

4 Diskussion

Die zur Zeit verfügbaren Daten zu AVMs stammen aus Patientendatenbanken von Zentren, die sich auf zum Teil unterschiedliche und konkurrierende Verfahren zur Behandlung von AVMs spezialisiert haben. Damit stellte sich die Frage,

1. ob und in welchem Ausmaß die Ergebnisse krankenhausbasierter Patientenkohorten durch einen Zuweisungsbias oder andere zentrumsspezifische Faktoren beeinflusst wurden,
2. ob dieser Einfluss sich in der Form von intervenierenden Variablen auf relevante Parameter auswirkte und
3. ob unterschiedlichen Standards im Prozess der Datenerhebung erkennbar waren, die ihrerseits Einfluss auf die Qualität der Ergebnisse nehmen konnten.

In der vorliegenden Arbeit wurden demographische, klinische und morphologische Merkmale von AVM-Patienten untersucht, die als Parameter in die klinische AVM-Forschung Eingang gefunden haben.

In einem Vergleich von an fünf Standorten erhobenen Daten aus drei etablierten AVM-Patientendatenbanken wurde aufzuzeigen versucht, ob und in welchem Umfang zentrumsspezifische Gegebenheiten die Verteilung dieser Merkmale beeinflussen. Bezüglich der untersuchten Merkmale wurden eine Reihe signifikanter Unterschiede zwischen den Zentren festgestellt, die im Folgenden kurz zusammengefasst werden sollen.

4.1 Zusammenfassung der Ergebnisse

Die Patienten der verschiedenen Zentren wurden bezüglich der demographischen Basisparameter Alter und Geschlecht verglichen. Es wurde festgestellt, dass sie sich signifikant in ihrem Durchschnittsalter unterschieden. Dabei waren insbesondere die Extremgruppen unterschiedlich belegt. In Paris wurden offenbar mehr Patienten unter 10 Jahren und in Berlin mehr Patienten über 59 Jahren behandelt als anderswo. Nicht in den Vergleich einbezogen werden konnten die

Patienten aus Toronto, da für sie keine Altersangaben vorlagen. Auch bei der Geschlechtsverteilung gab es Auffälligkeiten. New York war das einzige Zentrum, das einen höheren Anteil an weiblichen Patienten aufwies, in den anderen Zentren wurden dagegen mehr männliche als weibliche Patienten behandelt.

Als Grundlage des Vergleichs morphologischer Merkmale wurden die drei Elemente der für die Bestimmung des operativen Behandlungsrisikos häufig herangezogene Spetzler-Martin-Skala - Lokalisation der AVM in eloquenten Hirnarealen, Drainage der AVM und AVM-Größe - untersucht. Patienten aus Toronto wiesen einen signifikant höheren Anteil an AVMs mit einer Lokalisation in eloquenten Hirnarealen auf als Patienten anderer Datenbanken. Bezüglich der Drainage konnten keine signifikanten Unterschiede festgestellt werden. Für die Patienten aus Toronto lagen keine Angaben zur Drainage vor. Deutliche Unterschiede konnten bei der AVM-Größe festgestellt werden: Patienten aus Toronto wiesen signifikant häufiger kleine AVMs auf als die Patienten der anderen Datenbanken. AVMs können - außer als Zufallsbefund - durch eine intrakranielle Blutung, chronisch wiederkehrende Kopfschmerzen, fokale oder generalisierte Krampfanfälle, ein persistierendes, progredientes oder reversibles neurologisches Defizit auffällig werden.

Die Häufigkeit intrakranieller Blutungen unterschied sich in der bivariaten Analyse signifikant zwischen den Zentren. Die Anzahl von Blutungen war in Toronto höher als in den anderen Zentren. Dieser Zusammenhang wurde jedoch in der multivariaten Analyse nicht bestätigt. Ein persistierendes neurologisches Defizit wurde in dem New Yorker Zentrum häufiger festgestellt als in den anderen Zentren. Eine statistisch signifikante Häufung des Vorliegens eines durch multiplen kleinen Schädigungen durch Steal-Effekte oder klinisch inapparente Blutungen im Alter hervorgerufenen persistierenden neurologischen Defizits konnte dagegen nicht nachgewiesen werden. Der χ^2 -Test wies auch für den Zusammenhang zwischen reversiblen neurologischem Defizit und Zentrum auf signifikante Unterschiede hin, die Bonferroni-korrigierten Irrtumswahrscheinlichkeiten der Post-Hoc Analyse erlaubten jedoch keine Aussage, welche Zentren besonders betroffen waren. Bei den anderen genannten Präsentationsmodi waren

keine signifikanten Unterschiede zwischen den Zentren festzustellen. Für die Patienten aus Toronto lagen nur Angaben zur intrakraniellen Blutung vor.

