

1 Grundlagen und Entwicklungslinien der Risikoadjustierung im Kontext medizinischer Bewertungen

Franz-Josef Fischer

Dieser Beitrag bildet die Entwicklung der Bedeutung einer Risikoadjustierung in Praxis und Forschung für die Medizin nebst ausgewählten Verfahren und Messmethoden/Skalen ab. Der Leser erfährt dabei viel über Methodenentwicklung und Bewertungsskalen zur Risikoanalyse und sollte staunen, wenn er erfährt, wie fortschrittlich gegenüber heute bereits die Reichsversicherungsordnung RVO³ von 1911 im Thema war. Der eher narrative Review wird an einigen Stellen mit konkreten Beispielen für ungewollte Desinformation und andererseits wachsender Transparenz der Operationalisierung medizinischer Prozess- und Ergebnisqualität ergänzt.

1.1 Entwicklung und Abgrenzung risikoanalytischer Bewertungsansätze

Bei der Frage der Risikoadjustierung medizinischer Evaluationsmethoden geht es primär um die Frage der Vergleichbarkeit von Ergebnisgrößen und der damit verbundenen Absicherung von Werturteilen und Handlungsempfehlungen unter eingeschränkt kontrollierbaren Bedingungen des Praxisalltags. Statistische Verfahren haben hier den Nachteil, dass diese für die Bewertung des Einzelfalles kaum ad hoc valide Ergebnisse liefern können. Daher wurden in der jüngeren Historie der Medizin immer wieder Verfahren entwickelt, die auf der Basis kumulierter großer Datenmengen eine – je nach Zweck und

Fragestellung unterschiedliche – prädiktive Aussage zur medizinischen Fallschwere und der damit verbundenen Risikoposition eines Patienten ermöglichen haben. Damit konnte auch für den Einzelfall schnell eine Einordnung getroffen werden, was insbesondere für kritische Behandlungsphasen ein entscheidender Vorteil ist. Ein weiterer Vorteil derartiger Verfahren war ein in der Zeitachse der Entwicklung eher geringer werdender Bedarf an varianzerklärenden Inputvariablen, wie das Beispiel des APACHE-Scores zeigt. In der Praxis spiegelt sich dies oftmals in einer Verbesserung der prädiktiven Aussagen durch

3 Reichsversicherungsordnung, gegründet 1911 (RVO); ab 1975 schrittweise integriert in das Sozialgesetzbuch SGB.

eher weniger Variablen wider, wie die aktuelle Europäische Leitlinie zur ACS-Diagnostik^{4,5} darlegt. Hier sind die prädiktiven Variablen auf lediglich 3 – aber sehr belastbare – Inputgrößen geschrumpft, was eine medizinische Bewertung hinsichtlich weiterer Behandlungsschritte erheblich erleichtert: 12-Kanal-EKG, hochsensitives TroponinT bzw. I und nachfolgend bei Bedarf eine echokardiographische Untersuchung. Die Risikoadjustierung des Einzelfalles gelingt durch eine weltweit kontinuierliche Weiterentwicklung entsprechender Scores nicht zuletzt deswegen, weil wir für die Weltbevölkerung den gleichen Anatomie-Atlas zu Grunde legen können. Dies mag zunächst banal erscheinen, ist aber für die Entwicklung von Systemen zur medizinischen Risikoadjustierung nicht unbedeutend, zumal dann die Ergebnisse auch über die Grenzen von Gesundheitssystemen (z. B. vergütungsrelevante und damit zum überwiegenden Teil an Prozeduren ausgerichtete DRG-Vorgaben) aus eher limitierten und hinsichtlich der Datenerfassungsmodalitäten divergierenden nationalen Quellen hinweg nutzbar sind. Ein weltweit einheitliches DRG-System dürfte wegen der sehr unterschiedlichen Rahmenbedingungen der Klinik-Finanzierung und Datenerfassung wohl eher ein Wunschtraum bleiben; weltweit einheitliche Systeme zur medizinischen Risikoadjustierung sind aber definitiv schon Realität. Natürlich bestimmt der jeweilige Zweck und das Ziel der spezifisch medizinischen Risikoadjustierung unterschiedliche Ausprägungen

dieser Systeme. Grundsätzlich können aber Systeme mit eher wenigen Patienten von der Entwicklungsarbeit in Gesundheitssystemen mit einer hohen Patientenzahl und entsprechend transparenter Datenbereitstellung profitieren. Dies gilt insbesondere bei prädiktiven Modellen auf der Grundlage von ICD-kodierten Diagnosen und der Bewertung des Einflusses von komplexen Komorbiditäten.

