

Abschlussbericht

Forschungsprojekt:

Apoplex Prädikt - Anwendung von multiparametrischen Methoden des maschinellen Lernens zur Identifikation von Prädiktoren von Schlaganfällen und Schlaganfallrezidiven in medizinischen Routinedaten

Projekt: 02/16

Laufzeit: 01.10.2016 – 30.06.2018

Professor Dr. med. Jochen Schmitt, MPH, Zentrumsdirektor¹

PD Dr. med. Jesscia Barlinn, FÄ Neurologie²

Dr. rer. nat. Thomas Datzmann, Biometriker¹

¹Zentrum für evidenzbasierte Gesundheitsversorgung, Universitätsklinikum Carl Gustav Carus Dresden und TU Dresden

²Klinik und Poliklinik für Neurologie, Universitätsklinikum Carl Gustav Carus Dresden

Telefon: 0351 458 6495, Mail: Jochen.Schmitt@uniklinikum-dresden.de

Zusammenfassung

Im vorliegenden Projekt konnten neue Erkenntnisse (1) zur Epidemiologie des Schlaganfalls und –Rezidivs in Sachsen, (2) zu regionalen Unterschieden mit Hinweisen auf mögliche Unterversorgung, sowie (3) methodisch zur Analyse von GKV-Daten mittels Methoden des maschinellen Lernens gefunden werden. Es konnten aktuelle Zahlen zur Schlaganfallhäufigkeit, Rezidivrate, der schlaganfall-bezogene Mortalität in Sachsen, sowie Risikofaktorenkonstellationen bestimmt werden. Besonders hervorzuheben sind hier auch die altersstratifizierten Analysen zu Inzidenz und Mortalität, in denen auch Versicherte im Alter über 85 Jahren noch berücksichtigt wurden. Die vorliegende Arbeit stellt damit erstmals, unseres Wissens nach, diese epidemiologischen Maßzahlen auch für sehr alte Menschen dar. Dabei ist z.B. aufgefallen, dass beim ischämischen Schlaganfall zwar in den Altersstufen unter 80 Jahren stets die Inzidenz bei Männern höher ist, jedoch ältere Frauen häufiger einen Schlaganfall bekommen. Insgesamt konnten 4759 Patienten mit hämorrhagischem Schlaganfall (I61, I62), 29382 Patienten mit ischämischem S. (I63, I64) und 13253 Patienten mit einer TIA(G45) identifiziert werden. Davon hatten 716 Patienten ein Rezidiv nach hämorrhagischem S. (15,1%), 3686 Patienten ein Rezidiv nach ischämischem S. (12,6%) und 1095 Patienten ein Rezidiv nach TIA (8,3%) im Beobachtungszeitraum. Während beim hämorrhagischen S. Männer etwas häufiger betroffen sind (17% gegen 13%) und die Rezidivhäufigkeit mit dem Alter stark abnimmt (45-49 – 25% auf 85+ - unter 6%), haben beim ischämischen S. und bei der TIA beide Geschlechter gleich häufig ein Rezidiv und die Häufigkeit bleibt über alle untersuchten Altersgruppen nahezu identisch. Nur in der höchsten Altersgruppe 85+ (Tab.1) sinkt beim ischämischen S. auch die Rezidivhäufigkeit, was wiederum durch eine Unterdiagnostik leichter Fälle bei sehr alten Menschen bedingt sein kann. Diese epidemiologischen Zahlen konnten erstmals für die Region Sachsen in dieser umfassenden Weise erhoben werden.

Die regionale Analyse der Inzidenzen und der Schlaganfall-bezogenen Mortalität getrennt nach 3-stelligen Postleitzahlen ergab eine leicht inhomogene Verteilung des ischämischen Schlaganfalls innerhalb Sachsens. Die gefundene Verteilung mit niedrigeren Werten (Inzidenz & Mortalität) für die Metropolregionen im Vergleich zu ländlichen Regionen könnte einmal für vorhandene regionale Versorgungsunterschiede akuter Schlaganfallpatienten, bzw. eine regional unterschiedliche Wirksamkeit von Präventivmaßnahmen sprechen.

Mit Hilfe verschiedener Methoden des maschinellen Lernens konnte nun überprüft werden, ob statistisch auffällige Korrelationen zwischen Schlaganfallereignis bzw. -Rezidiv und medikamentösen Therapien sowie vorhandenen Komorbiditäten in den Daten nachweisbar sind. Zunächst konnten die bekannten Risikofaktoren für inzidente Schlaganfälle im Datensatz repliziert werden, was die Validität der Studie stützt und Voraussetzung für die explorativen Data Mining Analysen darstellt. Univariate Analysen der häufigsten vaskulären Risikofaktoren, aber auch weiterer assoziierter Komorbiditäten, zeigten die höchsten Risiken für folgende bekannte Risikofaktoren (adjustiert für Alter und Geschlecht): Hypertonie (OR 4,39) und

Vorhofflimmern (OR 4,41) für ischämischen und hämorrhagischen Schlaganfall und TIA, zusätzlich schädlicher Alkoholgebrauch (OR 5,01) für den hämorrhagischen Schlaganfall. Interessanterweise zeigte sich bei einigen dieser Risikofaktoren eine erniedrigte Mortalitätsrate, sodass weitere Analysen des metabolischen Syndroms als Risikofaktor für Inzidenz sowie Mortalität erfolgten. Die gefundenen doch eher paradoxen Beziehungen zwischen erhöhtem Erkrankungsrisiko bei Vorliegen einer Erkrankung aus diesem Symptomenkomplex und einer erniedrigten Mortalität, lassen sich am plausibelsten mit einer Unterdiagnostik dieser Patientengruppe im Vorfeld erklären. Das Vorliegen einer Hypertonie hatte in allen Modellen den stärksten Einfluss auf das Rezidivrisiko. Die Einnahme von Medikamenten konnte bei sieben Präparaten als Risikofaktor für das ischämische Schlaganfallrezidiv belegt werden, wobei vier Präparate davon bei Vorliegen einer Hypertonie als Präventivfaktoren wirken und das Rezidivrisiko, zwar nur geringfügig, senken können.

Zusammenfassend konnte mit der vorliegenden Arbeit gezeigt werden, dass sich mit multiparametrischen Methoden des maschinellen Lernens Prädiktoren von Schlaganfällen und Schlaganfallrezidiven in medizinischen Routinedaten identifizieren lassen, jedoch die Hinzunahme weiterer komplementärer Datenquellen nötig ist, um verlässliche Vorhersagemodelle für das Gros der Bevölkerung aufstellen zu können.

1. Forschungsvorhaben

Obwohl die Mortalität des akuten Schlaganfalls in den vergangenen Jahren aufgrund von therapeutischen Erfolgen deutlich reduziert werden konnte (1), stellt diese Erkrankung in den westlichen Industrienationen derzeit die vierthäufigste Ursache für Tod dar. In Deutschland ist der akute Schlaganfall die häufigste Ursache für dauerhafte Invalidität im Erwachsenenalter, wodurch sich abgesehen von individuellen Folgen auch erhebliche gesundheitsökonomische Belastungen ergeben.

Einige Risikofaktoren von Schlaganfällen sind aus epidemiologischer Forschung bekannt und prinzipiell durch Präventivmaßnahmen beeinflussbar. So sind beispielsweise einzelne vaskuläre Risikofaktoren wie arterielle Hypertonie, Diabetes mellitus, Hypercholesterinämie oder auch das Schlafapnoe Syndrom als unabhängige Risikofaktoren für einen Schlaganfall bekannt und weitestgehend gut untersucht, letztlich ist deren Einfluss auf die Erkrankung jedoch in der Zusammenschau untereinander oder in Verbindung mit weiteren Erkrankungen, oder z.B. Verschreibungsverhalten von Medikamenten unerforscht. Gerade hier vermuten die Antragssteller durch das Auflösen von komplexen Zusammenhängen wichtige und bisher ungeklärte Zusammenhänge, bei denen in der Zukunft ein hohes Potenzial für primär oder sekundär präventive Therapieansätze in der Zukunft abgeleitet und genutzt werden können.

Im Rahmen des geplanten Forschungsprojekts sollen neue Data Mining/Analysis Methoden bei der wissenschaftlichen Analyse von Routinedaten der medizinischen Regelversorgung genutzt werden, um neue

Erkenntnisse zum komplexen Zusammenspiel bekannter und unbekannter Risikofaktoren und Präventivfaktoren von Schlaganfällen und Schlaganfallrezidiven zu gewinnen.

2. Fragestellung und Ziel des Forschungsvorhabens

Das Ziel der Studie war die wissenschaftliche Nutzung von BigData Methoden zur Erkennung neuer Risikofaktoren und Risikofaktor-Konstellationen für den Schlaganfall und das Schlaganfallrezidiv in medizinischen Routinedaten einer Kohorte von rund 2,1 Millionen Personen und einem Beobachtungszeitraum von 7 Jahren.

Speziell sollten durch die innovative Methode neue, komplexe Zusammenhänge von Vorerkrankungen, medikamentösen Therapien, der Inanspruchnahme von Gesundheitsleistungen und individuellen Patientencharakteristika bezüglich des Risikos eines Schlaganfalls und eines Schlaganfallrezidivs identifiziert werden.

