarbeiter Martin Zeitz Centrum
für Seltene Erkrankungen, UKE, Betreuung des
Projekts „tNAMSE“ und „Studierendenklinik“

12/2016 – 12/2017

Werksstudent der Universität Hamburg,
Hochschulsport Hamburg

06/2015 – 02/2019

„Medical Assistant“ OrionPharma, Datenbankpflege
und wissenschaftliche Literaturrecherche, Abteilung
Simdax (Levosimendan)

Praktische Erfahrungen

11/2020 – 10/2021

Praktisches Jahr

08/2014 – 09/2014

Physiotherapie - Praktikum Klinik für Rheumatologie
Campus Charité Berlin - Mitte

03/2014 – 04/2014

Physiotherapie - Pflichtpraktikum
Orthopädie/Chirurgie in Reha - Tagesklinik Berlin -
Pankow

07/2011 – 08/2011

Praktikum auf der Allg. - und Viszeralchirurgie des
Marienhospitals Aachen

1.12 Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: 