Natürlich hat die Methodik prospektiver und randomisierter klinischer Studien Ihren übergeordneten Stellenwert in Fragestellungen der medizinischen Forschung und ist für eine prospektive Evaluation, beispielsweise der Analyse der Wirksamkeit eines Arzneimittels oder der Effektivität eines medizinischen Verfahrens als Standard für Fragen der Zulassung oder Indikationserweiterungen von Therapieverfahren anzusehen. Dies gilt insbesondere dann, wenn es in der Testanordnung gelingt, alle relevanten Aspekte des Therapieerfolges, auch diejenigen, die von Patientenseite berücksichtigt werden müssen, in die Ergebnisbetrachtung einzubeziehen und damit eine möglichst hohe Kausalität der Ergebnisse jenseits einer reinen Korrelation der Ergebnisgröße von der Inputvariable zu erreichen. Dazu wird in der Regel ein Testverfahren eingesetzt, bei dem die Rahmenbedingungen unabhängig von weiteren erkennbaren oder auch nicht direkt erkennbaren externen Variablen möglichst unter Ausschluss des Zufalls kontrolliert werden können, so dass vom Ergebnis des Testverfahrens eine signifikante zweiseitige Kausalitätsbeziehung und eben nicht nur eine einseitige Korrelation zur Inputvariable hergestellt werden kann.

Der dazu meist als Goldstandard⁶ eingesetzte Test eines randomisierten (zufallsgesteuerten) und (wenn möglich) doppelten Verblindungsansatzes (Randomised Control Trial; RCT) zwischen einer interventionellen

4 Collet, J.-P. et al (2020): 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. *European Heart Journal* 42 (14): 1289–1367. <https://doi.org/10.1093/eurheartj/ehaa575>

5 Nef, H. M., Achenbach, S., Birkemeyer, R. et al. (2021): Manual der Arbeitsgruppe Interventionelle Kardiologie (AGIK) der Deutschen Gesellschaft für Kardiologie – Herz- und Kreislaufforschung e.V. (DGK). *Kardiologie* 15: 542–584. <https://doi.org/10.1007/s12181-021-00504-6>

6 Good Clinical Practice (<https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-clinical-practice>, Zugriff am 15.02.2024).

Verum- und nicht interventionellen Placebo-Gruppe, erlaubt in Abhängigkeit von Testumfang (ausreichende Probandenzahl da Mittelwertvergleich etc.), absoluter Effektgröße und entsprechender Verfügbarkeit der Ergebnisse im Idealfall (eingedenk einer vorher bestimmten akzeptablen Irrtumswahrscheinlichkeit) eine Darstellung einer Kausalbeziehung zwischen Input und Output; immer vorausgesetzt die Kontrollvariablen und Rahmenbedingungen konnten während des Experimentes bzw. Testansatzes konstant gehalten werden und alle relevanten Faktoren betreffend des Testergebnisses, waren den Studieninitiatoren (zumindest weitgehend) bekannt. Dies stellt eine valide statistische Auswertung von Studienergebnissen hinsichtlich zentraler Qualitätskriterien allerdings regelmäßig vor große Herausforderungen, da insbesondere in den über einen längeren Zeitraum aufgelegten prospektiven Studien häufig fehlende Werte durch Nichterhebung oder durch bereits aus der Studie ausgeschiedene Probanden bei der Studienaushwertung zu berücksichtigen sind. In derartigen Fällen sind durch potentiell ergebnisverzerrende statistische Korrekturen auch Einschränkungen für die Aussagekraft naheliegend⁷. Zudem ist für die beobachteten Effektgrößen häufig eine differenzierte Betrachtung der Ausgangsbasis bzw. eine eingeschränkte Homogenität der Vergleichspopulationen zu diskutieren. Konnte der vermutete Effekt definitiv nicht gefunden werden oder war er nur angesichts eines zu geringen Absolutwertes oder einer nicht ausreichenden Größe der Stichprobe nicht zu erkennen? War der Effekt unabhängig von zufälligen Ergebnissen? War die gewählte

Skala der Ergebnisdarstellung überhaupt geeignet, die Effekte für weitere Auswertungen zugänglich zu machen (z. B. Notwendigkeit einer Intervallskalierung für ausgewählte Testverfahren etc.)? Wurde hinsichtlich der Ergebnismessung berücksichtigt, dass unterschiedliche Patientengruppen (Alter, Geschlecht, Vorerkrankungen etc.) in der Analyse auch unterschiedlich betrachtet wurden und nicht fälschlicherweise in eine gemeinsame medizinisch-prädiktiv homogene Gruppe einbezogen wurden? Insbesondere kleine Effektgrößen lassen häufig Raum für zufallsgesteuerte Aussagen. Kurzum: Während wir für den Bereich prospektiver klinischer Studien über Werkzeuge verfügen, eine weitgehend belastbare Kausalitätsbeziehung darzustellen, ist dies bei Erhebungen im Praxisalltag medizinischer Interventionen oftmals nicht oder nur eingeschränkt möglich. Dies betrifft aber den weitaus größten Teil der Gesundheitsversorgung, die Ihrer Verantwortung nur gerecht werden kann, wenn andere Verfahren als z. B. RCT zur Risikoadjustierung der relevanten Patientenkollektive und eingesetzten Interventionen genutzt werden können, um eine valide Messung von Qualitäts- und Outcome-Parametern zu erhalten. Hier schlägt nun die Stunde der medizinischen Risikoadjustierung.

