

Der Score Bias

L. Dubs

Definitionen

Ein *Score* ist ein numerisches Bewertungssystem, das den Zustand eines Patienten mittels Punktwerten für ausgewählte Kriterien und einer daraus resultierenden Gesamtpunktzahl zu einem bestimmten Zeitpunkt beschreiben soll [1].

Ein *Bias* (aus dem Englischen übersetzt: Schräglage/Schiefelage) ist ein systematischer Denkfehler, ein unkontrollierter Einfluss oder eine verdeckte Voreingenommenheit innerhalb wissenschaftlicher Studien [2–4].

Der *Score Bias* ist die Vermischung von Messwerten der Organebene mit solchen der Individuumebene, bzw. die Vermischung der Schweregradeinschätzungen von Schädigungen mit solchen der Fähigkeitsstörungen (ICIDH-Modell).

Die *ICIDH-Ebenen* «Organ – Individuum – Gesellschaft» sind Grundlage für die seit 1980 bestehende Internationale Klassifikation der Schädigungen, Fähigkeitsstörungen und Beeinträchtigungen (ICIDH = International Classification of Impairments, Disabilities and Handicaps). Dieses WHO-Handbuch ist 1996 in deutscher Sprache erschienen [5] und liefert die Klassifikationsgrundlagen für das Management chronischer Erkrankungen bzw. der Krankheitsfolgenmanifestationen.

Probleme mit den Kniescores

Zur Beurteilung eines Gesundheitszustandes oder eines Behandlungsergebnisses werden in der Medizin häufig Scorewerte angegeben und mit der Steigerung von Scorewerten einem Behandlungserfolg gleichgesetzt. Hefti hat 1996 [6] auf den möglichen «Selbstbetrug für Fortgeschrittene» hingewiesen und darauf aufmerksam gemacht, dass es kuriose Wege geben kann, mit Scorebewertungen zu einem guten Resultat zu kommen. Am Beispiel der Scores für Knieinstabilitäten nach Verletzungen bzw. Operation des vorderen Kreuzbandes soll gezeigt werden, wie gefährlich die Anwendung eines oder mehrerer Kniescores bei der Resultatbeurteilung sein kann, wenn man die Schichtungen der ICIDH-Klassifikation nicht beachtet. Was interessiert den Patienten? Er möchte eigentlich nur wissen, was er wieder kann wie z.B. Treppensteigen, Kauern, Laufen, Wandern über län-

gere Distanzen. Er interessiert sich also nur für seine Fähigkeiten bzw. Fähigkeitsstörungen («disabilities») auf der mittleren Ebene der ICIDH (Individuumebene) bzw. für die dadurch entstehende Benachteiligung in der Gesellschaft. Den Arzt interessieren natürlich auch seine Befunde auf Organebene wie Bewegungsmessung, Stabilitätsprüfung, Röntgenbildwerte, Schwellungsgrad und die Beurteilung einer Deformität. Diese einzelnen Parameter werden mit einer von Score zu Score unterschiedlichen Punktzahl versehen, wobei die Gesamtpunktzahl meistens 100 beträgt. Dem Patienten kann es letztlich gleichgültig sein, wie die Messwerte auf Organebene ausfallen, solange er seine gewünschten Fähigkeiten erkennen kann. In der gleichen Punktesummation sind also Werte aus zwei (selten drei) verschiedenen ICIDH-Ebenen enthalten. Es ist einleuchtend, dass der Erfinder eines Scores davon ausgegangen ist, dass eine Verbesserung eines Organmesswertes automatisch zu einer Verbesserung der Patientenfähigkeiten führen müsse.

Abbildung 1 zeigt als Beispiel den Score des Hospital for Special Surgery (HSS). Organmesswerte, welche über den Schweregrad der Schädigung («impairment») Auskunft geben, sind Schwellung, Gefühllosigkeit, «giving way», Knieuntersuchung, Quadrizepsumfang, Bewegungsausmass, Stabilitätstests. Damit kann man 78 Punkte erreichen bzw. 78% abdecken.

Fähigkeitsmesswerte, welche den Schweregrad der Patientenbehinderung ausdrücken, sind Treppensteigen, Sportrückkehr, Fähigkeitstests und ergeben einen Anteil von 2%.