3. Methodik

3.1 Untersuchungskohorte

Datenbasis der Fall-Kontroll Studie bildeten pseudonymisierte Routinedaten von 2,1 Millionen sächsischen Versicherten der AOK PLUS über den Beobachtungszeitraum von 2008 bis 2014. Ausgangskohorte (Originalpopulation) waren unter allen Versicherten diejenigen, die in den Jahren 2008 und 2009 keinen Schlaganfall (ischämisch, hämorrhagisch, TIA) gemäß ICD-Codes I61-I64, G45 hatten und im Jahr 2010 ein Mindestalter von 45 Jahren erreicht hatten (kindliche und juvenile Schlaganfälle sollten nicht untersucht werden). Fälle wurden in den Jahren 2010 bis 2014 registriert. Die Gruppe der Personen der Ausgangskohorte, die auch bis zum Ende des Beobachtungszeitraums keinen Schlaganfall hatten bildete die Kontrollkohorte der Studie.

Die weiteren Populationen waren wie folgt definiert:

- Inzidenter Schlaganfall im Zeitraum 2010 bis 2014
- Inzidenter Schlaganfall mit letalem Ausgang im Zeitraum 2010 bis 2014
- Schlaganfallrezidiv im Zeitraum 2010 bis 2014
- Schlaganfallrezidiv mit letalem Ausgang im Zeitraum 2010 bis 2014

Folgende Expositionsfaktoren / Einflussgrößen auf die Zielkriterien wurden untersucht:

- Alter
- Geschlecht
- Region (nach 4-stelliger PLZ des Wohnorts des Versicherten)
- Arztkontakte/Rezepte

Mit Schwerpunkt auf

- Komorbiditäten und
- Medikamentösen Therapien soll geprüft werden ob sich auffällige Korrelationen zwischen diesen und dem Schlaganfallereignis/-rezidiv in den Daten nachweisen lassen, um eventuell Wechselwirkungen zu finden, die das Schlaganfallrisiko erhöhen oder senken können. Insgesamt wurden die häufigsten 500 Komorbiditäten der Ausgangskohorte, sowie 1500 am häufigsten verordneten Medikamente (PZNs) untersucht.

3.2 Datenanalyse

Die am ZEGV in einer klassischen relationalen Datenbank vorliegenden pseudonymisierten Routinedaten wurden aus den verschiedenen Tabellen (stationäre und ambulante Versorgungsdaten, soziodemographische Daten, Diagnosen, Medikamente) zusammengefügt und in eine Projektdatenbank überführt. Im analytischen Teil wurde zunächst eine Beschreibung der Baseline Charakteristiken der Ausgangs- und Studienkohorten erstellt basierend auf inzidenten Patienten und relative Risiken für verschiedene Einflussfaktoren berechnet. Kumulative 5-Jahres Inzidenzen und 30-Tage Mortalitätsraten wurden berechnet, sowie univariate Regressionen zu bekannten Risikofaktoren durchgeführt.

Methoden des maschinellen Lernens

Für die Analysen mit den Methoden des maschinellen Lernens (ML) wurden aus den identifizierten Kohorten neue Analysedatensätze generiert. Um auch zeitliche Bezüge zu berücksichtigen wurde eine spezielle Fall-Kontroll Studie konzipiert (Abb.1). Hierbei konnte jeweils ein Zeitraum von 2 Jahren vor einem Schlaganfallereignis genutzt werden, um medikamentöse Vortherapien und Komorbiditäten jedes Patienten zu erfassen. Der eigentliche Krankenhausaufenthalt mit Schlaganfallereignis wurde dabei bewusst ausgeschlossen um Folgen des Schlaganfalls nicht mit den Ursachen zu verwechseln, ansonsten interpretiert der Algorithmus nämlich eine Hemiparese (G81), eigentlich eine Folge eines Schlaganfalles, als Hauptursache für das Schlaganfallereignis.

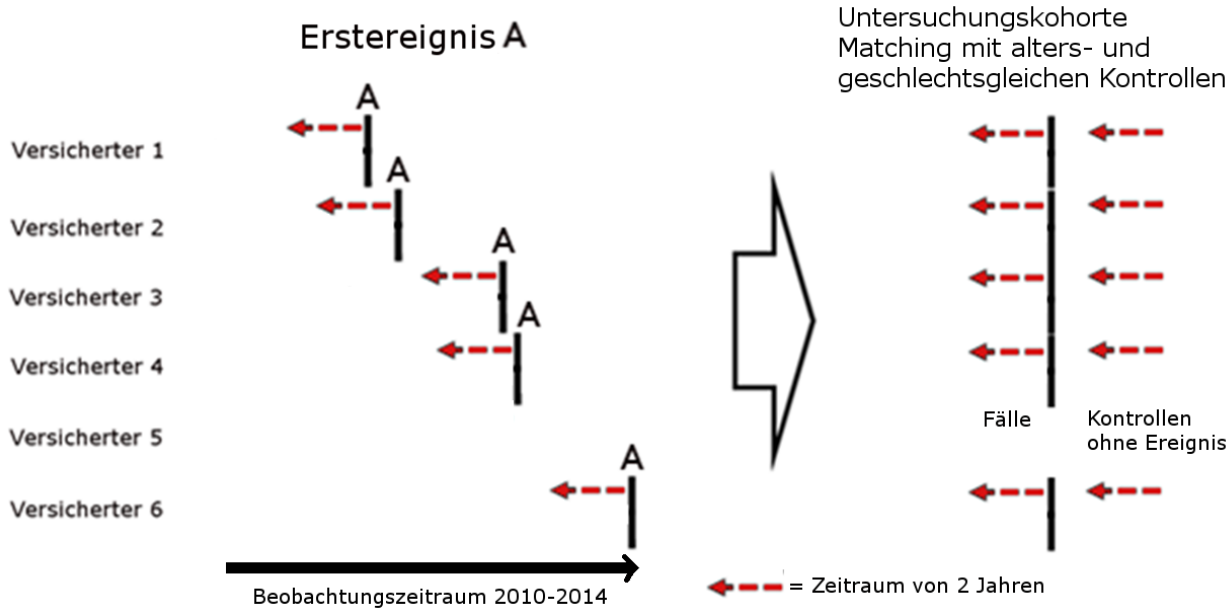


Abb.1. Fall-Kontroll Studie: ML-Analysen des Schlaganfallereignisses; jedem Patienten mit Schlaganfallereignis (Fall) wurde eine Kontrolle durch Matching (1:1) zugewiesen. Für jeden Patienten konnte ein Zeitraum von 2 Jahren vor Erstereignis untersucht und mit demselben Zeitraum einer Kontrolle verglichen werden (tag-genau).

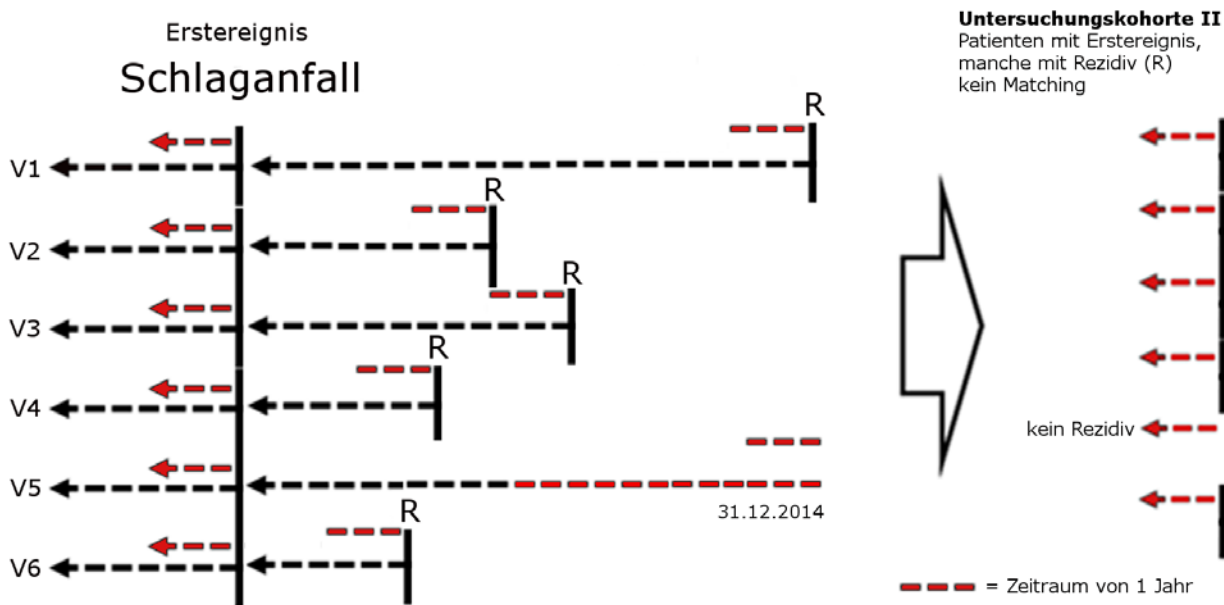


Abb.2. Analyse des Schlaganfallrezidivs (R); alle Versicherten mit Erstereignis (V1-6) bilden hier die Untersuchungskohorte; es wird jeweils ein Zeitraum von 1 Jahr vor dem Rezidiv untersucht. Für Patienten ohne Rezidiv wurde der 31.12.2014, das Ende des Beobachtungszeitraums, als zeitlicher Vergleichspunkt gewählt, von dem aus ein 1 Jahr zuvor nach Komorbiditäten und medikamentösen Therapien jedes Patienten geschaut wurde.