4.2 Zentrums- und datenbankspezifische Einflussfaktoren

Für die zwischen den Zentren vorgefundenen Unterschiede ergeben sich verschiedene Erklärungsansätze. Neben dem Vorliegen eines Zuweisungsbias könnten insbesondere divergierenden Datenbankstandards und lokale Besonderheiten eine Rolle gespielt haben.

4.2.1 Zuweisungsbias

Vermutlich ist die Spezialisierung auf die Bestrahlungstechnik der Grund, dass der Datensatz aus Toronto eine relativ hohe Anzahl an Patienten mit kleinen AVMs und eine relativ niedrige Anzahl an Patienten mit einer mittelgroßen AVM enthielt. Wie bereits dargestellt wurde, ist die Chance auf eine erfolgreiche Obliteration der AVM umso größer, je kleiner die AVM ist. Dies dürfte zur Folge haben, dass ein Großteil der Zuweisungen an das Zentrum in Toronto Patienten mit eher kleine AVMs betreffen.

Der hohe Anteil an Kindern in dem Pariser Datensatz dürfte Ergebnis einer Spezialisierung auf die Behandlung pädiatrischer vaskulärer Erkrankungen an diesem Zentrum sein.

4.2.2 Divergierende Datenbankstandards

Eine Ursache für die Divergenzen in den Daten der AVM-Zentren könnte in der unterschiedlichen Definition wichtiger Variablen liegen. Einen Hinweis darauf gibt der Umstand, dass in der Toronto-Datenbank trotz der dort festgestellten großen Anzahl kleiner AVMs häufiger als anderswo AVMs in eloquenten Hirnregionen diagnostiziert wurden. Da die Wahrscheinlichkeit einer AVM, in einer eloquenten Hirnregion zu liegen mit zunehmender Größe ansteigt, muss angenommen werden, dass es zwischen den Datenbanken erhebliche Unterschiede in der Definition des Begriffes "Eloquenz" gibt.

Auch bei den klinischen Charakteristika werden Hinweise auf deutliche Unterschiede in den Datenbankstandards gefunden. Dass persistierende neurologische Defizite im New Yorker Datensatz signifikant häufiger festgehalten worden sind, könnte auf Unterschiede im Studienprotokoll - speziell der Erfordernis einer unabhängigen neurologischen Evaluation der Patienten in New York - zurückzuführen sein.

4.2.3 Lokale Besonderheiten

Die große Anzahl an Patienten über 60 Jahren in dem Berliner Datensatz ist wahrscheinlich auf den Zustrom ostdeutscher Patienten zurückzuführen, für die es aufgrund der begrenzten Ressourcen in den vergangenen Jahrzehnten keine Möglichkeit einer adäquater Diagnostik und Behandlung gab.

Der hohe Anteil an Frauen in der New Yorker Datenbank stellt ebenfalls eine lokale Besonderheit dar, die sich jedoch einer plausiblen Erklärung bislang entzieht.

4.3 Der Einfluss intervenierender Variablen auf das Merkmal intrakranielle Blutung

Die intrakranielle Blutung ist von ihrer klinischen Bedeutung, ihrer Häufigkeit als Präsentationsmodus und ihrer prognostischen Bedeutung für Rezidivblutungen als wichtigstes klinisches Merkmal dieser Untersuchung zu betrachten. Sie stellt das zentrale Glied in der Kette der Argumentation für den Einsatz einer der Behandlungsoptionen Operation, Embolisation oder Bestrahlung dar. Diese invasiven Verfahren sind mit einem nicht unerheblichen Behandlungsrisiko behaftet – wie eingangs berichtet liegt die Komplikationsrate bei Operationen zwischen 3% (niedriger Spetzler-Martin Wert) und 20% (hoher Spetzler-Martin Wert).