Abbildung 2 zeigt die geläufigsten Scores bei den Knieinstabilitäten und ihre prozentuale Aufteilung der Punkte auf die ICIDH-Ebenen «Organ und Individuum». Unter den 70% Organmesswerten werden mit 30 bis 40% die objektiven Stabilitätstests gewichtet. Durch operative Verbesserung dieser Messwerte ist ein Gewinn von 30 Scorepunkten möglich, auch wenn der Patient durch die Schwierigkeiten beim Laufsport allenfalls 5 Scorepunkte einbüsst.

Probleme mit der Schmerzmessung

Ein spezielles Problem ist die Beurteilung und Messung von Schmerzen. Gehört Schmerz auf die Organ- oder Individuumsebene? Raspe [7] hat mit guten Gründen dafür argumentiert, dass der Schmerz sinnvollerweise als eine Schädigung zu interpretieren sei, welche unter dem Begriff der sensorischen Schädigung bzw. «sensory impairment» (I 98) klassifiziert werden kann. Schmerz als solcher ist nicht einfach zu messen, da dieser einer subjektiven Erlebnisqualität entspricht. Ebenso unzuverlässig lässt sich die Zuordnung der Schmerzintensität zur Fähigkeitsstörung erstellen. Bei akuten Schmerzen gelingt die Korrelation in der Regel gut, bei den häufigeren, chronifizierten Schmerzen eher schlecht. Schmerzen können sich also sehr unterschiedlich im Fähigkeitsbereich ausdrücken, so dass der Miteinbezug einer allfälligen Handicapsituation in der Gesellschaft (gesellschaft-

Korrespondenz:

Dr. med. Luzi Dubs

Spezialarzt FMH für Orthopädische Chirurgie

Merkurstrasse 12

CH-8400 Winterthur

Abbildung 1

Kniescore des Hospital for Special Surgery (HSS).

Kriterien	Gewichtung innerhalb des Scores	ICIDH-Ebenen		
		I	D	H
Schmerz	2%	2%		
Schwellung	2%	2%		
Gefühllosigkeit	2%	2%		
Giving way	8%	8%		
Treppensteigen	2%		2%	
Sportrückkehr	6%		6%	
Fähigkeitstests	14%		14%	
Knieuntersuchung	14%	14%		
Quadrizepsumfang	4%	4%		
Bewegungsausmass	6%	6%		
Stabilitätstests (LCL, MCL, ACL, PCL)	40%	40%		
		78%	22%	

Abbildung 2

Kniescoresysteme und ICIDH-Ebenen.

Score	Schädigung	Fähigkeitsstörung
Lysholm	75	25
Marshall	78	22
Noyes	50	50
OAK	75	25
IKDC	71	29
Total	70%	30%

liche Rehabilitationshindernisse, Traumatisierungen in der Lebensbiographie usw.) ergänzend notwendig ist, um den Schweregrad der Schmerzwahrnehmung zu erfahren. In den verschiedenen Kniestabilitäts-Scores werden Schmerzen durchschnittlich zu 30% (2–45%) bewertet. Wer ein steifes Knie hat oder den Schongang wegen Schmerzen einnimmt und dadurch schmerzfrei wird, kann also schon allein dadurch 30 Scorepunkte gewinnen, derweil er wegen der reduzierten Fähigkeit höchstens 5 Punkte verliert. Durch die Dominanz der Schmerzbewertung und der Untersuchungsbefunde durch Messung von Organparametern kann ein Zerrbild der Patientenrealität entstehen, welches für die schlüssige Beurteilung des Interventionsnutzens unzureichend und sogar irreführend ist!

