

2 Bedeutung klinisch gewonnener Information für ärztliche Handlungsoptionen

2.1. Forschungsergebnisse

2.1.1. Kernaussagen

- Etablierte physiologische Scores wie der Glasgow Coma Scale (GCS) erklären einen Großteil der Varianz ärztlicher Entscheidungen in der frühen Phase der Schwerverletztenversorgung. In bestimmten Fällen können klinische Handlungs- bzw. Leitlinienempfehlungen vereinfacht werden.
- Häufig als Indikatorverletzungen (z.B. Rippenfrakturen) propagierte Befunde besitzen hingegen (bei konsequenter Anwendung bildgebender Referenzverfahren) nur eine geringe Aussagekraft in der Diagnostik stumpfer Bauchverletzungen beim Polytraumatisierten. Sie sollten daher nicht als Schrittmacher für Schockraum-Algorithmen verwendet werden.

2.1.2. Publikationen

- **Stengel D**, Seifert J, Braatz F, Beneker J, Ekkernkamp A, Matthes G. Modellierung kritischer Informationsmengen für unfallchirurgische Entscheidungen. Das Sequenzielle Informations-Angebots-Modul (SIAM). *Unfallchirurg* 2005;108:551-558.
- Mutze S, Rademacher G, Matthes G, Hosten N, **Stengel D**. Blunt cerebrovascular injury in patients with blunt multiple trauma: diagnostic accuracy of duplex Doppler US and early CT angiography. *Radiology* 2005;237:884-892.
- Matthes G, Seifert J, Bogatzki S, Steinhage K, Ekkernkamp A, **Stengel D**. Alter und Überlebenswahrscheinlichkeit nach Polytrauma. „Local tailoring“ des DGU-Prognosemodells. *Unfallchirurg* 2005;108:288-292.
- Matthes G*, **Stengel D***, Seifert J, Rademacher G, Mutze S, Ekkernkamp A. Predictive factors of liver injury in blunt multiple trauma. *Langenbeck's Arch Surg* 2005 Nov 1;:1-5 [Epub ahead of print]. *Erstautoren
- Matthes G*, **Stengel D***, Seifert J, Rademacher G, Mutze S, Ekkernkamp A. Blunt liver injuries in polytrauma: results from a cohort study with the regular use of whole-body helical computed tomography. *World J Surg* 2003;27:1124-1130. *Erstautoren

- Rademacher G, **Stengel D**, Siegmann S, Petersein J, Mutze S. Optimization of contrast agent volume for helical CT in the diagnostic assessment of patients with severe and multiple injuries. *J Comput Assist Tomogr* 2002;26:113-118.

2.2. Identifizierung handlungsrelevanter Information

2.2.1. Informationsflut und Intuition

Der Arzt sieht sich in seiner täglichen Praxis mit einer Fülle häufig redundanter oder überflüssiger diagnostischer Ergebnisse konfrontiert. Eine mögliche Definition intuitiver Entscheidungsprozesse ist die unbewusste Filterung und Gewichtung dieser Informationen zu einer bedeutenden Teilmenge, die ausschlaggebend für oder gegen eine bestimmte Handlung ist. Theoretische Konstrukte sehen Intuition als die Erkennung von Merkmalen, ihre Zuordnung zu Mustern und Handlungsketten an (s. Abbildung 1).¹²