Es wurden aus den identifizierten inzidenten Fällen (Schlaganfallereignis im Beobachtungszeitraum 2010-2014) Kohorten aufgebaut und ein probabilistisches Matching zur Identifikation entsprechender Kontrollen durchgeführt („propensity score“ Matching mit „nearest neighbor“ Algorithmus $k=1$) (2). Durch das Fall-Kontrolldesign konnte ein Beobachtungszeitraum von maximal 2 Jahren vor dem Schlaganfallereignis untersucht werden (auch bei den tag-genau gematchten Kontrollen).

Dazu wurden Medikamentenabgaben (bis zu 1500 häufigste PZN-Codes) und Komorbiditäten (500 häufigste Diagnosen) abgefragt und in einen Analysedatensatz überführt. Damit war die Anzahl gleichzeitig betrachteter Variablen doch sehr beträchtlich.

Zur Analyse des Rezidiv-Risikos wurde kein Matching durchgeführt, hier wurden alle Patienten mit Erstereignis berücksichtigt, inklusive darin enthaltener Kontrollen ohne Rezidiv. Der Beobachtungs-/Vergleichszeitraum betrug für alle Patienten 1 Jahr vor Rezidiv (rote gestrichelte Linie in Abb.2).

Im Anschluss wurden verschiedene ML-Techniken (3) eingesetzt, um möglichst prädiktive Modelle für Schlaganfallereignis und –rezidiv abzuleiten (hohe Accuracy), mit dem Ziel relevante bekannte Risikofaktoren zu replizieren sowie unbekannte Risiko- und Präventivfaktoren, sowie Kombinationen davon, zu identifizieren. Die korrekte Vorhersage aller Patienten mit Ereignis (high „true positive rate“) und die Vorhersage aller ohne Ereignis (high „true negative rate“) war dabei von höchster Bedeutung und wurde als prozentuale Übereinstimmung (Accuracy) angegeben und optimiert.

Folgende Algorithmen wurden eingesetzt: Logistische Regression mit L2 Regularisation (LRL2) (4, 5), Entscheidungsbäume (DT) (6, 7), Random Forests (RF) (8), Support Vector Machines (SVM) (9, 10), Neuronale Netze (NN) (11, 12), Deep Learning (DNN) (13, 14).

4. Wissenschaftliche Ergebnisse des Projekts

4.1 Untersuchungskohorten

Aus der Ausgangskohorte von Patienten ohne Schlaganfall in den Jahren 2008/ 2009 ($n=929152$) konnte eine Kontrollpopulation (kein Schlaganfall 2010-2014) von 884 035 Patienten sowie drei Untersuchungskohorten mit 4759 Patienten mit hämorrhagischem Schlaganfall (I61, I62), 29382 Patienten mit ischämischem S. (I63, I64) und 13253 Patienten mit einer TIA (G45) identifiziert werden. Die Verteilung der Untersuchungskohorten in Altersklassen inklusive Rezidivraten sowie relevante Komorbiditäten sind in Tab.1 dargestellt.

Die ermittelte Schlaganfallrate unserer Population AOK PLUS-Versicherter Patienten in Sachsen steht in Einklang mit Literaturangaben über die Schlaganfallhäufigkeit in Deutschland (Erlanger Schlaganfallregister) (15).

Die Ausgangskohorte (siehe Tab.1) hat ein Durchschnittsalter von 65,8 Jahren im Jahr 2010 und einen Frauenüberschuss (56,7%). Die daraus hervorgehende Kontrollpopulation ohne Schlaganfall im Beobachtungszeitraum 2010 bis 2014 ist der Ausgangspopulation in Alters- und Geschlechtsverteilung sehr ähnlich und auch in der Häufigkeitsverteilung der Komorbiditäten, der Inanspruchnahme des Gesundheitssystems, sowie den gemittelten Bruttoausgaben aller Krankenhausaufenthalte vergleichbar. Die Untersuchungskohorten sind im Mittel 8 bis 10 Jahre älter. Das Geschlechterverhältnis ist beim ischämischen S. und bei der TIA vergleichbar mit der Ausgangs- und Kontrollpopulation, weicht jedoch beim hämorrhagischem S. davon ab. Hier sind beide Geschlechter etwa gleich häufig betroffen. Das Risiko einen Schlaganfall zu erleiden steigt mit dem Alter, dies ist aber beim ischämischen S. und bei der TIA ausgeprägter als beim hämorrhagischem S. Der prozentuale Anteil an Patienten mit bestimmten Komorbiditäten ist generell in allen Untersuchungskohorten im Vergleich zur Ausgangs- und Kontrollkohorte erhöht (Patienten sind aber auch deutlich älter im Mittel). Das Risiko einen Schlaganfall zu erleiden ist bei Vorhandensein von Bluthochdruck (Hypertonie) stark erhöht (RR über 8). Diabetes mellitus und Hyperlipidämie erhöhen zumindest für den ischämischen S. und für eine TIA das Risiko (RR bis ~2).

4.2 Arztkontakte/Leistungen

Die Variablen zur Inanspruchnahme des Gesundheitssystems zeigen keine deutliche Zunahme in den Untersuchungskohorten. Allein die Kontakte mit ambulanten Neurologen sind höher in den Untersuchungskohorten, nicht aber die Kontakte mit Allgemeinmedizinern im Vergleich zur Kontrollkohorte. Die Kosten für die stationäre Versorgung erhöhten sich bei Patienten mit ischämischem Schlaganfall im Mittel um etwa 50%, bei Patienten mit hämorrhagischem Schlaganfall um über 200% im Vergleich zur Kontrollkohorte (Tab.1).

4.3 Kumulative Inzidenzen und Mortalität über 5 Jahre

Die 5-Jahres kumulative Inzidenzen nach Altersgruppen sind jeweils getrennt für hämorrhagischen, ischämischen Schlaganfall und TIA in Abb.4 dargestellt. Die Altersgruppe 75-79 Jahre zeigte die höchste Inzidenz, mit einem Abfall in den höheren Altersgruppen. In der höchsten Altersgruppe (95+) fällt die Inzidenz auf sehr geringe Werte ab und dass sowohl beim hämorrhagischen und ischämischen S., als auch bei der TIA. Ein ähnliches Muster ist von Krebserkrankungen bekannt (16), hier fallen sowohl Inzidenzen, so wie teilweise auch die Mortalitätsraten in den höchsten Altersgruppen für fast alle untersuchten Krebsarten deutlich ab. In der Literatur werden verschiedene potentielle Gründe dafür diskutiert. Es ist nicht bekannt, ob dahinter biologische Prozesse stecken und damit ein kausaler Zusammenhang besteht, oder ob es sich um nicht kausale Zusammenhänge handelt, die durch Selektionseffekte durch Stratifizierung der Patienten entstanden sind (16). Zum einen könnten sehr alte Menschen ein geringeres Risiko für ein Schlaganfallereignis im Vergleich zu jüngeren Menschen besitzen, da durch die Seneszenz andere Zell- sowie Immunprozesse ablaufen (kausaler Pfad). Zum anderen (für uns wahrscheinlicher) könnte der Rückgang der Inzidenzen im hohen Alter mit einer Unterdiagnostik leichter Schlaganfälle erklärt werden, zumal der Anteil letaler Schlaganfälle in den höheren Altersgruppen weiter ansteigt (Abb.5). Es ist bekannt, dass die Hospitalisierungsrate wegen Schlaganfallereignis mit dem Alter absinkt (17, 18) und anscheinend im hohen Alter leichte Schlaganfälle nicht mehr in dem Maß stationär behandelt werden (Abb.3). Die Autoren versuchten dies mit Bettenknappheit und einem Mangel an Stroke Units und dem Fehlen von thrombolytischer Therapie zur Studienzeit zu erklären. Es handelt sich hierbei um Daten aus der Rotterdam Study (single center, 1990-1999). In unseren Daten fehlen alle Fälle von inzidentem

häuslichen Schlaganfall mit unmittelbarem Versterben und daher unterschätzen wir die Inzidenzen im hohen Alter zusätzlich noch.

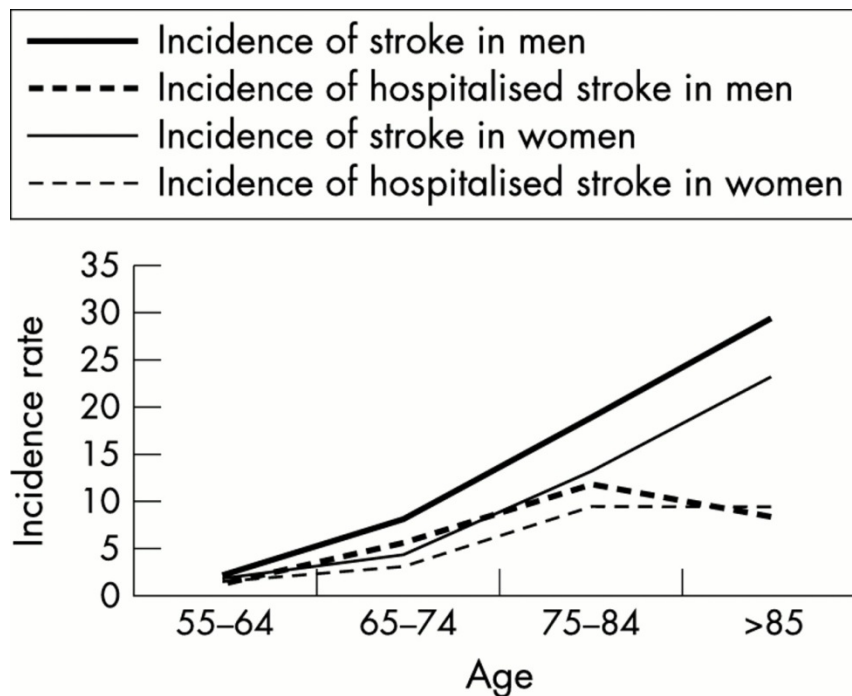


Abb.3: Hospitalisierungsrate nach Schlaganfallereignis nach Hollander et al., 2003 (17) Trotz einer weiteren Zunahme der Inzidenz im Alter nimmt die Rate der Krankenhausaufenthalte im Alter von 80 Jahren und darüber ab oder bleibt konstant. Daten stammen aus der Rotterdam Studie.