Jenseits von umfangreichen statistischen Verfahren zur Absicherung eines Studienergebnisses wird es in der von hoher Komplexität geprägten Realität eines medizinischen Praxisalltages aber kaum möglich sein, alle Voraussetzungen für eine gut kontrollierte »Laborsituation« einer RCT sicherzustellen. Dies sollte aber dann nicht zur Konsequenz haben, dass in diesem Bereich keine Erkenntnisgewinnung mehr möglich ist. Denn oftmals sind es ja gerade praxisnahe Situationen, die eine hohe Relevanz für den Krankheitsalltag von Patienten haben und damit auch das Interesse der Fachöffentlichkeit wecken. Dies dürfte eines der zentralen Aufgabenfelder einer Risikoadjustierung sein, die dort einsetzt, wo sowohl der Einfluss der medizini-

7 Zur Missing Value-Problematik s. a. Urban, D., Mayerl, J., Wahl, A. (2016): Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder Nicht-Imputation? Eine Anleitung für die Regressionspraxis mit SPSS. Stuttgart: Universität Stuttgart. (<https://www.sowi.uni-stuttgart.de/dokumente/forschung/siss/SISS44.pdf>, Zugriff am 15.02.2024).

schen Intervention selbst als auch die Einflussfaktoren, die ein Patient für das Behandlungsergebnis mitbringt, unabhängig voneinander bewertet und nach Möglichkeit in eine Kausalbeziehung zum Behandlungsergebnis gebracht werden müssen.

Wenn nun in einer Klinik für eine ausgewählte Patientengruppe eine Inhouse-Mortalitätserwartung zum einen korrekt berechnet und zum anderen in einer kausalen Abgrenzung der Einflussfaktoren gegen den beobachteten Wert abgeglichen werden soll, wird erkennbar, welchen Einfluss eine bestmögliche Risikoadjustierung auf das Qualitätsurteil hat. So sollte durch den Einbezug aller relevanten Komorbiditäten und Komplikationserwartungen und deren Interaktionen die Morbiditätssituation des Patienten ebenso umfassend beschrieben werden wie die Risikopotentiale der medizinischen Prozedur, die auch unabhängig von der Morbiditätslast des zu behandeln Patienten zu bewerten ist. Aber auch weitergehende Risikofaktoren wie die Dauer der Intervention, die Beherrschbarkeit von Komplikationen durch vorhandene oder nicht vorhandene Medizintechnik sind hin-

sichtlich einer Kausalbeziehung zu berücksichtigen. Sollen all diese Einflussfaktoren hinsichtlich einer Mortalitätserwartung berücksichtigt werden, steht die Frage von Messoptionen und geeigneten Skalen im Mittelpunkt. Denn ein Qualitätsurteil im Bereich medizinischer Outcomes⁸ ist in den seltensten Fällen eine Bewertung analog technischer Vorgaben im Sinne von erfüllt und/oder nicht erfüllt. Hier sind Aussagen erst dann belastbar, wenn auf der Patientenseite durch die risikoadjustierte Bildung homogener klinisch-prädiktiver Gruppen geeignete Vergleichskollektive gebildet wurden, die zufallsunabhängig eine Aussage zur Kausalität erlauben. Erst mit Hilfe dieser Auswertungstechniken kann dann belastbar unterschieden werden, ob der Patient verstorben ist, weil der ärztliche Eingriff mangelhaft durchgeführt wurde oder aber weil der Patient ein zu hohes initiales Behandlungsrisiko mitbrachte. In vielen Fällen ist aber der Grund für die schlechten Outcomes weder auf der Seite der Klinik noch auf der Seite des Patienten, sondern auf Grund einer weiteren Einflussquelle gegeben, nach der dann gesucht werden muss.

1.2 Die Bedeutung der Risikoadjustierung für das medizinische Qualitätsmanagement

Durch den Einsatz von geeigneten Verfahren zur Risikoadjustierung verändert sich die Blickrichtung des medizinischen Qualitätsmanagements ausgehend von der Betrachtung einer Prozess- und Ergebnisqualität über die nominalen Messergebnisse im »Zähler«

einer Verhältniszahl hinaus auf den »Nenner« des Qualitätsurteils. Ist z. B. die Mortalität in einem ausgewählten Fachbereich numerisch gleich zwischen einer Klinik der Maximalversorgung und einem Krankenhaus der Regelversorgung, beruhigt dies die Klinik der Re-

8 Die in diesem Kontext immer wieder auftretende Fragestellung nach einer objektiven Qualitätsbewertung (s. a. Lauterbach-Interview in der Talk-Show Markus Lanz im Juli 2023, der die irreführende Behauptung einer einfachen Gegenüberstellung der medizinischen Ergebnisse als ausreichend betrachtete) ist von hoher Bedeutung sowohl für gesundheitspolitische Fragestellungen wie auch für Haftungsaspekte. Genau dies hatte in den USA der 80er und 90er Jahren eine Welle von unberechtigten Klagen gegen Kliniken ausgelöst.