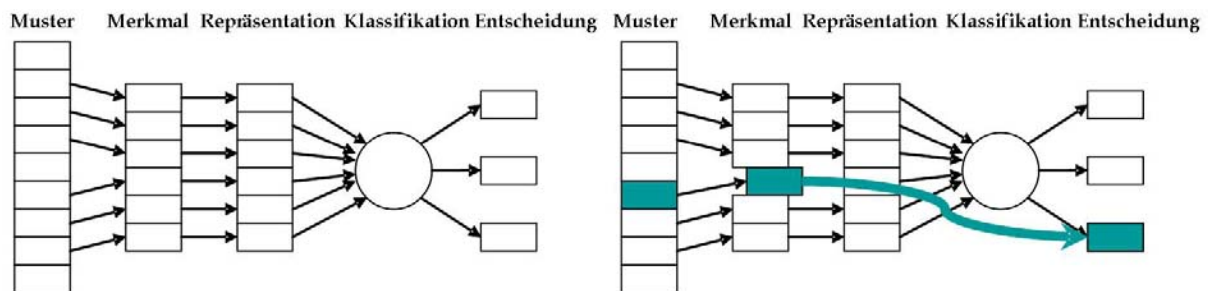


Abbildung 1 Theoretische Modelle der Mustererkennung und Entscheidungsfindung. Links: Annahme einer notwendigen Grundmenge von Information. Rechts: alternative Annahme einer Handlungsauslösung durch ein einziges wesentliches Kriterium. Nach Buckingham und Adams.¹²

Eine praktische Umsetzung dieser Überlegungen wurde bisher jedoch nicht angestrebt. Es wurde postuliert, dass im klinischen Alltag, insbesondere jedoch in Grenzsituationen eine minimale Informationsmenge ausreicht, um eine messbare Aktion nach sich zu ziehen.

Um diese Theorie zu überprüfen, wurde ein Modell entwickelt, das auf der stufenweisen Präsentation unterschiedlich komplexer Information beruht.

¹² Buckingham CD, Adams A. Classifying clinical decision making: interpreting nursing intuition, heuristics and medical diagnosis. *J Adv Nurs* 2000;32:990-998.

2.2.2. Modellierung kritischer Informationsmengen

2.2.2.1. Methoden

Es wurden folgende Phasen eines Protokolls festgelegt, mit dem die Quantifizierung bestimmter Domänen innerhalb ärztlicher Handlungsketten möglich sein könnte:

- 1) In einem relevanten Szenario sollte den Testpersonen in mehreren Schritten zunehmend detaillierte klinische Information (Exposition) angeboten werden, auf deren Grundlage sie die Entscheidung für oder gegen eine eindeutig bezeichnete medizinische Intervention treffen müssen (Ereignis). Dieses Ereignis ist dichotom- es tritt ein oder tritt nicht ein. Die Stärke der Exposition, das Ereignis auszulösen, sollte gemessen und mit anderen Expositionen verglichen werden. In statistischen Begriffen: es sollte der Grad der Varianz der Ereignishäufigkeit beschrieben werden, der durch die dargebotene Information bedingt wird.
- 2) Um robuste Rückschlüsse auf Entscheidungsmuster zu erlauben, sollte der Grad der Übereinstimmung zwischen verschiedenen Testpersonen (d.h., der Eintritt des Ereignisses in Abhängigkeit von bestimmten Informationen bzw. Informationsmengen) jenseits einer zufälligen Übereinstimmung liegen.
- 3) Die Entscheidungen sollten reproduzierbar sein. Bei einer nochmaligen Konfrontation mit dem gleichen Szenario sollte also auch wieder die gleiche Entscheidung getroffen werden.
- 4) Die schrittweise angebotenen Informationsmengen sollten jeweils einer Merkmalsdimension zugeordnet sein, mit anderen Worten: Informationen, die ein physiologisches Zustandsbild beschreiben, sollten sich klar von anatomischen Angaben unterscheiden.

Grundlage dieser Pilotstudie bildeten Fallpräsentationen von 98 konsekutiven Schwerverletzten, deren Daten der lokalen DGU-Registrierung entnommen wurden und in Tabelle 1 zusammengefasst sind.

Die Fälle wurden drei Notärzten (einer Oberärztin für Unfallchirurgie und Rettungsmedizin, einem Facharzt für Chirurgie mit der Zusatzbezeichnung Rettungsmedizin, einem leitenden Notarzt und Oberarzt für Anästhesie) in unterschiedlichem Umfang präsentiert.

Hierzu wurden einzelne Aspekte des physiologischen Zustandsbildes am Unfallort (Glasgow Coma Scale [GCS], Schockindex), des Unfallmechanismus, der notärztlichen Einschätzung der Verletzungsschwere und des definitiven

Verletzungsmusters nach Komplettierung der bildgebenden Diagnostik zur Verfügung gestellt. Endpunkt war die Entscheidung für oder gegen eine endotracheale Intubation am Unfallort.