Beim hämorrhagischem S. liegt die alters-abhängige Inzidenz der Männer stets höher als die der Frauen, obwohl beide Geschlechter etwa gleich häufig betroffen sind (Tab.1, Abb.4). Beim ischämischen S. und bei der TIA haben zunächst die Männer ebenfalls eine höhere Inzidenz. In den höheren Altersgruppen jedoch zeigt sich ein Angleich beziehungsweise sogar höherer Anteil von weiblichen Patienten mit Schlaganfallereignis. Dieser Umstand liegt sowohl an der Demographie als auch der Ätiologie. Grundsätzlich sind Frauen in unserer Untersuchungskohorte vom ischämischen S. oder der TIA stärker betroffen als Männer (Tab.1), sie sind aber auch in der Grundgesamtheit der Ausgangspopulation bereits häufiger vertreten und somit ergibt sich nur ein leicht erhöhtes Risiko für Frauen in Summe. Bei den Frauen verschiebt sich zusätzlich, wahrscheinlich bedingt durch die höhere Lebenserwartung, der Anteil von Patienten mit Schlaganfallereignissen in höhere Altersstufen im Vergleich zu Männern (Abb.4). Die vorliegende Arbeit ist die erste Studie die fallende Inzidenzen beim Schlaganfall in den höchsten Altersstufen zeigt. Die meisten Arbeiten geben die höchste Altersgruppe mit 85+ an oder

untersuchten sehr alte Patienten überhaupt nicht (19). Dadurch konnten diese Studien die fallenden Inzidenzen gar nicht entdecken.

Age-dependent cumulative incidences per 100,000 persons

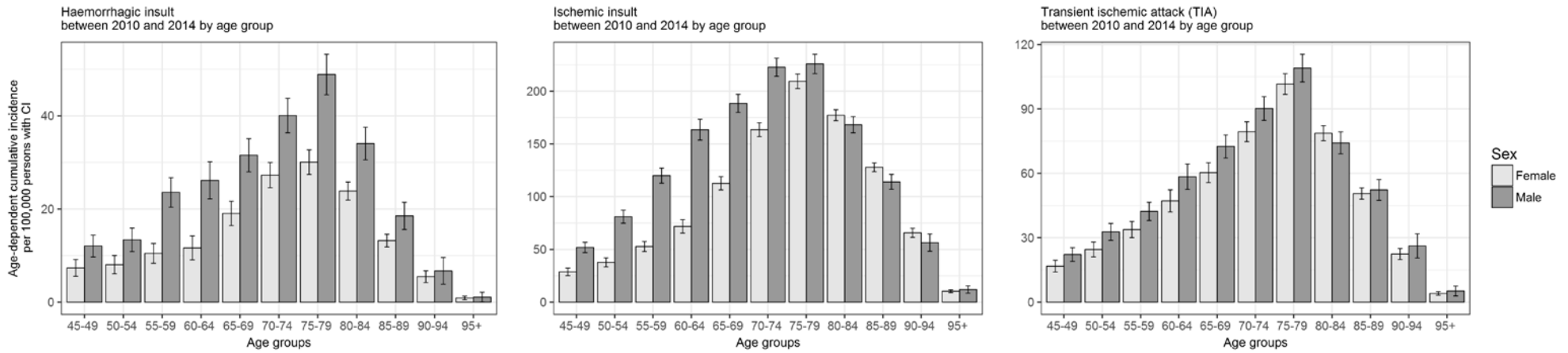


Abb.4. 5-Jahres kumulative Inzidenz

Age-dependent mortality per 1,000 persons

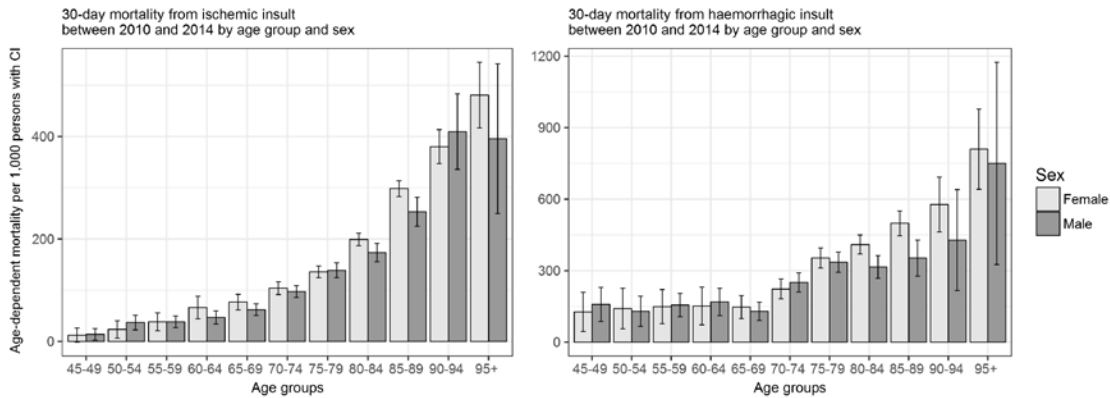


Abb.5. Schlaganfallmortalität

4.4 Risikofaktorenanalyse

Zunächst konnten die bekannten Risikofaktoren für inzidente Schlaganfälle im Datensatz repliziert werden, was die Validität der Studie stützt und Voraussetzung für die explorativen Data Mining Analysen darstellt.

Univariate Analysen der häufigsten vaskulären Risikofaktoren, aber auch weiterer assoziierter Komorbiditäten, zeigten die höchsten Risiken für folgende bekannte Risikofaktoren (adjustiert für Alter und Geschlecht): Hypertonie (OR 4,39) und Vorhofflimmern (OR 4,41) für ischämischen und hämorrhagischen Schlaganfall und TIA, zusätzlich schädlicher Alkoholgebrauch (OR 5,01) für den hämorrhagischen Schlaganfall (Tab.2-4).

Tab.2. Univariate Regression: Risikofaktoren für den hämorrhagischen Schlaganfall

Univariate crude analysis of risk factors for haemorrhagic insult and 30-day mortality (adjusted for sex and age)

Risk factor (ICD-10-GM Code)	OR for presence of haemorrhagic insult (95% CI) *	p-value	OR for 30-day mortality after insult (95% CI) *	p-value
Hypertension (I10/I11)	3.26 (2.89-3.67)	<0.001	0.78 (0.62-0.99)	0.036
Diabetes mellitus (E10-E14)	1.39 (1.28-1.51)	<0.001	1.08 (0.95-1.23)	0.247
Disorders of lipoprotein metabolism (E78)	1.23 (1.14-1.34)	<0.001	0.76 (0.67-0.87)	<0.001
Sleep disorders (G47)	1.16 (0.99-1.36)	0.063	0.67 (0.52-0.87)	0.003
Chronic ischemic heart disease (I25)	1.36 (1.25-1.49)	<0.001	1.14 (0.99-1.30)	0.059
Heart failure (I50)	2.31 (2.08-2.56)	<0.001	1.14 (0.98-1.31)	0.083
Atrial fibrillation and flutter (I48)	3.55 (3.20-3.96)	<0.001	1.03 (0.90-1.18)	0.653
Other chronic obstructive pulmonary disease (J44)	1.28 (1.12-1.46)	<0.001	1.09 (0.89-1.33)	0.399
Nicotine dependence (F17)	2.06 (1.63-2.61)	<0.001	1.59 (1.13-2.21)	0.007
Overweight and obesity (E66)	1.24 (1.10-1.40)	<0.001	0.76 (0.62-0.92)	0.006
Alcohol related disorders (F10)	5.01 (3.87-6.57)	<0.001	1.27 (0.93-1.71)	0.131
Myocardial infarction (I21/I22)	1.44 (1.17-1.78)	<0.001	1.24 (0.93-1.64)	0.145
Major depressive disorder (F32/F33)	1.35 (1.20-1.52)	<0.001	0.83 (0.69-0.99)	0.043
Amyloid angiopathy (E85+I68)	NA	NA	0.18 (0.04-0.49)	0.004
Other coagulation defects (D68)	14.42 (11.48-18.38)	<0.001	1.19 (1.01-1.40)	0.033
Congenital malformations of cardiac septa (Q21)	2.08 (1.09-4.17)	0.03	1.04 (0.39-2.46)	0.935
Migraine (G43)	0.74 (0.52-1.04)	0.083	0.55 (0.26-1.07)	0.097
Cancer (C00-C97, without C44)	1.21 (1.08-1.35)	0.001	1.10 (0.92-1.31)	0.278
Parkinson's disease (G20-22)	1.46 (1.15-1.87)	0.002	0.89 (0.63-1.25)	0.512
Dementia (F00-03/G30)	2.13 (1.90-2.39)	<0.001	1.26 (1.08-1.47)	0.004

* adjusted for age and sex

Tab.3. Univariate Regression: Risikofaktoren für den ischämischen Schlaganfall

Univariate crude analysis of risk factors for ischemic insult and 30-day mortality