Zusammenhänge anstatt Ursachen

Eine kausale Beziehung zwischen Organ- und Fähigkeitsmesswert wird in der naturwissenschaftlichen Medizin gerne als logisch angesehen. Schliesslich entspricht z. B. die Umwandlung eines instabilen Gelenkes in ein stabiles einer Organreparatur mit kausal-kurativer Logik [8], und nur die Rekonstruktion offenbart die Grundlagen für die Heilung. Ähnlich wie in der Inneren Medizin medikamentöse Korrekturen gewisser EKG-Arrhythmien und Cholesterinmesswerte allein der Evidenzprüfung nicht standgehalten und zu häufigeren Todesfällen geführt haben [3], ist auch davon auszugehen, dass «Gelenkrekonstruktionen» nicht a priori mit einem günstigen Patientennutzen vergesellschaftet sein müssen, indem die Organmesswerte verändert werden. Eine isolierte Stabilitätsprüfung als Kriterium für den Patientennutzen muss analog als ein Epiphänomen oder Surrogat angesehen werden. Ihre starke Berücksichtigung im Score birgt die Gefahr eines Surrogattrugschlusses [3, 4].

Anstelle von kausalen Beziehungen scheint es wichtiger, von mehr oder weniger starken (funktionalen) *Zusammenhängen* zu sprechen, wie dies Zycha in seinem kybernetischen Modell dargestellt hat [9].

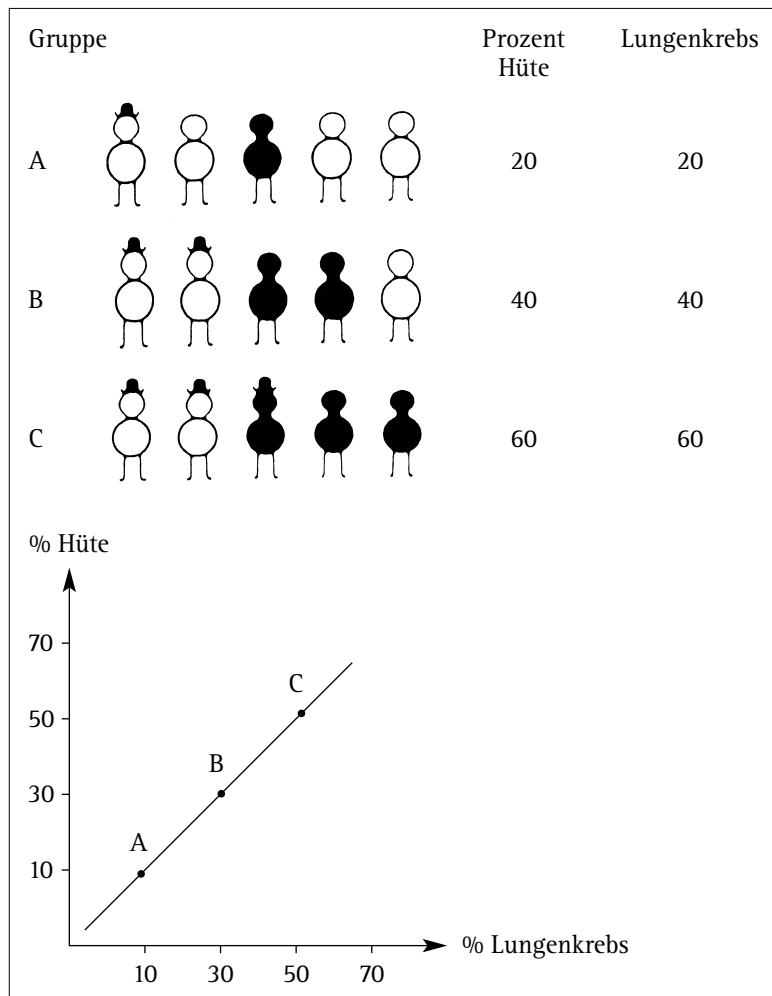
Am Beispiel der Kniebandverletzungen können diverse Faktoren am Endresultat beteiligt sein. So gibt es sicher einen Zusammenhang mit dem Schweregrad des Traumas, einen solchen mit dem Trainingszustand, mit der allfälligen Fremdschuld bei der Traumatisierung oder der geschwächten psychosozialen Resistenzlage des Verunfallten, welcher für sich allein schon von prognostischer Bedeutung sein kann. Diese möglichen Handicaps sind in den geläufigen Kniescores nicht enthalten. Sobald man aber von Zusammenhängen anstelle von Ursachen spricht, muss man sich wiederum klar darüber werden, wie streng die Zusammenhänge sind. Ist der Zusammenhang zwischen einem guten Knieergebnis strenger an die Werte der Stabilitätsmessungen gebunden oder entscheidet mehr der Schweregrad des Unfalls über die Güte des Resultats?