Das Experiment gliederte sich in vier Schritte, die jeweils durch eine Latenzphase von einer Woche getrennt wurden.

Im ersten Schritt wurden Verletzungsmechanismus (z.B. Verkehrsunfall, Sturz aus großer Höhe), GCS, systolischer Druck und Herzfrequenz angeboten. Die in den weiteren Schritten zur Verfügung gestellten Informationen sind in Abbildung 2 dargestellt.

Den Abschluss bildete eine Überprüfung der Retest-Reliabilität und internen Test-Konsistenz anhand einer zufälligen Stichprobe aus 30% der in allen Schritten angebotenen Patientendaten.

Merkmal	Studienpopulation:	DGU 2002:
	Mittel (95% KI)	Mittel (95% KI)
n	98	2983
Alter (Jahre)	36,6 (32,7 - 40,6)	41,7 (40,9 - 42,4)
männlich (%)	54,1 (43,7 - 64,2)	71,9 (69,5 - 74,2)
ISS	37,3 (35,2 - 39,4)	23,8 (23,3 - 24,3)
Verkehrsunfall	56,1 (45,7 - 66,1)	63,3 (61,6 - 65,1)
Sturz aus >3 m Höhe	17,3 (10,4 - 26,3)	13,8 (12,6 - 15,1)
intubiert, präklinisch (%)	79,6 (70,3 - 87,1)	54,7 (52,7 - 56,7)
GCS, intubierte Patienten	6,9 (5,9 - 7,9)	3,5 (3,4 - 3,6)
GCS, nicht-intubierte Patienten	13,6 (12,1 - 15,1)	13,4 (13,2 - 13,6)
Beatmungsdauer (Tage)	13,3 (10,4 - 16,2)	10,0 (9,8 - 10,2)
ITS Liegezeit (Tage)	16,2 (13,2 - 19,3)	12,2 (11,9 - 12,5)
Letalität (%)	30,6 (21,7 - 40,7)	15,5 (13,9 - 16,6)

Tabelle 1 Eckdaten der für die Präsentation verwendeten Fälle.