Angesichts der relativen Seltenheit der AVMs musste ein Mangel an prospektiven Langzeituntersuchungen von ausreichend großen, nicht selektierten und unbehandelten Stichproben konstatiert werden, die verlässlich Auskunft über die Auswirkungen intrakranieller Blutungen geben könnten. Es überwiegen teilweise retrospektive Untersuchungen kleiner Patientenkollektive.

Angesichts der für einen großen Teil der AVMs zudem eher günstigen Spontanprognose stellt sich die Frage nach der Berechtigung der Forderung einer grundsätzlichen interventionellen Behandlung von AVMs. Neuere Ansätze propagieren eine individuelle modellbasierte Risikoüberprüfung und Allokation einer interventionellen oder nicht-interventionelle Form der Therapie.

Vor diesem Hintergrund war die Feststellung, dass der bivariate Test auf signifikante Unterschiede zwischen den Zentren hinwies, von großer Bedeutung. Der Datensatz aus Toronto wies mehr intrakranielle Blutungen und zugleich mehr kleine AVMs auf als die anderen Datensätze. Aus der Literatur war jedoch bekannt, dass die AVM-Größe einen unabhängigen Effekt auf die Häufigkeit des Präsentationsmodus intrakranielle Blutung hat. Kleine AVMs werden offenbar häufiger durch Blutungsereignisse auffällig als große AVMs, wobei – wie eingangs erläutert - einiges dafür spricht, dass kein absolut höheres Risiko für kleine AVMs vorliegt, sondern dass die kausale Verknüpfung zwischen der Größe der AVM und dem Auftreten anderer Präsentationsmodi - wie zum Beispiel Krampfanfälle – hier interveniert.

Solange das Grundproblem der AVM-Forschung, dass in der Regel nur symptomatische AVMs überhaupt erkannt werden, bestehen bleibt, kann eine Aussage über Unterschiede zwischen den Blutungsraten der verschiedenen AVM-Größenklassen letztlich nicht getroffen werden, denn dadurch fehlt ein Bestandteil des Quotienten. Angesichts der geringen Prävalenz von AVMs scheint eine Lösung dieses Dilemmas, etwa durch systematische Reihenuntersuchungen mit bildgebenden Verfahren, jedoch nicht in Sicht.

Der in der Literatur beschriebene Zusammenhang zwischen Größe und Blutungshäufigkeit konnte auch für die vorliegende Studie gezeigt werden. Kleine AVMs bluteten signifikant häufiger als mittelgroße und große AVMs.

Als Konsequenz wurde der Zusammenhang zwischen dem Präsentationsmodus intrakranielle Blutung und den Zentren unter Berücksichtigung der intervenierende Variable „Größe der Blutung“ analysiert. Durch den Test auf bedingte Unabhängigkeit ließ sich nachweisen, dass kein

direkter Zusammenhang zwischen der Häufigkeit intrakranieller Blutungen und dem Ort ihrer Registrierung bestand.

Die Feststellung der Abhängigkeit der Blutungshäufigkeit von einer intervenierenden morphologischen Variablen, die ihrerseits durch das Vorliegen eines Zuweisungsbias bezüglich der Spezialisierung der Zentren auf bestimmte Behandlungsmethoden gekennzeichnet ist, verstärkt zum einen die Zweifel an der Berechtigung der Forderung einer grundsätzlichen interventionellen Therapie von AVMs.

Zum anderen bietet die Kenntnis dieser Abhängigkeiten die Möglichkeit, die intervenierende Variable durch geeignete Verfahren wie den modellbasierten Test auf bedingte Unabhängigkeit zu kontrollieren, dadurch zu besseren Schätzungen der tatsächlichen Inzidenz von intrakraniellen Blutungen und somit auch zu validen Modellen des individuellen Behandlungsrisikos zu gelangen, die zu einer verbesserten Allokation von interventionellen Maßnahmen wie Operation, intraarterielles interventionelles Verfahren oder Bestrahlung führen könnten.