Dass selbst streng erscheinende lineare Zusammenhänge genauer durchleuchtet werden müssen, zeigt das bekannte Beispiel des Umwelttrugschlusses aus dem vortrefflichen Buch «Torheiten und Trugschlüsse in der Medizin» von Skrabanek [4] in Anlehnung an Rosen et al. [10], was in Abbildung 3 dargestellt ist.

In einer Bevölkerungsgruppe A tragen 20% einen Hut, und 20% haben Lungenkrebs. In der Bevölkerungsgruppe B tragen 40% einen Hut, und man findet bei 40% Lungenkrebs. In der Gruppe C beträgt die jeweilige Rate 60%. Es besteht also offensichtlich eine lineare Beziehung zwischen dem Tragen eines Hutes und dem Auftreten von Lungenkrebs, so dass

Abbildung 3

Der Umwelttrugschluss (nach Skrabanek und Rosen et al. [4]).



logischerweise zu folgern wäre, das Huttragen zu verbieten. Die Illustration zeigt jedoch sehr schön, dass diese Beziehung keineswegs verbindlich sein muss. Es ist einleuchtend, dass hier kein strenger Zusammenhang vorliegt. Sollte in gleicher Weise ein linearer Zusammenhang zwischen Rückenschmerzen und radiologischen Bandscheibenveränderungen zustande kommen, ist wohl niemandem übelzunehmen, daraus abzuleiten, dass Rückenschmerzen die Folge von Bandscheibenveränderungen seien und umgekehrt.

Anders könnte es aussehen, wenn die gleiche lineare Beziehung zwischen der Fusslänge von 28 cm und dem Tragen der Schuhgrösse 46 angegeben wird. Es ist kaum anzunehmen, dass solche mit kleinen Füßen die grossen Schuhe tragen und solche mit grossen Füßen sich in kleine Schuhe hineinzwängen. Somit besteht ein relativ strenger logischer Zusam-

menhang zwischen diesen Messwerten. Letztlich geht es immer wieder darum, wissenschaftliche Beobachtungen, ein Testresultat mit dem Kranksein bzw. dem Interventionsnutzen in einen möglichst verbindlichen Zusammenhang zu bringen.

Die im rein pathogenetischen Modell zu denken gewohnte Medizin nach den Newton-Galileischen Prinzipien von Ursache und Wirkung wird sich künftig intensiv mit der Frage von Biases und Trugschlüssen auseinanderzusetzen haben, um möglicherweise schon bald zu jenen Erkenntnissen zu kommen, welche verschiedene biologische Grundlagenforscher des 20. Jahrhunderts vorbereitet haben [11–16]. Lebewesen als «nicht triviale Maschinen» [17] unterliegen biosemiotischen und kybernetischen Prinzipien, den Phänomenen der Selbstorganisation innerhalb ihrer verschiedenen Schichtungen. Diese Selbstorganisation wird durch das salutogenetische Denkmodell reflektiert [18]. Resultatbeurteilungen werden zunächst immer am Goldstandard der «natural history» Mass nehmen müssen. Wenn wir unsere «trivialen Maschinen» zur Organreparatur mit Erfolg einsetzen wollen, um die Überlegenheit unserer Eingriffe gegenüber der «natural history» zu dokumentieren, haben wir stets streng darauf zu achten, das Resultat möglichst trugschlussfrei zu präsentieren.