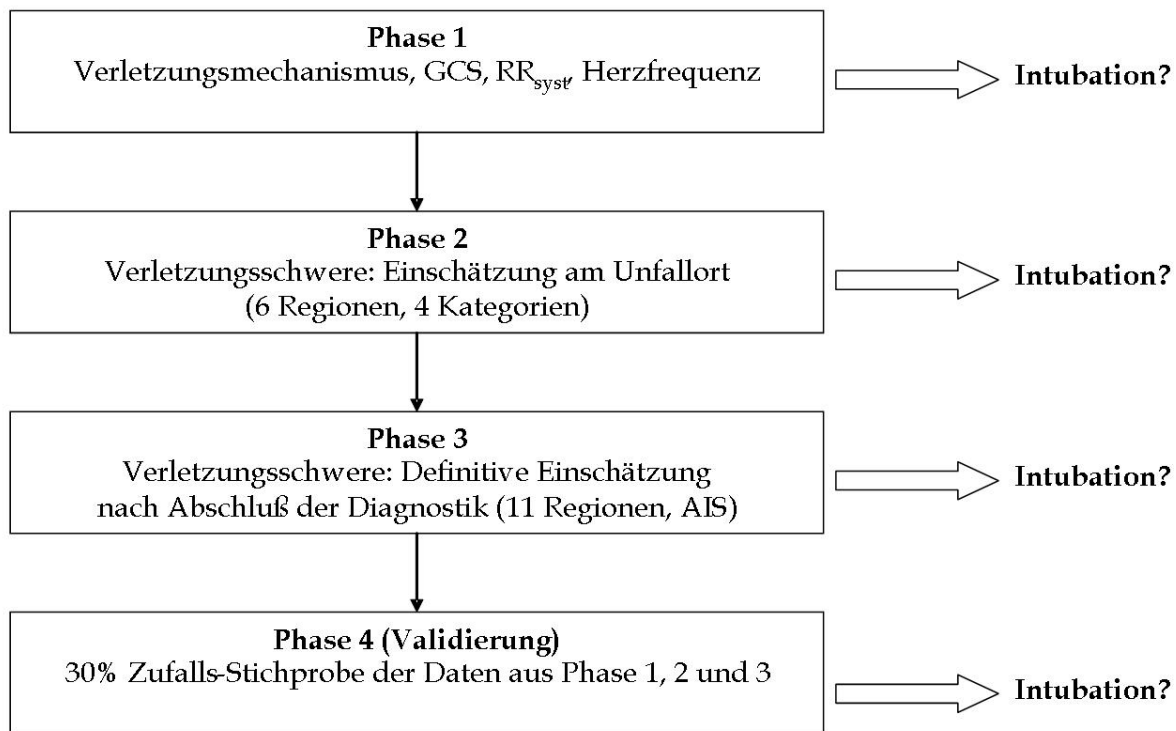


Abbildung 2 Studienprofil und Informationsangebot in den einzelnen Phasen.

Der Einfluss der präsentierten Information (Exposition) auf die Wahrscheinlichkeit einer Intubation (Ereignis) wurde im univariaten und multivariaten logistischen Regressionsmodell überprüft. Neben den Angaben der Studienteilnehmer wurde auch die tatsächlich erfolgte Intubation am Unfallort berücksichtigt.

Für die Generierung multivariater Modelle gelten nach Hosmer und Lemeshow¹³ andere Eingangskriterien und Grenzwerte als bei üblichen statistischen Testverfahren. Expositionen, die in der univariaten Testung einen p-Wert unter 0,25 liefern, gelten üblicherweise als viel versprechende Kandidaten für die multivariate Prüfung. Die Modellierung erfolgte im Backward-Selektionsverfahren sowohl unter statistischen als auch klinischen Kriterien; Variablen wurden ab einem Grenzwert von $p=0,1$ wieder verworfen. Bei nominalen Skalen (Verletzungsmuster) wurden Dummy-Variablen generiert, um den Einfluss einzelner Variablenklassen auf die Ereigniswahrscheinlichkeit zu prüfen (also z.B. Sturz aus großer Höhe im Vergleich zu Verkehrsunfall). Zudem wurden Interaktionen erster Ordnung durch Multiplikation einzelner Variablen geprüft.

¹³ Hosmer DL, Lemeshow S. *Applied logistic regression*. New York: Wiley, 2000.

Die Modellgüte wurde durch das Bayes-Informationskriterium (dieses nimmt mit zunehmender Anpassung ab) und die Flächen (AUC) unter den Receiver Operating Characteristics (ROC) Kurven bestimmt. Ein Modell mit einer AUC=100% erklärt den Eintritt eines Ereignisses perfekt, eine AUC=50% entspricht einem Münzwurf. Die Beobachter-Übereinstimmung jenseits des Zufalls wurde innerhalb jeder experimentellen Phase durch ungewichtete kappa-Statistik bestimmt.

Die Retest-Reliabilität beschreibt die Korrelation von zwei Messwertreihen.¹⁴ Diese Korrelation entspricht dem Anteil der Gesamtunterschiedlichkeit der Testergebnisse, der auf wahre Merkmalsunterschiede zurückzuführen ist- mit anderen Worten, wie häufig gleiche Entscheidungen bei einem nochmaligen Testdurchlauf getroffen werden.

Die Retest-Reliabilität r_{t1t2} errechnet sich als

$$[1] \quad r_{t1t2} = \text{cov}(t1, t2) / s_{t1} \times s_{t2}$$

wobei cov die Kovarianz, t1 das Ergebnis ersten Testergebnisses, t2 das Ergebnis ersten Testergebnisses und st1 und st2 deren jeweilige Standardabweichungen darstellen.

Die interne Testkonsistenz gibt an, wie gut ein Variablensatz eine Klasse von Merkmalen beschreibt. Skalen setzen sich häufig aus Unterskalen zusammen (in der Lebensqualitätsmessung z.B. soziale, physische und psychische Dimension), die wiederum durch einzelne Komponenten abgebildet werden sollen. Passen diese Komponenten nicht zueinander, misst die Unterskala möglicherweise nicht das, was sie messen soll. Diese Passgenauigkeit wird durch Cronbachs alpha beschrieben.¹⁵

Formal erfasst alpha den auf eine Merkmalsdimension zurückgehenden Anteil aller Items (sog. Homogenitätsindex).