gelversorgung sicherlich zunächst. Leider kann das Urteil über die Qualität der Versorgung nach einer adäquaten Risikoadjustierung danach auch grundsätzlich anders ausfallen. Wenn im Klinikum der Maximalversorgung durch die Positionierung der Einrichtung als höchste Versorgungsstufe für Akutfälle wesentlich mehr komplexe und damit komplikationsträchtige Fälle eingeliefert wurden, kann durchaus auch eine höhere Mortalität erwartet werden. Dann hat die Klinik der Regelversorgung ein Problem. Es ist aber auch denkbar, dass die höheren Risiken durch weitere Faktoren (z. B. Erreichbarkeit) bedingt auch in ausreichender Zahl im Klinikum der Regelversorgung eintreffen, womit diese Klinik hinsichtlich des Outcomes aufgewertet wird. Wenn nun weder die Klinik der Maximalversorgung noch die Klinik der Regelversorgungsstufe alle relevanten Komorbiditäten der aufgenommenen Patienten kodieren oder keine geeignete Methode zur Risikoadjustierung eingesetzt wird, ist hinsichtlich der Mortalitätsbewertung eigentlich gar keine Aussage mehr möglich, da keine prädiktiv valide Bezugsgröße für das Gesamtkollektiv (auch haftungsrechtlich) belastbar ermittelt werden kann.

Mit dem Ziel höherer Objektivität richten wir mit den Techniken der Risikoadjustierung den Blick nun eher auf den Nenner der Aussage. Waren die Ausgangssituationen zwischen zwei Patienten hinsichtlich der vorliegenden Risiken wirklich identisch oder gab es Abweichungen, die für das Ergebnis einer medizinischen Intervention oder Therapie von hoher, u. U. kausaler Relevanz waren. Kann diese Frage nicht eindeutig beantwortet werden, können auch keine diesbezüglichen Bewertungen valide vorgenommen werden. Medizinern ist diese Voraussetzung nur allzu bewusst: »My patients are sicker!«⁹ ist eine oft

zu hörende Aussage von Ärzten im Kontext von Outcomes-Evaluationen, die zwingend objektiv bewertet werden muss – sei es, um den verantwortlichen Mediziner hinsichtlich der Behandlungsprivilegien für seinen Patienten zu unterstützen, oder sei es, um ihm wegen der nicht Verifizierbarkeit seiner Aussage zusätzliche Ressourcen für die Behandlungsführung in Abrede zu stellen.

Dass die Bewertung der ärztlichen Aussage »eines besonders schwer erkrankten Patienten« (die ihm damit auch ein höheres Behandlungsrisiko aufbürdet) keine triviale Aufgabe ist, kann ein einfaches Zahlenbeispiel verdeutlichen. Nehmen wir an, dass dem fraglichen Patienten jeweils eine Hauptdiagnose in drei unterschiedlichen medizinisch-prädiktiven Schweregraden zugeordnet werden könnte. Dazu kommen dann drei Nebendiagnosen¹⁰, die ebenfalls in jeweils drei unterschiedlichen Schweregraden vorliegen können und alle relevant für das Ergebnis der Behandlung sein können. Die möglichen Beeinflussungen der Schweregrade der Nebendiagnose untereinander sind zudem zu berücksichtigen. Im vorliegenden Fall könnte der Patient bereits bei dieser noch überschaubaren Ausgangssituation in 27 (!) unterschiedlich medizinisch-prädiktive Fallkomplexitäten eingeordnet werden. Nun werden sicherlich nicht alle 27 möglichen Cluster die gleiche Therapierelevanz besitzen, so dass u. U. nur ein Drittel der Cluster als letztendlich relevant eingeordnet werden muss, aber neben dieser einen ICD-10-Hauptdiagnose gibt es weitere ca. 16.000 ICD-10-WHO-Diagnosen, die zusätzlich zum großen Teil neben einer Hauptdiagnose auch als relevante Nebendiagnosen in Frage kommen können. Es ist leicht einzusehen, dass hier die Instrumente zur Risikoadjustierung hinsichtlich der

9 Übersetzt: »Meine Patienten sind kränker« (d. h., sie stellen höhere medizinische Risiken dar).

10 Ist eher defensiv geschätzt, da die Zahl der Nebendiagnosen bei Patienten > 75 Jahre in > 80 % der Fälle bei ca. acht Nebendiagnosen liegt.

Patientenzahl, auf deren Basis diese entwickelt wurden, sehr umfangreich ausfallen müssen, denn die Anzahl der Nebendiagnosen ist insbesondere bei älteren Patienten in der Regel deutlich größer als drei. Damit entstehen für die begleitende Statistik zur Fundierung der entsprechenden Risikoadjustierungsmodelle wirklich »gewaltige« Herausforderungen hinsichtlich der Verfügbarkeit von Patientendaten, die allein auf nationaler Ebene wegen der Begrenztheit der Datenvolumina¹¹ oftmals gar nicht zu leisten sind.