Fähigkeitsassessment als Ausweg

Die Einsicht in die grossen Vorteile der ICIDH als ein mehrschichtiges Klassifikationsmodell der Krankheitsfolgeerscheinungen mit der separaten Schweregradeinteilung auf den drei Ebenen «Organ, Individuum und Gesellschaft» macht es erst verständlich, dass wir – einstweilen auf die Orthopädie bezogen – in der Resultatbewertung mit Scores der Gefahr ausgesetzt sind, Zerrbilder der Patientenrealität zu erstellen und den Gewinn an Scorepunkten automatisch mit Patientennutzen gleichzusetzen. Zweifellos behalten Messwerte auf Organebene eine Bedeutung für das Verständnis der Schädigung. Um sie in ihrer eigenen Wertigkeit bestimmen und berechnen zu können, müssen sie sich jedoch an einer reinen patientenorientierten Fähigkeitseinschätzung, dem sogenannten Fähigkeitsassessment, orientieren. Solche Messinstrumente werden mehr und mehr eingesetzt und erfordern zuerst noch ausgiebige Validierungen. Es ist zu fordern, dass die Mischscores, welche Werte verschiedener Ebenen der Natur in eine hypothetische «pseudokausale» Beziehung bringen, nicht mehr ohne weiteres als Wirksamkeitsnachweis von Interventionen angesehen werden dürfen. Es bleibt auch nachzuprüfen, ob andere Fachdisziplinen ähnlich konzipierte Scores für die klinische Resultatbeurteilung einsetzen.

Literatur

- 1 Krämer KL, Maichl FP. Scores, Bewertungs-Schemata und Klassifikationen in Orthopädie und Traumatologie. Stuttgart-New York: Thieme; 1993.
- 2 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidenzbasierte Medizin (Übers. Kunz R, Fritsche L). München, Bern, Wien, New York: W. Zuckschwerdt Verlag; 1999.
- 3 Schmidt JG, Bucher HC, Steurer J. «Critical appraisal» und Methoden der klinischen Epidemiologie für die praxisbezogene Bewertung von Studien. Kurskriptum Version 2f. Einsiedeln: Paracelsus-Heute-Stiftung; 1998.
- 4 Skrabanek P. Torheiten und Trugschlüsse in der Medizin. 4. Auflage. Mainz: Verlag Kirchheim; 1995.
- 5 World Health Organisation. International Classification of Impairments, Disabilities and Handicaps (Übers. Matthesius RG). Berlin, Wiesbaden: Ullstein Mosby; 1995.
- 6 Hefti F. Numerische «Scores» zur Evaluation von Behandlungsergebnissen am Bewegungsapparat. Schweiz Ärztezeitung 1996;77:492-6.
- 7 Raspe HH. Mindestanforderungen an das ärztliche Gutachten zur erwerbsbezogenen Leistungsfähigkeit von Kranken mit chronisch-unspezifischen Schmerzen. Versicherungsmedizin 1997;49:118-25.
- 8 Dubs L. Alles getan – Patient noch nicht zufrieden: Einführung in eine Evidence Based Surgery. Swiss Surg 1999; 5:160-6.
- 9 Zycha H. Theorie und Erfahrung in der Medizin: Orthodoxe Wissenschaft und ganzheitliche Kybernetik. Forsch Komplementärmed 1998;5(Suppl. 1):52-9.
- 10 Rosen M, Nystrom L, Wall S. Guidelines for Regional Mortality Analysis. Int J Epidemiol 1985;14:292-9.
- 11 Antonoswsky A. Health, stress and coping: New perspectives on mental and physical wellbeing. San Francisco: Jossey-Brass; 1979.
- 12 Foss L, Rothenberg K. The second medical revolution. From biomedicine to infomedicine. Shambala. Boston, New York: New Science Library; 1987.
- 13 Ginzburg C. Spurensicherungen. Berlin: Wagenbach; 1983.
- 14 von Uexküll T. Was ist und was will «Integrierte psychosomatische Medizin?» In: von Uexküll T (Hrsg.). Integrierte Psychosomatische Medizin. 3. Auflage. Stuttgart, New York: Schattauer; 1994. S. 17-34.
- 15 Weiner H. The dynamics of organism: Implications of recent biological thought for psychosomatic theory and research. Psychosomatic Medicine 1989;51:608-35.
- 16 von Weizsäcker V. Gesammelte Schriften. Frankfurt/M.: Suhrkamp; 1986. Bd. 8, S. 92-93.
- 17 Foerster HV. Das Konstruieren einer Wirklichkeit. In: Segal L. Das 18. Kamel. München: Piper; 1988.
- 18 Schüttel W, Brucks U, Johnen R, Köllner V, Lamprecht F, Schnyder U (Hrsg.). Handbuch der Salutogenese. Wiesbaden: Ullstein Medical; 1998.