Abschließend erfolgten die Ermittlung der Überlebenswahrscheinlichkeit intubierter und nicht-intubierter Patienten nach der Kaplan-Meier-Methode und die Überprüfung von Differenzen mittels Log-Rank-Test.

Die statistische Auswertung erfolgte mit dem Software-Paket STATA 8.0 für Windows-Rechner. Da das Programm keine Varianz (var) der Flächen unter den ROC-Kurven liefert, wurden diese durch Normalapproximation mit Hilfe der Formel

¹⁴ Bortz J, Döring N. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer, 2000.

¹⁵ Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.

[2] $\text{var}(\text{AUC}) = \text{AUC} \times (1 - \text{AUC}) / n$

bestimmt, um 95% Konfidenzintervalle zu berechnen.

2.2.2.2. Ergebnisse

In der ersten Phase der Studie erwies sich bei den Studienärzten der GCS-Wert an der Unfallstelle als einziger Prädiktor für die Intubationswahrscheinlichkeit (s. Abbildung 3). Unfallmechanismus, Herzfrequenz und Blutdruck trugen demgegenüber nicht zur Gesamtvarianz bei. Lediglich bei den Notärzten vor Ort erhöhte der systolische Blutdruck gering die Modellstabilität. Für die Beobachter-Übereinstimmung zwischen allen Ärzten wurde ein kappa-Wert von 0,74 ermittelt; zwischen den Studienärzten war die Übereinstimmung perfekt (kappa=0,94).

In weiteren Studienphasen war eine deutlich geringere Beobachter-Übereinstimmung zu verzeichnen (alle Ärzte: kappa 0,49 und 0,23, nur Studienärzte: kappa 0,84 und 0,40). Die Flächen unter den ROC-Kurven erreichten nach Modellierung der geschätzten Verletzungsschwere und des definitiven Verletzungsmusters nicht diejenigen des GCS (s. Tabelle 2).

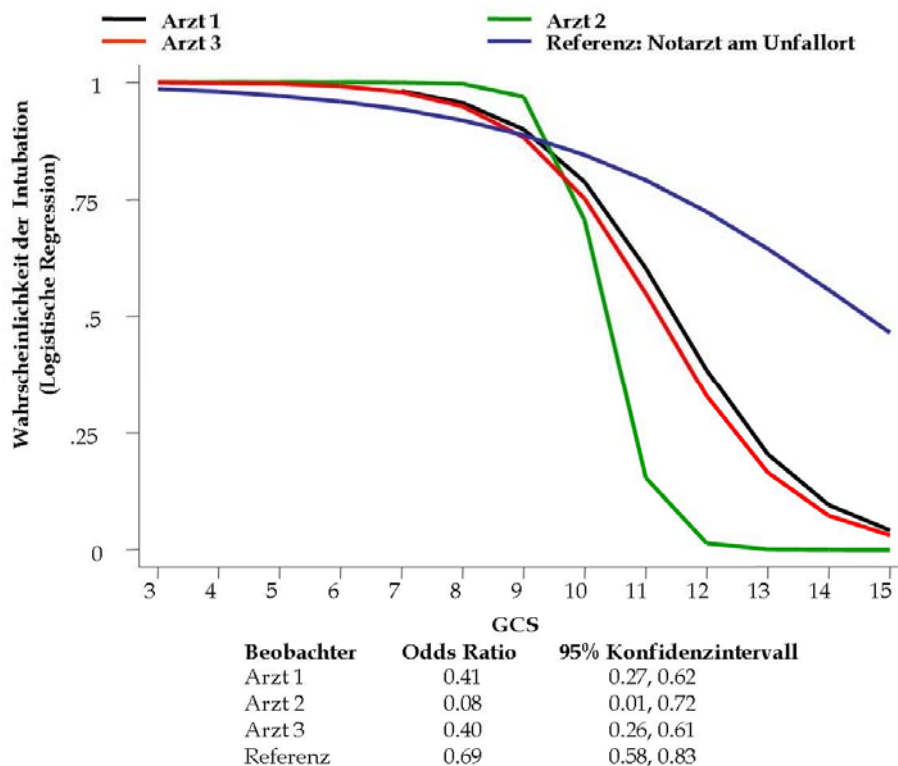


Abbildung 3 Vorhersagewahrscheinlichkeit einer Intubation in Abhängigkeit vom GCS-Wert vor Ort.