Risk factor (ICD-10-GM Code)	OR for presence of ischemic insult (95% CI) *	p-value	OR for 30-day mortality after insult (95% CI) *	p-value
Hypertension (I10/I11)	4.39 (4.16-4.63)	<0.001	0.65 (0.57-0.74)	<0.001
Diabetes mellitus (E10-E14)	1.76 (1.70-1.82)	<0.001	1.21 (1.13-1.29)	<0.001
Disorders of lipoprotein metabolism (E78)	2.37 (2.29-2.45)	<0.001	0.54 (0.50-0.57)	<0.001
Sleep disorders (G47)	1.23 (1.15-1.31)	<0.001	0.87 (0.77-0.99)	0.038
Chronic ischemic heart disease (I25)	1.47 (1.42-1.52)	<0.001	1.31 (1.22-1.40)	<0.001
Heart failure (I50)	2.32 (2.23-2.42)	<0.001	1.76 (1.64-1.88)	<0.001
Atrial fibrillation and flutter (I48)	4.41 (4.23-4.60)	<0.001	1.67 (1.56-1.78)	<0.001
Other chronic obstructive pulmonary disease (J44)	1.19 (1.13-1.25)	<0.001	1.39 (1.26-1.53)	<0.001
Nicotine dependence (F17)	3.06 (2.79-3.36)	<0.001	1.15 (0.94-1.40)	0.17
Overweight and obesity (E66)	1.44 (1.38-1.51)	<0.001	1.02 (0.93-1.12)	0.675
Alcohol related disorders (F10)	2.95 (2.66-3.28)	<0.001	1.64 (1.33-2.00)	<0.001
Myocardial infarction (I21/I22)	1.74 (1.60-1.89)	<0.001	1.84 (1.62-2.08)	<0.001
Major depressive disorder (F32/F33)	1.39 (1.34-1.45)	<0.001	0.88 (0.80-0.96)	0.003
Amyloid angiopathy (E85+I68)	4.00 (2.15-8.16)	<0.001	0.56 (0.13-1.57)	0.335
Other coagulation defects (D68)	2.16 (1.94-2.41)	<0.001	1.67 (1.42-1.95)	<0.001
Congenital malformations of cardiac septa (Q21)	7.71 (5.83-10.41)	<0.001	0.37 (0.20-0.61)	<0.001
Migraine (G43)	0.90 (0.78-1.04)	0.14	0.53 (0.34-0.79)	0.003
Cancer (C00-C97, without C44)	1.10 (1.05-1.15)	<0.001	1.28 (1.17-1.39)	<0.001
Cancer (C00-C97, without C44 & C77-C79)	1.10 (1.05-1.15)	<0.001	1.27 (1.16-1.39)	<0.001
Cancer (C00-C97, without C44 & C77-C79) - 2-days mortality	---	---	1.10 (0.90-1.32)	0.357
Parkinson's disease (G20-22)	1.42 (1.29-1.56)	<0.001	1.30 (1.11-1.52)	0.001
Dementia (F00-03/G30)	2.04 (1.95-2.13)	<0.001	1.49 (1.39-1.60)	<0.001

* adjusted for age and sex

Tab.4. Univariate Regression: Risikofaktoren für eine TIA

Univariate crude analysis of risk factors for transient ischemic attack

Risk factor (ICD-10-GM Code)	OR for presence of ischemic attack (95% CI) *	p-value
Hypertension (I10/I11)	3.57 (3.32-3.85)	<0.001
Diabetes mellitus (E10-E14)	1.47 (1.40-1.55)	<0.001
Disorders of lipoprotein metabolism (E78)	2.46 (2.35-2.59)	<0.001
Sleep disorders (G47)	1.30 (1.19-1.43)	<0.001
Chronic ischemic heart disease (I25)	1.47 (1.40-1.55)	<0.001
Heart failure (I50)	2.18 (2.05-2.31)	<0.001
Atrial fibrillation and flutter (I48)	2.68 (2.52-2.86)	<0.001
Other chronic obstructive pulmonary disease (J44)	1.20 (1.11-1.30)	<0.001
Nicotine dependence (F17)	2.10 (1.81-2.44)	<0.001
Overweight and obesity (E66)	1.34 (1.25-1.44)	<0.001
Alcohol related disorders (F10)	1.97 (1.67-2.34)	<0.001
Myocardial infarction (I21/I22)	1.33 (1.17-1.52)	<0.001
Major depressive disorder (F32/F33)	1.49 (1.39-1.59)	<0.001
Amyloid angiopathy (E85+I68)	1.80 (0.62-5.86)	0.292
Other coagulation defects (D68)	1.62 (1.37-1.93)	<0.001
Congenital malformations of cardiac septa (Q21)	8.77 (5.39-15.31)	<0.001
Migraine (G43)	1.12 (0.93-1.34)	0.233
Cancer (C00-C97, without C44)	1.10 (1.02-1.17)	0.011
Parkinson's disease (G20-22)	1.71 (1.49-1.96)	<0.001
Dementia (F00-03/G30)	2.21 (2.07-2.36)	<0.001

* adjusted for age and sex

Interessanterweise zeigte sich bei einigen dieser Risikofaktoren eine erniedrigte Mortalitätsrate, sodass weitere Analysen des metabolischen Syndroms (20) als Risikofaktor für Inzidenz sowie Mortalität erfolgten. Tatsächlich zeigte sich bei Vorhandensein mehrerer Risikofaktoren aus dem Symptomenkomplex des metabolischen Syndroms (Hypertonie, Übergewicht, Hyperlipoproteinämie, gestörte Glukoseintoleranz bzw. Diabetes Typ II) zwar eine Erhöhung des Risikos für das Auftreten eines Schlaganfalls, jedoch eine erniedrigte Mortalität sowohl für den ischämischen (signifikant) als auch hämorrhagischen Schlaganfall (nicht signifikant). Dieses scheinbar paradoxe Ergebnis lässt sich am ehesten auf einen Selektionseffekt und potentieller Unterdiagnostik dieser Begleiterkrankungen bei Patienten mit tödlichem Schlaganfall zurückführen (Tabelle 5). In der Literatur ist dieses Phänomen unter den Stichwörtern „obesity paradox“ oder „Adipositas Paradoxon“ sowie „reverse epidemiology/causation“ bekannt (21-24). Zusammenfassend wurde in zahlreichen Studien Folgendes gefunden: Übergewicht schützt vor bestimmten Erkrankungen, u.a. bei chronischer Herzinsuffizienz, koronarer Herzkrankheit, arterieller Hypertonie, Diabetes mellitus Typ 2, Alzheimer und Schlaganfall (23). Insbesondere Outcomes wie erkrankungsspezifische Mortalität oder Komplikationsraten nach OP sind davon betroffen. Während einige Autoren an einem kausalen Zusammenhang festhalten (25-27), plädiert die Mehrheit für einen nicht kausalen Selektionseffekt der diesem scheinbar paradoxen Phänomen zugrunde liegt (21-24, 28, 29). Wenn zur Ermittlung des Gewichtsstatus der BMI (body mass index) genutzt wird, führt dies per se zum Paradoxon. Bei Nutzung des Hüftumfangs oder direkter Messung des viszeralen Fettes kommt es nicht zu den paradoxen Erscheinungen (23). Sehr niedrige BMI-Werte (Untergewicht) sind häufig mit schlechteren Outcomes assoziiert. Daher scheinen Patienten mit Übergewicht bessere Outcomes zu haben als Vergleichspatienten mit Normalgewicht/Untergewicht. Diese beiden Gruppen (Normalgewicht/Untergewicht) werden bei der Nutzung von Adipositas-Codes (E66) in eine Gruppe einsortiert und dadurch entstehen diese verzerrten Ergebnisse. Nicht Übergewicht schützt, sondern Untergewicht bzw. Gewichtsverlust schadet oder ist sogar eine Folge der Erkrankung nicht der Auslöser. Zusätzlich ist denkbar (24), dass in der Gruppe mit einer Erkrankung aus dem Symptomenkomplex des metabolischen Syndroms insgesamt der Anteil leichter Schlaganfälle höher ist, als in der Gruppe ohne diese Risikofaktoren und durch diese Selektionsverzerrung die Mortalität in der Gruppe mit metabolischen Erkrankungen scheinbar erniedrigt ist. Könnten wir für den Schweregrad des Schlaganfalles ordentlich kontrollieren, müssten die gefundenen paradoxen Beziehungen dadurch aufgelöst werden können. Wir halten aber eine Unterdiagnostik der Begleiterkrankungen aus dem

Symptomenkomplex bei Patienten mit tödlichem Schlaganfall für relevant und denken, dass dies allein schon zu den verzerrten Ergebnissen führt. Es handelt sich hierbei ja um den inzidenten Schlaganfall und die 30-Tages Mortalität, daher ist es plausibel, dass ein Teil dieser Patienten deutlich unterdiagnostiziert war zum Zeitpunkt des Todes im Hinblick auf das metabolische Syndrom. Eine Abwertung aller biologischen Erklärungsversuche sollte zugunsten einer Biasinduktion durch Unter- oder Fehlcodierung in Betracht gezogen werden. Bei einigen primär nicht metabolischen Erkrankungen wie Schlafstörungen (G47), Migräne (G43), angeborene Fehlbildungen von Herz-Septa (Q21) oder Depression (F32/33) zeigen unsere Daten (Tab.2, Tab.3) ein vermindertes Risiko für das Versterben nach einem inzidenten Schlaganfall, was kausal (biologisch), zumindest auf den ersten Blick, nicht nachvollziehbar ist und daher unseres Erachtens am besten durch eine selektive Unterdiagnostik von Schlaganfallpatienten mit tödlichem Ausgang und damit einhergehender Unterversorgung im Vorfeld erklärt werden könnte.