Lenken wir nun den Blick einmal auf eine traditionelle Outcomes-Bewertung im medizinischen Qualitätsmanagement¹², die ebenfalls ohne umfassende Risikoadjustierung schnell ihre Aussagekraft verlieren kann. Zudem ist wegen der kleinen Effektgrößen im nachfolgenden Beispiel die Ergebnisbetrachtung hinsichtlich des Ausschlusses von rein zufallsbedingten Effekten nur über einen längeren, mehrjährigen Zeitraum überhaupt erst möglich. Von 100 Knie-TEPs¹³ wurden im Jahr 2020 in Deutschland laut GBA-Berichterstattung 1,3 Knie-TEPs – gleiches prädiktives Risiko vorausgesetzt – wegen Komplikationen erneut operiert¹⁴. Um signifikante Qualitätsunterschiede und nicht reine Zufallseffekte zu beschreiben, muss ein Krankenhaus zur

Dokumentation einer signifikant besseren Versorgung aber überhaupt erst mindestens 447 Knie-TEPs pro Jahr durchführen.¹⁵ Diese Fallzahl erreichen aktuell aber nur ca. 2,5 % aller Kliniken in Deutschland. Dies lässt erkennen, dass in vielen medizinisch-qualitätspolitischen Fragestellungen und Klinikbewertungen (Mortalität, Komplikationen, seltene Therapieversager etc.) ohne ausreichende Risikoadjustierung und bei kleinen Erwartungswerten eine Bewertung von Behandlungsergebnissen auf einer Datenbasis von nur einem Jahr sehr oft gar nicht möglich ist. Je kleiner die entsprechende Klinik und je weniger spezialisiert diese ist, desto schwieriger werden Aussagen zum Qualitätsstandard in der betreffenden Klinik ohne adäquate Risikoadjustierung. Dies betrifft aber leider auch eine positive Beweisführung der betreffenden Krankenhäuser, die nur mit einer ausgeprägten Risikoadjustierung überhaupt Aussagen zur eigenen, unterstellten »guten« medizinischen Performance belegen können. So ist es ohne adäquate Risikoadjustierung einer Klinik bis ca. 20.000 stationären Behandlungsfällen insgesamt im Jahr eigentlich gar nicht möglich, ihre gute Qualität in den einzelnen Fachbereichen und Leistungen durch belastbare Outcomes unabhängig vom Zufall und dem Einfluss der patientenseitigen Morbiditätslast darzu-

11 In Deutschland stehen jährlich ca. 19 Mio. Patientendaten für die Auswertungen des INeK zur Verfügung. Im Vergleich mit dem US-amerikanischen Disease Staging (Jefferson Medical School), dessen Entwicklung auf > 300 Mio. Patientendaten aufsetzt, würde die Eigenentwicklung einer »deutschen« Skala hier ca. 20 Jahre in Anspruch nehmen. Da ist es schon vorteilhaft, dass jeder 6. US-Amerikaner deutsche Vorfahren hat.

12 Nach Aussage des BMG künftig ein zentraler Faktor zur Marktberichtigung von Anbietern im Sinne von Kliniken oder Fachabteilungen.

13 TEP steht für Total endoprothetische Versorgung (kompletter Gelenkersatz).

14 Gemeinsamer Bundesausschuss (2020): Geschäftsbericht 2020. (https://www.g-ba.de/downloads/17-98-5148/2021-07-01_G-BA_Geschaeftsbericht_2020_bf.pdf, Zugriff am 15.02.2024).

15 Mit steigendem Anteil von schlechten Outcomes, schrumpft die Voraussetzung einer großen Grundgesamtheit für die Signifikanzermittlung. Ist die Komplikationsrate z. B. bei 3 %, bedarf es nur noch ca. 200 vergleichbarer Interventionen. Bei einer Komplikationsrate von 10 % reichen schon ca. 50 Fälle aus, um die Zufallsunabhängigkeit des Outcomes belastbar darzustellen. Es kommt aber sehr (!!) selten vor, dass Ärzte eine so schlechte Arbeit machen und der Grund allein auf der Seite der Behandlung liegt. Ohne qualifizierte Risikoadjustierung laufen daher Ansätze zur Qualitätssicherung oftmals ins Leere und verunsichern Patienten und Leistungsträger. In den USA wurden auf dieser Basis viele öffentlichen Portale mit Klagen der betroffenen Kliniken konfrontiert und erhebliche Schadensersatzleistungen erzielt.

stellen. Ohne umfassende Risikoadjustierung und daher mithilfe eher prozedural aggregierten prädiktiven Clustern gelingt dies meist nur den größeren Kliniken und den leistungsspezifisch stark eingegrenzten Fachkliniken, die entsprechende statistische »Power« der Vergleichspopulationen zum Benchmark durch die prozedurale Basis der Cluster und darin enthaltenen allgemein anzunehmenden Risikoverteilungen bereits auf Basis eines einzelnen Auswertungsdatenjahres besitzen. Dennoch erfreuen sich Klinik-Benchmarks mit Outcome-Bezug und ohne umfassende Risikoadjustierung sowie diesbezügliche Jahresvergleiche großer Beliebtheit. Sie sind aber für eine objektive Diskussion hinsichtlich der Bewertung medizinischer Outcomes nicht nur ungeeignet, sondern potentiell gefährlich, da sie insbesondere in der Laienöffentlichkeit erhebliche Wahrnehmungsverzerrungen und ungünstige Voreinstellungen bzw. Vorverurteilungen von Kliniken verstärken können.