4.4 Der Einfluss fehlender Werte auf die untersuchten Merkmale

Ob Unterschiede zwischen den Datenbanken und Zentren auch in Bezug auf die Qualität des Datenerhebungsprozesses bestehen, sollte durch eine Analyse der fehlenden Werte mittels eines multivariaten loglinearen Modells überprüft werden. Dabei wurde davon ausgegangen, dass fehlende Werte im Wesentlichen die Folge von individueller Nachlässigkeit oder unklar definierten Zuständigkeiten im Datenerhebungsprozess sind und somit keine inhärente Systematik vorliegt.

Das zunächst auf der Basis eines transformierten Datensatzes mit den Variablen „Merkmalsgruppe“ (X), „fehlender Wert“ (Y) und „Zentrum“ (Z) getestete Modell (XY,XZ,YZ) enthielt alle möglichen Haupteffekte und Interaktionseffekte mit Ausnahme der Drei-Wege-Interaktion. Der Fit dieses Modell erwies sich jedoch als nicht hinreichend.

Da einerseits das Hinzufügen der Drei-Wege-Interaktion zu dem trivialen Ergebnis eines perfekten Fits geführt hätte und andererseits das „Abspecken“ eines Modells entsprechend der Logik der „backward elimination“ keine Verbesserung der Anpassung erbringen konnte, waren die Möglichkeiten der Modellierung über das Hinzufügen oder Entfernen von Interaktionseffekten damit erschöpft.

Nach Inspektion der adjustierten Residuen des „keine Drei-Wege-Interaktion“-Modells konnten jedoch Kontraste definiert werden, die bei gleicher Zahl von Freiheitsgraden zu einem Modell mit einem insgesamt akzeptablen Fit führten. Dieses Modell legt lokale, auch innerhalb desselben Datenbankstandards „Paris-Berlin-Asien“ anzutreffende und die drei untersuchten Gruppen der demographischen, morphologischen und klinischen Merkmale umspannende Ursachen für die Abweichungen nahe.

So steht zu vermuten, dass in New York die klinischen oder morphologischen und in Paris die morphologischen Daten von dem aus Neurologen, interventionellen Neuroradiologen und Neurochirurgen bestehenden medizinischen Fachpersonal nicht lückenlos in die Datenbank eingegeben wurden, während die demographischen Daten möglicherweise im Rahmen eines zuverlässigeren Routineprozesses aus dem Krankenhaussystem entnommen wurden. Bei den Daten aus Asien scheint hingegen die Übernahme der demographischen Daten eher zu Auslassungen geführt zu haben als die Eingabe der klinischen und morphologischen Daten.

Es stellt sich auf dieser lokalen Ebene jedoch auch die bei statistischen Signifikanztests von großen Stichproben häufig auftretende Frage, ob dem hochsignifikant höher als erwarteten Anteil fehlender demographischer Werte im Datensatz aus Asien tatsächlich eine praktische Relevanz zukommt, da hier – absolut betrachtet - nur 13 Mal eine Auslassung festgestellt wurde (3% aller Angaben bzw. 8,5% der demographischen Werte aus Asien) und da demographische Variablen zudem typischerweise nur als erklärende Variablen eingesetzt werden.

Bei einem Anteil von 26,5% Beobachtungen unvollständiger Angaben in der Gruppe der klinischen Merkmale aus New York, von 19,6% in der Gruppe der morphologischen Merkmale

aus New York und von 16,8% in der Gruppe der morphologischen Merkmale aus Paris kann von einer praktischen Relevanz im Sinne einer vorhandenen Beeinflussung klinischer und morphologischer Parameter schon eher die Rede sein. Ein durch fehlende Angaben bedingter Bias könnte vorliegen, sollten die Eingabe-Unterlassungen nicht rein zufällig zustande gekommen sein, sondern einer unbekanntem Systematik folgen.

4.5 Anmerkungen zu den eingesetzten statistischen Methoden

Verfahren, die in der Beurteilung von randomisierten Studien eingesetzt werden, werden insbesondere nach ihrer Trennschärfe bezüglich klar definierter Endpunkte ausgewählt. Der Methodeneinsatz im Rahmen dieser Arbeit zielte dagegen eher auf die Identifizierung von Mustern der wechselseitigen Abhängigkeit zwischen den Daten und den Erzeugern dieser Daten.