Tab.5. Analyse des metabolischen Syndroms als RisikofaktorAnalysis of metabolic syndrome¹ as risk factor for stroke and 30-day mortality after stroke

as risk factor for ischemic insult	OR for presence of ischemic insult (95% CI) ²	OR for 30-day mortality after insult (95% CI) ²
0/4 diseases (ref)	---	---
1/4 diseases	3.48 (3.22-3.77)	0.58 (0.48-0.71)
2/4 diseases	6.63 (6.14-7.16)	0.48 (0.40-0.57)
3/4 diseases	9.27 (8.56-10.04)	0.42 (0.35-0.51)
4/4 diseases	11.15 (10.09-12.34)	0.47 (0.38-0.58)
0/4 diseases (ref)	---	OR for 2-day mortality after insult (95% CI) ²
1/4 diseases	---	0.46 (0.34-0.64)
2/4 diseases	---	0.34 (0.25-0.46)
3/4 diseases	---	0.30 (0.22-0.41)
4/4 diseases	---	0.32 (0.21-0.48)
as risk factor for hemorrhagic insult		OR for 30-day mortality after insult (95% CI) ²
0/4 diseases (ref)	---	---
1/4 diseases	2.74 (2.35-3.21)	0.94 (0.69-1.29)
2/4 diseases	3.07 (2.64-3.58)	0.81 (0.60-1.10)
3/4 diseases	3.91 (3.32-4.62)	0.74 (0.54-1.02)
4/4 diseases	3.87 (3.06-4.90)	0.70 (0.47-1.06)
as risk factor for transient ischemic attack		
0/4 diseases (ref)	---	---
1/4 diseases	3.13 (2.80-3.51)	---
2/4 diseases	5.87 (5.27-6.56)	---
3/4 diseases	7.41 (6.61-8.31)	---
4/4 diseases	8.51 (7.35-9.88)	---

¹ metabolic syndrome (hypertension I10+I11; overweight and obesity E66; disorders of lipoprotein metabolism E78; diabetes mellitus type 2 E11)² adjusted for age and sex

4.5 Inzidenzen und Mortalität im regionalen Vergleich

Es erfolgte eine Analyse der Inzidenzen und der Schlaganfall-bezogenen Mortalität getrennt nach 3-stelligen Postleitzahlen in Sachsen. Hier zeigte sich eine leicht inhomogene Verteilung des ischämischen Schlaganfalls mit niedrigerer Häufigkeit und Mortalität in Ostsachsen sowie in dem Stadtgebiet um Leipzig (Abb.6,7). Nordsachsen (Gebiet um Torgau) liegt im sächsischen Vergleich am oberen Rand der Mortalitätsskala, gefolgt von leicht erhöhten Werten über dem Durchschnitt für das mittlere Erzgebirge (Gebiet um Annaberg-Buchholz). Die drei Metropolregionen Chemnitz, Leipzig und Dresden (Maximalversorger bzw. Universitätsklinikum) liegen deutlich unter dem Durchschnittswert. Die leichte Clusterung der Mortalitätsraten könnte ein Anzeichen für regionale Versorgungsunterschiede im Management akuter Schlaganfallpatienten sein, da mit der Entfernung von einem Uniklinikum die Raten steigen.

Insbesondere könnte auch die inhomogene Verteilung der Inzidenzen mit erhöhten Werten für Westsachsen und dem Nordosten von Sachsen (Abb.6) eine regional unterschiedliche Wirksamkeit von Präventivmaßnahmen bedeuten (Raten sind alters-standardisiert).

**Age-standardized Cumulative Incidence of ischemic insult
(2010-2014) per 100.000 persons (CAR model)**

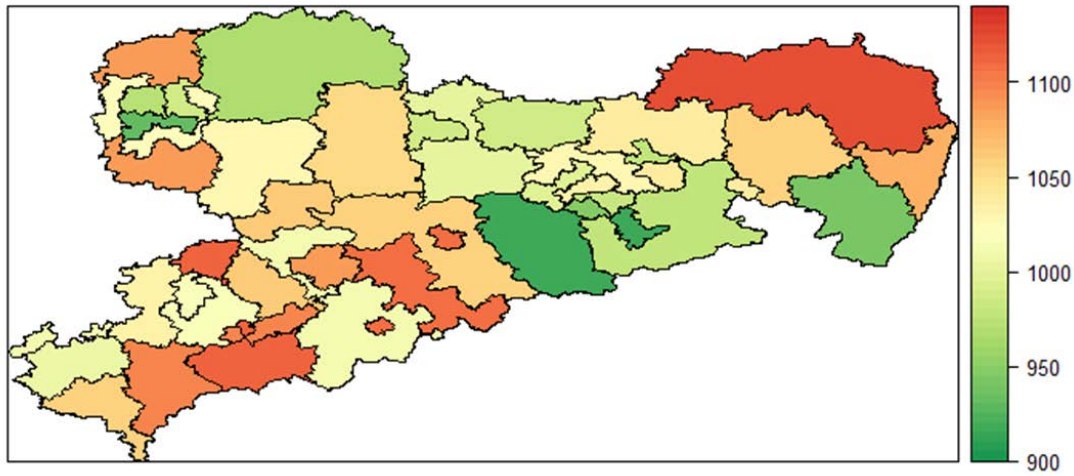


Abb.6. Kumulative 5-Jahres Inzidenz in Sachsen per 3-stelligem PLZ-Gebiet

**Age-standardized 30-day Mortality after ischemic insult
(2010-2014) per 1,000 incident cases (CAR model)**

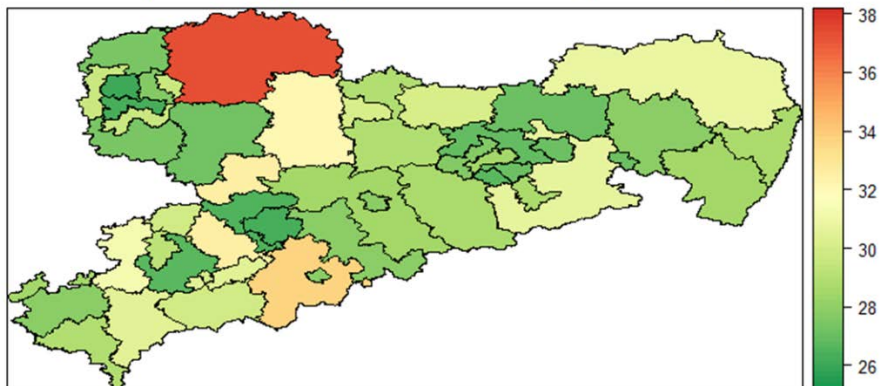


Abb.7. 30-Tages Mortalität nach ischämischen Schlaganfall (kumuliert über 5 Jahre)

4.6 Analyse des Schlaganfallereignisses

Mit Hilfe verschiedener Methoden des maschinellen Lernens sollte nun überprüft werden, ob statistisch auffällige Korrelationen zwischen Schlaganfallereignis und medikamentösen Therapien sowie vorhandenen Komorbiditäten in den Daten nachweisbar sind. Die maximale prozentuale Übereinstimmung (Accuracy) lag um 60% (Generalisierungsfehler ca. 40% - Tab.6). Es konnte somit etwa 10% der Varianz durch die Modelle aufgeklärt werden (binäres Outcome). Das Verhältnis von informativen zu uninformativen („Noise“) Variablen war sehr hoch, was die Analysen einschränkte. Risiko- sowie Präventivfaktoren für das Schlaganfallereignis sind anscheinend sehr heterogen und daher kann nur ein gewisser Anteil der Fälle über die Wirkung medikamentöser Therapien sowie vorhandener Komorbiditäten erklärt werden. Dies trifft auf alle drei untersuchten Entitäten zu (ischämischer Schlaganfall, hämorrhagischer S., TIA). Ob 10% Varianzaufklärung nun als viel oder wenig zu werten sind ist unklar. Trotz anfänglicher Euphorie zeigte sich in Analysen zum Beitrag der Genetik an Erkrankungen, dass auch dort Varianzaufklärungen von wenigen Prozent für einzelne Riskloci (Genlocus) üblich sind (30, 31) und Summeneffekte aller Genloci kombiniert üblicherweise Werte von 10-20% Heritabilität (h^2) nicht überschreiten (30, 32, 33). Diese Ergebnisse lassen die vorliegende Analyse von GKV-Daten zum Schlaganfallereignis mit einer Varianzaufklärung von ca. 10% als eher vielversprechend erscheinen. Insbesondere da in der vorliegenden Analyse Verhaltensdaten, genetische Informationen (Biomarker), sowie klinische Befunde (u.a. Laborwerte) zu den Patienten fehlten, gehen wir davon aus, dass durch Verknüpfung (Linkage) dieser Datenquellen, vertiefte Einsichten in das wahrscheinlich komplexe Zusammenspiel verschiedener Risiko- und Präventivfaktoren des Schlaganfalls gewonnen werden können und auch die Vorhersagegüte der Modelle deutlich verbesserbar ist.