Vergleiche mit Aussagen zu »guten und schlechten Kliniken oder Ärzten« in der Laienpresse (Focus, FAZ, Talkshows im öffentlichen Fernsehen u. a.) haben generell, wie alle retrospektiven Untersuchungsansätze (Fall-Kontroll-Studien etc.), immer das Problem, dass die absolute *Vergleichbarkeit sowohl der Ausgangsbasis wie auch des Ergebnisses* zwischen Individuen, untersuchten Populationen oder überprüften Kliniken gewährleistet werden muss. Dies stellt besondere Anforderungen an die im Verfahren bereitgestellten Informationen und insbesondere an diejenigen Informationen, die den Gesundheitszustand des Patienten und seine spezifischen Risikofaktoren auch in deren Abhängigkeit untereinander betreffen. Grundsätzlich könnte eine Klinik ein gutes Rating für ihre Leistungen schon allein dadurch erlangen, dass in bestimmten Fachbereichen primär Patienten mit einer prognostisch signifikant geringeren medizinischen Fallschwere bzw. Mortalitätserwartung aufgenommen werden und die komplexeren Fallschweren nach Möglichkeit an die nächsthöhere Versorgungstufe weitergeleitet wer-

den. Dies führt dann in Klinik-Benchmarks regelhaft zu einer Aufwertung der risikoaversen Versorgungseinrichtungen und auf der anderen Seite zur höheren Wahrscheinlichkeit für Kritiken an der Behandlungsperformance bei Einrichtungen der Schwerpunkt- und Maximalversorgung, die die betreffenden Risiken ja meist nicht mehr weiterverlegen können. Natürlich würde jede Klinik nun diesen Verdacht weit von sich weisen. Doch mittlerweile haben sich ähnlich wie in den USA auch auf EU-Ebene Vereine und Stiftungen gegründet, deren Ziel es ist, mit validierten Methoden zur Risikoadjustierung Kliniken ohne hohe Morbiditätslast zu detektieren und auf der anderen Seite eine Empfehlung zur Förderung derjenigen Kliniken abzugeben, die gezielt höhere medizinische Risiken erfolgreich behandeln¹⁶. Werden also unterschiedliche Kliniken in einen Performance-Vergleich einbezogen, ist eine adäquate maximale Risikoadjustierung das oberste Gebot für einen validen Vergleich, insbesondere dann, wenn an dem betreffenden Benchmark auch noch ein fachliches oder öffentlichkeitsbezogenes Reporting hängt. Ist dies nicht gegeben, ist die Gefahr einer manipulativen Instrumentalisierung des Qualitätsbegriffes naheliegend (z. B. vermeintlich »gute« Prozess- und Ergebnisqualität medizinischer Interventionen ohne Einbezug der Morbiditätslast der Patienten). Die durchgeführte Risikoadjustierung muss also für den Bereich des gesamten stationär-ambulant Versorgungskontextes mind. die beiden folgenden Kriterien erfüllen:

- Sie muss das Risiko des Patienten für das Behandlungs-Outcome während des Klinikaufenthaltes bzw. in einem ambulanten Setting in einer *mindestens ordinal-kalibrierten Skala* abbilden können, und

16 S. a. EuroSCORE, asbl, Luxemburg/Brüssel (gemeinnützige Stiftung) u. a.

- die *Methodik der Klassifikation muss für Dritte absolut transparent* sein, so dass zumindest ein Facharzt genau nachvollziehen

kan kann, warum die eingesetzte Risikobewertung genau zu diesem Ergebnis kommt.

1.3 Risikobewertung im Zentrum einer ambulanten Behandlungsverlagerung

Es erscheint selbstverständlich, dass eigentlich alle relevanten Informationen schon zu Beginn der Entscheidung für oder gegen eine stationäre Aufnahme eines Patienten vorliegen sollten und dadurch die Entscheidung der Mediziner eine maximale Absicherung erfährt. In Versorgungssystemen, in denen aber erst anl. der Entlassung eines Patienten alle relevanten Informationen zur Abrechnung vorliegen müssen (häufig bei Systemen mit DRG-basierter Leistungsabrechnung gegeben), wird dieses Ziel häufig erst im weiteren Verlauf einer Behandlung erreicht. Belegarztsysteme wie in den BeNeLux-Ländern, den USA oder in Skandinavien haben hier offensichtlich einen zentralen Wettbewerbsvorteil gegenüber sektoral stärker getrennten Versorgungsansätzen (wie in Deutschland oder Frankreich), was sich in einem deutlich höheren prozentualen Anteil ambulant versorgter Patienten gegenüber stationär aufgenommenen Patienten äußert¹⁷. Aber auch bereits diese Aussage bedarf einer maximalen Risikoadjustierung, hinsichtlich aller patientenbezogenen Faktoren, denn sonst könnte sofort die Kritik aufkommen, dass hier unterschiedliche Risiken gegenübergestellt und folglich