Statt eines modellbasierten Tests auf bedingte Unabhängigkeit hätte im vorliegenden Fall beispielsweise auch der in der Medizinstatistik häufig verwendete Mantel-Haenszel-Test eingesetzt werden können, zumal er effizienter als das loglineare Modell ist und auch bei geringen Fallzahlen eingesetzt werden kann. Der modellbasierte Test bot sich jedoch an, da zum einen bei den gegebenen Stichprobengrößen nicht von einem Fallzahlproblem ausgegangen werden musste und zum anderen das modellbasierte Verfahren a priori mehr Möglichkeiten zur Überprüfung von Hypothesen bot. So hätte bei Vorliegen einer signifikanten Drei-Wege-Interaktion statt des bedingten Tests auf Unabhängigkeit der direkte Test $G^2[(XZ, YZ)]$ eingesetzt werden können.

Im Fall der Analyse der fehlenden Werte hätte auch die logistische Regressionsanalyse als Verfahren zur Analyse kategorischer, insbesondere dichotomer abhängiger Variablen auf die vorliegende Problemstellung angewandt werden können. Ein Einsatz der umgekehrten Helmert-Kontraste wäre ebenfalls möglich gewesen. Hier hat letzten Endes die einfache Darstellbarkeit der Häufigkeiten und adjustierten Residuen sowie die sich dadurch ergebenden Möglichkeiten zur inhaltlichen Interpretation der Resultate im Rahmen eines explorativen Vorgehens den Ausschlag für den Einsatz loglinearer Modelle gegeben.

4.6 Ausblick

Es konnte in der vorliegenden Arbeit gezeigt werden, dass zwischen den Zentren zum Teil erhebliche Unterschiede bezüglich relevanter demographischer, morphologischer und klinischer Parameter bestehen. Einige dieser Unterschiede konnten auf systematische Verzerrungen der Stichprobe aufgrund eines Zuweisungsbias zurückgeführt werden. Die Möglichkeit von Verzerrungen durch divergierende Standards, lokale Besonderheiten und fehlende Werte im Rahmen sich unterscheidender Qualitätsstandards wurden dargestellt und diskutiert.

Insbesondere das Vorliegen eines Zuweisungsbias dürfte nicht nur die Vergleichbarkeit der Resultate der gegenwärtigen, zentrenbasierter AVM-Forschung beeinträchtigen, sondern auch bei der Erforschung anderer seltener Krankheiten eine Rolle spielen.

Allgemein lässt sich bezüglich des Zuweisungsbias folgende Schlussfolgerung ziehen: Existieren **1)** bei einer bestimmten Erkrankung bevölkerungsbasierte Daten nicht in hinreichendem Umfang, lässt sich **2)** diese Erkrankung nur an entsprechend spezialisierten und technisch ausgestatteten tertiären Zuweisungszentren erfolgsversprechend therapieren; gibt es **3)** mehrere Behandlungsoptionen, die für bestimmte Untergruppen von Patienten unterschiedlich geeignet sind und **4)** unterscheiden sich wichtige klinische Merkmale in diesen Untergruppen, dann muss im Rahmen der krankenhausbasierten, klinischen Erforschung dieser Krankheit mit dem Vorliegen eines Zuweisungsbias in den Daten unterschiedlich spezialisierter Zentren gerechnet werden.

Sind die Zuweisungsmuster bekannt, die aus einer unterschiedlichen methodischen Spezialisierung der Zentren resultieren, sollte auch bei diesen Erkrankungen die Möglichkeit bestehen, einen resultierenden Zuweisungsbias durch den Einsatz geeigneter Verfahren zu korrigieren. Zumindest aber besteht die Möglichkeit, im Rahmen der Publikation der Forschungsergebnisse auf bekannte oder vermutete Abweichungen von der Grundgesamtheit aller Patienten hinzuweisen und die Resultate diesbezüglich zu diskutieren.

Als weitere praktische Konsequenz der hier dargestellten Ergebnisse sollten Datenbankbetreiber außerdem die Standards, nach denen sie vorgehen, nach Möglichkeit angleichen. Zumindest sollten die Standards einschließlich der getroffenen Maßnahmen zur Qualitätssicherung offengelegt werden.