Tab.6. Beispiel „Confusion matrix“ (random forest - ischämisches Schlaganfallereignis)

true	predicted	
	0	1
0	16666	11903
1	11636	17746

4.7 Analyse des Schlaganfallrezidivs

Insgesamt konnten 716 Patienten mit Rezidiv nach hämorrhagischem S. (15,1%), 3686 Rezidive des ischämischen (12,6%) und 1095 TIA-Rezidive (8,3%) im Beobachtungszeitraum identifiziert werden. Rezidive des hämorrhagischen S. sind somit etwas häufiger als Rezidive des ischämischen Typs. Während beim hämorrhagischen S. Männer etwas häufiger betroffen sind (17% gegen 13%) und die Rezidivhäufigkeit mit dem Alter stark abnimmt (45-49 – 25% auf 85+ - unter 6%), haben beim ischämischen S. und bei der TIA beide Geschlechter gleich häufig ein Rezidiv und die Häufigkeit bleibt über alle untersuchten Altersgruppen nahezu identisch. Nur in der höchsten Altersgruppe 85+ (Tab.1) sinkt beim ischämischen S. auch die Rezidivhäufigkeit, was wiederum durch eine Unterdiagnostik leichter Fälle bei Hochaltrigen bedingt sein kann. Der prozentuale Anteil an Patienten mit bestimmten Komorbiditäten nimmt in der Gruppe der Patienten mit Rezidiv im Vergleich zur Gruppe aller Patienten mit Erstereignis nochmals (deutlich) zu, bis auf eine Ausnahme (Anteil Patienten mit Koronarer Herzkrankheit – I25 - nimmt leicht ab beim hämorrhagischen Schlaganfall). Diese Zunahme kann durch ein höheres Alter der Rezidivgruppe im Vergleich zur Erstereignisgruppe begründet sein, oder aber den schlechteren Gesundheitszustand (höhere Komorbidität) der Patienten mit Rezidiv repräsentieren.

Bei der Analyse von Risiko- bzw. Präventivfaktoren für das Schlaganfallrezidiv wurden keine univariaten Analysen (Regressionen) bekannter Einflussfaktoren analog zum Erstereignis durchgeführt, sondern gleich Methoden genutzt, um den Einfluss, einzelner Komorbiditäten und medikamentöser Therapien nach Vorselektion, auf das Rezidivrisiko zu quantifizieren.

Die Bedeutsamkeit der Einflussfaktoren auf das Rezidivrisiko des ischämischen Schlaganfalls wurde mittels RF- und SVM-Analyse durchgeführt (Abb.8). Dabei hat die Komorbidität Hypertonie (I10 - ICD-10 GM) in allen Modellen den stärksten Einfluss auf das Rezidivrisiko. Sieben (Abb.8, Tab.7) der 1500 getesteten Medikamente (PZNs) haben ebenfalls einen Einfluss, aber mit untergeordneter Bedeutung. Um zu entscheiden, ob es sich dabei um risiko-erhöhende oder Präventivfaktoren handelt, wurden Wahrscheinlichkeiten der Klassenzugehörigkeiten berechnet. Folgende Variablen (Tab.8), einzeln betrachtet, erhöhen das Rezidivrisiko der Patienten (absteigende Reihenfolge nach Einfluss): Vorliegen einer Hypertonie (i10), Einnahme von Simvastatin, Pantoprazol („Proxy für ACC“), Metamizol-Natrium, Bisoprolol. Bei Vorliegen

einer Hypertonie (Partial Dependence), konnten folgende Faktoren das Risiko ein Rezidiv zu erleiden geringfügig senken (absteigende Reihenfolge nach Einfluss): Bisoprolol, Amylodipin, Omeprazol, Torasemid.

Tab.8 gibt den Einfluss (Wahrscheinlichkeitszunahme bzw. -abnahme) nach Vorselektion bestimmter Variablen an. Diese sogenannten „Partial Dependence Plots“ (34) quantifizieren den Einfluss ausgewählter Prädiktoren für die Vorhersagekraft des Rezidives und die Prädiktoren lassen sich nach deren Einflusswahrscheinlichkeit sortieren. Dieses Vorgehen wurde aber nur mit den Daten des ischämischen Schlaganfalls durchgeführt. Neben der Hypertonie, konnten wenige Medikamente als Risikofaktoren für das Auftreten eines Rezidivs erkannt werden. Simvastatin, ein Mittel zur Blutfettsenkung, zeigte den stärksten Effekt auf das Risiko ein Rezidiv zu erleiden. Ob es sich allerdings um einen kausalen Effekt handelt, ist fraglich, eher könnte wieder ein Selektionseffekt Grundlage für dieses Ergebnis sein. Patienten mit Einnahme von Simvastatin könnten eben im Mittel mehr Schlaganfallrezidive erleiden, als Patienten ohne oder mit geringeren Problemen der Blutfettwerte. Simvastatin ist hier nicht der Auslöser, sondern die Folge eines schwereren Verlaufs und damit höherem Rezidivrisikos („reverse causation“). Umgekehrt konnten einige Medikamente bei Vorliegen einer Hypertonie das Rezidivrisiko senken, wie Bisoprolol, ein Mittel zur Kräftigung der Herzfrequenz und Senkung des Blutdrucks. Hier könnte der Zusammenhang kausal sein, da es plausibel ist, dass Patienten ohne Medikament (bei Vorliegen derselben Hintergrunderkrankung) schlechtere Outcomes haben (Tab.8).

Bei der Vorhersage des Rezidivrisikos konnten höhere prozentuale Übereinstimmungen im Vergleich zum Erstereignis erzielt werden. Ein Verfahren zur Variablen-Vorselektion (35) konnte erfolgreich folgende 15 Variablen (die sich auf eine Komorbidität und 7 Medikamente reduzieren lassen – Tab.7) identifizieren, die bei ca. einem Drittel der Untersuchungspopulation (8852 von 29382) höchst prädiktiv waren (Accuracy~ 95%, kappa=0,86). Unter Einschluss aller ischämischen Schlaganfallereignisfälle konnte eine Accuracy von ca. 72% erreicht werden, was einer Varianzaufklärung von etwa 22% entspricht.

Tab.7. Identifizierte Einflussfaktoren auf das Rezidivrisiko des ischämischen Schlaganfalls nach Stability Selection.

Hypertonie
Pantoprazol
Amylodipin
Metamizol-Natrium
Torasemid
Bisoprolol
Simvastatin
Omeprazol

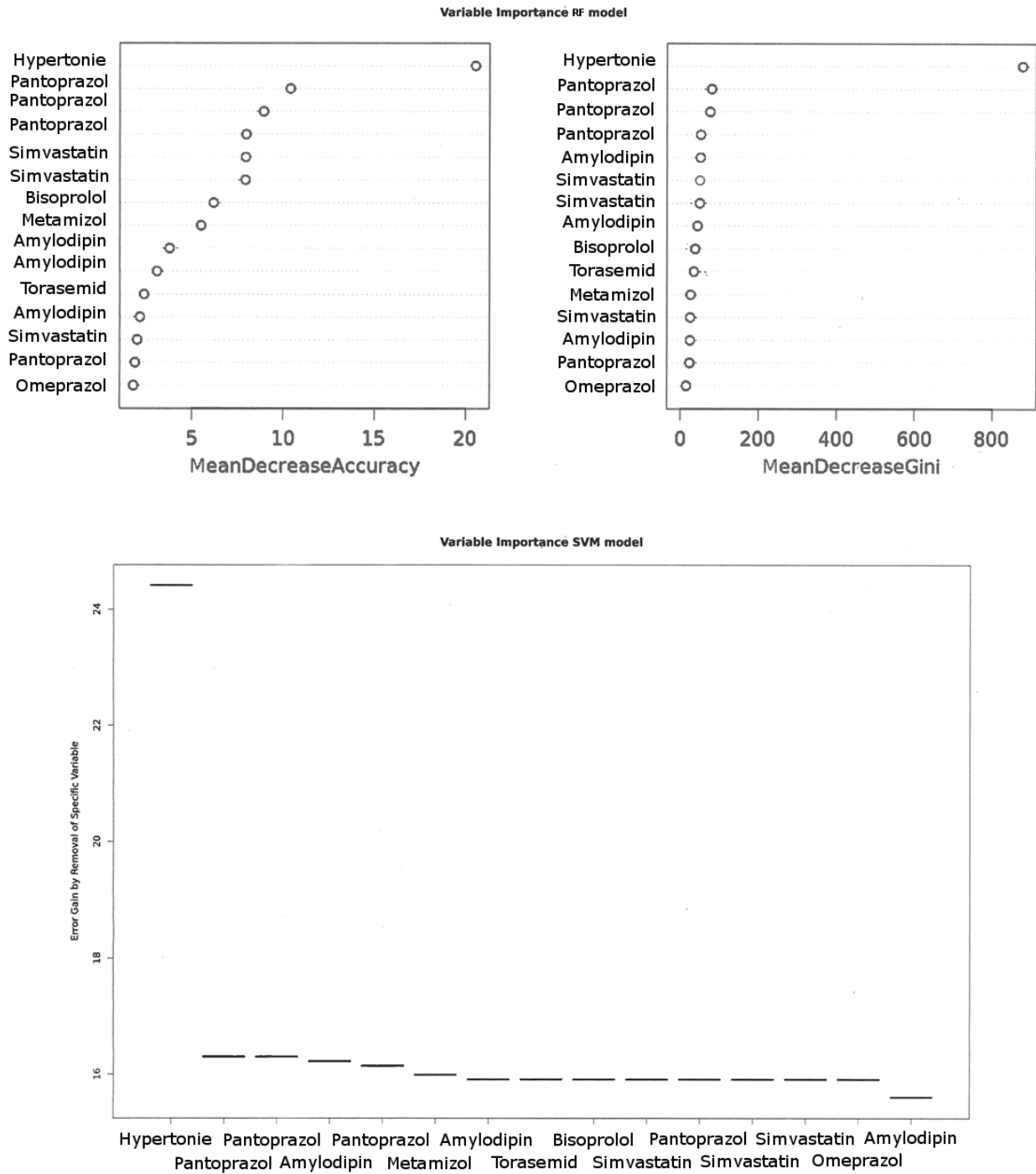


Abb.8. Die Bedeutsamkeit der Einflussfaktoren des ischämischen Schlaganfallrezidivs nach RF- (oben) und SVM-Modell (unten). Je höher die Werte, desto einflussreicher ist die Variable.