»Äpfel mit Birnen«¹⁸ verglichen werden. Da wir – epidemiologisch vorteilhaft – nun alle dem gleichen Anatomie-Atlas entsprechen, muss also die Nichtvergleichbarkeit der Benchmark Populationen durch noch fehlende Bereitstellung von Informationen im Aufnahmeprozess (Komorbiditäten, Laborwerte, ASA-Bewertungen etc.) bedingt sein. Für Gesundheitssysteme, die anl. einer stationären Aufnahmeentscheidung bisher eine eher retrospektive Risikoanalyse betrieben haben, bedeutet diese Vorgabe einen erheblichen Umstellungsprozess in der Datenbereitstellung. Die Forderung der Datenerhebung nach der Vorgabe »present-on-admission« ist nicht ohne Grund eine zwar eher banal klingende, aber umso mächtigere Forderung einer Patientenbeschreibung, wenn eine valide Entscheidung hinsichtlich der Ablehnung einer stationären Aufnahme juristisch verteidigungsfähig sein soll. Sich allein auf die Art der durchzuführenden Prozedur zu beziehen, ist keine Lösung, denn häufig bedingt das Vorliegen von Komorbiditäten oder situativen Beeinträchtigungen (ASA-Score etc.) eine stationäre Aufnahme von ansonsten hinsichtlich der medizinischen Komplexität überschaubaren Eingriffen. Die Ablehnung einer stationären Aufnahme stellt daher wesentlich höhere Anforderungen an eine maximale Erfassung aller Risikofaktoren als eine Ent-

17 Beispiel: Klinikum der Université Libre de Bruxelles (ULB; Hôpital Erasme in Brüssel-Anderlecht), in dem p. a. ca. 25.000 Patienten stationär, aber > 350.000 Patienten ambulant versorgt werden. S. a. À propos de l'hôpital (<https://www.erasme.ulb.ac.be/fr/a-propos-de-l-hopital>, Zugriff am 15.02.2024).

18 In der US-amerikanischen Literatur wird regelmäßig der Vergleich zwischen Äpfeln und Orangen (apples and oranges) formuliert.

scheidung zur stationären Aufnahme, da dann immer noch Zeit für die Bereitstellung der Informationen verbleibt und zudem der wichtige diagnostische Hilfsfaktor »Zeit« ausgenutzt werden kann.

Neben den Verfahren zur Risikoadjustierung hinsichtlich der Zuordnung von stationären bzw. ambulanten Therapieführungsoptionen spielen im Klinikalltag Skalen und Verfahren eine große Rolle, die in der Lage sind, prognostische Verbesserungen aber auch Verschlechterungen qualitativ und auch quantitativ in einem kurzen Zeithorizont darstellen zu können. Die Beispiele wie APA-

CHE IV^{19,20}, ASA-Score²¹, Apgar-Index²², SAPS²³- und TISS²⁴- und SOFA²⁵-Score werden im Folgenden kurz erläutert – sie konzentrieren sich primär auf Vitalparameter und kommen dabei häufig auch ohne den weitergehenden Einbezug von Diagnosen aus. Insbesondere ein höherer ASA-Score dürfte für deutsche Kliniken in der aktuellen Bewegung hin zu einem verstärkten ambulanten Operieren ein wichtiges Kriterium für die Absicherung einer entgegen der Normempfehlung dennoch erfolgten stationären Aufnahme des Patienten sein.

1.4 Historische Entwicklung der medizinischen Risikoadjustierung

»Prove and Improve« (Überprüfen und Verbessern) im Sinne der Verbesserung der Behandlungsprognose nach messbaren und fallsunabhängigen Ergebnisgrößen ist in allen Gesundheitssystemen die zentrale Forderung hinsichtlich der Akzeptanz qualitätspolitischer Vorgaben. Um dieses Ziel zu erreichen, wurden seit Beginn der 60er Jahre des vorigen Jahrhunderts insbesondere in den USA und

auch in Europa eine Vielzahl von Risikoadjustierungsansätzen für medizinische Entscheidungen entwickelt. Diese unterscheiden sich in vielerlei Hinsicht, können aber im Wesentlichen in drei Gruppen eingeteilt werden:

1. Risikoadjustierungssysteme zur Prognose der Entwicklung eines Patienten in kurzer

-
- 19 Markwart, R. et al. (2020): Epidemiology and burden of sepsis acquired in hospitals and intensive care units: A systematic review and meta-analysis. *Intensive Care Med.* 46: 1536–1551.
 - 20 Vincent, J. L. et al. (2014): Assessment of the worldwide burden of critical illness: The intensive care over nations (ICON) audit. *Lancet Respir. Med.* 2: 380–386.
 - 21 Doyle D. J., Hendrix J. M., Garmon, E. H. (2024): American Society of Anesthesiologists Classification. StatPearls Januar. (<https://www.ncbi.nlm.nih.gov/books/NBK441940/>, Zugriff am 15.02.2024); Committee on Economics (2014): Statement on ASA Physical Status Classification System. American Society of Anesthesiologists. (<https://www.asahq.org/standards-and-practice-parameters/statement-on-asa-physical-status-classification-system>, Zugriff am 15.02.2024).
 - 22 Apgar score (<https://medlineplus.gov/ency/article/003402.htm>, Zugriff im Juli 2023).
 - 23 SAPS II – Simplified Acute Physiology Score (<https://www.divi.de/empfehlungen/publikationen/intensivmedizin/saps-ii-simplified-acute-physiology-score>, Zugriff am 15.02.2024).
 - 24 TISS 28 – Therapeutic Intervention Scoring System (<https://www.divi.de/empfehlungen/publikationen/intensivmedizin/tiss-28-therapeutic-intervention-scoring-system>, Zugriff im Juli 2023).
 - 25 Sequential Organ Failure Assessment (SOFA) Score (<https://www.mdcalc.com/calc/691/sequential-organ-failure-assessment-sofa-score>, Zugriff im Juli 2023).

Zeitperspektive, z. B. Mortalitätserwartung, Komplikationsentwicklung etc.

2. Risikoadjustierungssysteme, die auf die Festlegung einer zumindest mittelfristig stabilen prädiktiven Gesamt-Krankheitsschwere ausgerichtet sind und unabhängig von Prozeduren und ärztlichen Interventionen sind
3. Risikoadjustierungssysteme speziell für versicherungsmathematische Berechnungen (Prämienkalkulation, Gegenfinanzierung von ausgewählten Therapieoptionen etc.)

Dies soll nun im Detail näher betrachtet werden:

1. Risikoadjustierungssysteme zur Prognose der Entwicklung eines Patienten in kurzer Zeitperspektive, z. B. in den kommenden Tagen (Mortalität, Komplikationsentwicklung etc.).

Hier geht es fast ausnahmslos um Patienten in einer kritischen Behandlungssituation, in der nicht so sehr die Diagnose im Zentrum steht, sondern Vitalfunktionen, Befundparameter aus Laborwerten und Bildgebung und ggf. Komorbiditäten und daraus erwachsender zeitnaher Risiken wie beispielsweise Gerinnungskomplikationen bei onkologischen Patienten. Beispiele für derartige Scores wären Glasgow Coma Scale²⁶, APACHE IV, SAPS- und TISS-Score oder auch der SOFA- bzw. der ASA-Score.

2. Die zweite Gruppe bilden Risikoadjustierungssysteme, die auf die Festlegung einer

prädiktiven Gesamt-Krankheitsschwere ausgerichtet sind und dieses Ziel insbesondere durch die Berücksichtigung möglichst aller Komorbiditäten eines Patienten und daraus resultierender Komplikationserwartungen in eher mittelfristiger Perspektive erreichen wollen.

Die resultierende Clusterung besitzt, durch den fast alleinigen Rückgriff auf Diagnosen und unter Einbezug einer eher geringen Anzahl von Prozeduren, beste Voraussetzungen, um die prädiktive Fallschwere unabhängig vom weiteren Behandlungsverlauf, dem Ressourcenverbrauch und der Vergütung abzubilden. So kann insbesondere der zeitpunktbezogene prädiktive Gesundheitszustand eines Patienten hinsichtlich einer kausalen Betrachtung des Behandlungsergebnisses unabhängig von den betreffenden Interventionen selbst und unabhängig von Art und Größe der behandelnden Einrichtung objektiviert werden. Die Haupteinsatzgebiete für derartige Scores (RAND – Charlson-Index, Elixhauser-Index, Disease Staging u. a.) sind die forensisch wichtige Nenner-Adjustierung für Qualitätsmessungen, die Begleitung von transsektoralen Versorgungsansätzen (wenn diese Skalen auch für den ambulanten Einsatz genutzt werden können wie z. B. Disease Staging) und die Verteidigung der Kliniken und der Ärzteschaft hinsichtlich nicht gerechtfertigter Anschuldigungen im Sinne von Schlechtbehandlung (Physician Malpractice²⁷). Sie werden häufig mit dem Akronym MCC (Multiple Chronic Conditions²⁸) beschrieben. Grundsätzlich gehören auch die überwiegend prozedural-orientier-

26 Teasdale, G., Jennett, B. (1974): Assessment of coma and impaired consciousness. A practical scale. In: *Lancet* 13 (2): 81–84. DOI: 10.1016/s0140-6736(74)91639-0; Abdulla, W. (1999): Interdisziplinäre Intensivmedizin. München u. a.: Urban & Fischer, S. 467; Leitlinien für Diagnostik und Therapie in der Neurologie. 4. Auflage. Georg Thieme Verlag, Stuttgart 2008, S. 654 ff.

27 Bal, B. S. (2009): An introduction to medical malpractice in the United States. *Clin Orthop Relat Res.* 467(2): 339–347. doi:10.1007/s11999-008-0636-2

28 Iezzoni, L. I. (Ed.) (2012), *Risk Adjustment for Measuring Health Care Outcomes*, 4th Ed., Health Administration Press, Chicago IL, S. 53 ff.