Tab.8. Partielle Abhängigkeiten der Einflussfaktoren. Oben: Einfluss der untersuchten Variable bei gleichzeitigem Festhalten der anderen Variablen auf dem Durchschnittswert. Unten:

Abhängigkeit der untersuchten Variable über den gesamten Wertebereich einer anderen Variable (hier: I10 Hypertonie).

Partial Dependence Data

Risk factors for ischemic stroke recurrence

Variable (present)	Class Probabilities	
	recurrence	no recurrence
Hypertonie	0.91	0.09
Simvastatin	0.9	0.1
Pantoprazol	0.88	0.12
Metamizol	0.86	0.14
Bisoprolol	0.61	0.39

Protective factors for ischemic stroke recurrence
(if hypertension is present)

Variable (present)	Probability reduction of recurrences
Bisoprolol	-0.08
Amylodipin	-0.07
Omeprazol	-0.06
Torasemid	-0.02

5. Abschließende Beurteilung des Projekts, Ausblick und Publikation

Das Projekt konnte erfolgreich abgeschlossen werden und liefert neue Erkenntnisse in drei Bereichen:

1. Epidemiologie des Schlaganfalls und –Rezidivs in Sachsen
2. Regionale Unterschiede und Hinweise für mögliche Unterversorgung
3. Methodisch: Data Mining funktioniert, aber weitere Datenquellen zu Verhalten (Rauchen, Ernährung, körperliche Aktivität), zu klinischen Faktoren (Laborwerte, Bildgebung, Blutdruck, etc) und zu Familienanamnese und Biomarkern offensichtlich notwendig, um einen größeren Teil der Varianz gerade bei Erstereignissen aufzuklären. Die Methode erscheint sehr vielversprechend und bietet einen innovativen Ansatz, das Konzept der Precision Medicine auf die Prävention zu übertragen.

Epidemiologie des Schlaganfalls und –Rezidivs in Sachsen

Im Rahmen des vorgelegten Projekts wurden anhand von Routinedaten u.a. mit Hilfe von Methoden des maschinellen Lernens die Schlaganfallhäufigkeit, Rezidivrate, die schlaganfallbezogene Mortalität und Risikofaktorenkonstellationen in Sachsen analysiert. Diese epidemiologischen Zahlen konnten erstmals für die Region Sachsen in dieser umfassenden Weise erhoben werden. Insgesamt stehen in der Literatur nur wenige epidemiologische Daten zum Schlaganfall in Deutschland zur Verfügung. Die letzte, bereits vor 10 Jahren publizierte Analyse von Zahlen aus dem Erlanger Schlaganfallregister zeigt vergleichbare Inzidenzen.

Regionale Unterschiede und Hinweise für mögliche Unterversorgung

Die Analyse der einzelnen Regionen innerhalb Sachsens wird noch weiter vertieft und könnte hilfreich sein im Hinblick auf die telemedizinische Versorgung von Schlaganfallpatienten durch drei getrennt arbeitende Netzwerke, mit möglichen Implikationen für eine Umgestaltung der Versorgungsstruktur.

Methodisch

Die im Rahmen des Projekts intendierte Selektion von bislang unbekanntem Risikokonstellationen mit nicht-bekanntem Faktoren oder Medikamenteninteraktionen konnte im Rahmen der Methoden des maschinellen Lernens erfolgreich durchgeführt werden, um allerdings einen höheren Anteil der Varianz, gerade beim inzidenten Erstereignis, aufklären zu können,

müssen noch weitere komplementäre Datenquellen mit hinzu gezogen werden. Verlässliche Hinweise auf unbekannte Risiko- und Präventivfaktoren des Schlaganfallrezidivs konnten aber gegeben werden.

Die Ergebnisse dieses geförderten Projekts werden zur Publikation in einem internationalen Journal eingereicht.

Literatur

1. Gueniat J, Breniere C, Graber M, et al. Increasing Burden of Stroke: The Dijon Stroke Registry (1987-2012). *Neuroepidemiology*. 2018;50(1-2):47-56.
2. Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci*. 2010;25(1):1-21.
3. Hastie T. *The Elements of Statistical Learning*. 2009.
4. Hastie T. *GLMNET*. 2009.
5. Park M, Hastie, T. Penalized Logistic Regression for Detecting Gene Interactions. *Biostatistics*. 2008;9(1):30-50.
6. Leo Breiman, Jerome Friedman, Charles J. Stone, et al. *Classification and Regression Trees*. Wadsworth: Belmont, CA: Chapman and Hall/CRC 1984.
7. Quinlan JR C4.5: Programs for machine learning. San Mateo: Morgan Kaufmann Publishers; 1993.
8. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32.
9. V V. *The nature of statistical learning theory*. New York: Springer; 1995.
10. Alexandros Karatzoglou, Alex Smola, Kurt Hornik, et al. kernlab -- An S4 Package for Kernel Methods in R. 2004;11:1--20.
11. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958;65(6):386-408.
12. Grossi E, Buscema M. Introduction to artificial neural networks. *Eur J Gastroenterol Hepatol*. 2007;19(12):1046-1054.
13. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85-117.
14. Tianqi Chen, Mu Li, Yutian Li, et al. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems 2015.
15. Heuschmann PU, Busse O, Wagner M, et al. Schlaganfallhäufigkeit und Versorgung von Schlaganfallpatienten in Deutschland. *Akt Neurol*. 2010;37:333-340.
16. Harding C, Pompei F, Wilson R. Peak and decline in cancer incidence, mortality, and prevalence at old ages. *Cancer*. 2012;118(5):1371-1386.
17. Hollander M, Koudstaal PJ, Bots ML, et al. Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study. *Journal of Neurology, Neurosurgery & Psychiatry*. 2003;74(3):317.
18. Nimptsch U, Mansky T. Trends in acute inpatient stroke care in Germany--an observational study using administrative hospital data from 2005-2010. *Dtsch Arztebl Int*. 2012;109(51-52):885-892.
19. Russo T, Felzani G, Marini C. Stroke in the very old: a systematic review of studies on incidence, outcome, and resource use. *J Aging Res*. 2011;2011:108785.
20. H T. Das Metabolische Syndrom - Beginn des "Tödlichen Quartetts"? *Austrian Journal of Cardiology*. 2005 12:6-7.
21. Ades PA, Savage PD. The obesity paradox: perception vs knowledge. *Mayo Clinic proceedings*. 2010;85(2):112-114.
22. Chang VW, Langa KM, Weir D, et al. The obesity paradox and incident cardiovascular disease: A population-based study. *PLOS ONE*. 2017;12(12):e0188636.
23. Hainer V, Aldhoon-Hainerová I. Obesity paradox does exist. *Diabetes care*. 2013;36 Suppl 2(Suppl 2):S276-S281.
24. Han SJ, Boyko EJ. The Evidence for an Obesity Paradox in Type 2 Diabetes Mellitus. *Diabetes Metab J*. 2018;42(3):179-187.
25. Schmidt DS, Salahudeen AK. Obesity-survival paradox--still a controversy? *Semin Dial*. 2007;20(6):486-492.

26. Uretsky S, Messerli FH, Bangalore S, et al. Obesity Paradox in Patients with Hypertension and Coronary Artery Disease. *The American Journal of Medicine*. 2007;120(10):863-870.
27. Vemmos K, Ntaios G, Spengos K, et al. Association between obesity and mortality after acute first-ever stroke: the obesity-stroke paradox. *Stroke*. 2011;42(1):30-36.
28. Khalangot M, Tronko M, Kravchenko V, et al. Body mass index and the risk of total and cardiovascular mortality among patients with type 2 diabetes: a large prospective study in Ukraine. *Heart*. 2009;95(6):454-460.
29. Pegueroles J, Jimenez A, Vilaplana E, et al. Obesity and Alzheimer's disease, does the obesity paradox really exist? A magnetic resonance imaging study. *Oncotarget*. 2018;9(78):34691-34698.
30. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nature reviews. Genetics*. 2014;15(11):765-776.
31. Chorney MJ, Chorney K, Seese N, et al. A Quantitative Trait Locus Associated with Cognitive Ability in Children. *Psychological Science*. 1998;9(3):159-166.
32. Wray NR, Maier R. Genetic Basis of Complex Genetic Disease: The Contribution of Disease Heterogeneity to Missing Heritability. *Current Epidemiology Reports*. 2014;1(4):220-227.
33. Minihane AM. The genetic contribution to disease risk and variability in response to diet: where is the hidden heritability? *Proc Nutr Soc*. 2013;72(1):40-47.
34. Friedman J, editor *Greedy Function Approximation: A Gradient Boosting Machine*. *Annals of Statistics*; 2000.
35. Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010;72(4):417-473.