	Phase 1		Phase 2	Phase 3
	GCS	GCS + RR _{sys}	Logistisches Modell	Logistisches Modell
Arzt 1	0,983 (0,957 – 1,000)	0,999 (0,992 – 1,000)	0,919 (0,864 – 0,973) Gesicht, Schädel, Thorax, Becken	0,981 (0,953 – 1,000) Schädel, Pneumothorax, Becken
Arzt 2	0,998 (0,990 – 1,000)	0,999 (0,993 – 1,000)	0,955 (0,913 – 0,996) Gesicht, Thorax, untere Extremität	0,833 (0,759 – 0,907) Gesicht, Schädel, Pneumothorax, Unterschenkel
Arzt 3	0,984 (0,960 – 1,000)	0,985 (0,961 – 1,000)	0,966 (0,930 – 1,000) Gesicht, Thorax, untere Extremität	0,978 (0,949 – 1,000) Schädel, Becken, Femur
Referenz	0,854 (0,784 – 0,924)	0,894 (0,833 – 0,955)	0,775 (0,692 – 0,857) Gesicht	0,756 (0,672 – 0,841) Alle Verletzungen

Tabelle 2 Aus der logistischen Regression abgeleitete Flächen unter den ROC-Kurven.

Die Retest-Reliabilität war erwartungsgemäß am höchsten für die in Phase 1 untersuchten Kriterien und betrug für die drei Teilnehmer 82%, 75% und 91%.

In der zweiten Phase waren die Entscheidungen nur von zwei Teilnehmern zu reproduzieren (r_{TT} 74% und 82%). In der dritten Phase waren Retest-Ergebnisse von 69%, 51% und 62% zu beobachten.

2.2.2.3. Wissenschaftlicher Erkenntnisbeitrag und Einschränkungen der Interpretation

Mithilfe des Modells konnte die Robustheit des GCS für die Vorhersage der Intubationswahrscheinlichkeit belegt werden. Alle übrigen Informationen trugen nicht oder nur in geringem Maß zur Modellgüte bei. Die Kenntnis *eines* Merkmals (ein Surrogat für die Bewusstseinslage) war für eine klinische Handlung ebenso bedeutsam wie eine große Menge anderer Information.

Der Arzt benötigt offensichtlich nur wenige Informationen, um sich für oder gegen eine Handlung zu entscheiden. Dies ist nützlich und gefährlich zugleich. Nützlich, weil es die durch Erfahrung erworbene ärztliche Fähigkeit verdeutlicht, wesentliche von unwesentlichen Wahrnehmungen zu trennen und in einen Handlungsablauf umzusetzen- der Arzt handelt ökonomisch. Gefährlich, weil zahlreiche als hinweisend angesehene Informationen keiner formalen Überprüfung unterzogen

wurden und damit eine auf den ersten Blick konsequente Handlungsweise möglicherweise zu einem unerwünschten Ergebnis führt.

Die vorgeschlagene Methode soll einen kleinen Baustein zum Verständnis und zur Einbindung ärztlicher Praxis in formale Entscheidungsbäume beitragen. Informationsflut kann in kleinere Bestandteile aufgetrennt und der Beitrag dieser Teile zur resultierenden Handlung gemessen werden.

Es sollte unterstrichen werden, dass die erste klinische Einschätzung einen wichtigen (wenn nicht den wichtigsten) diagnostischen Test darstellt, dessen Effektivität und Effizienz ebenso bewertet werden müssen wie diejenigen apparativer Untersuchungsverfahren. Dies gilt selbstverständlich nicht nur für das Szenario der schweren Verletzung

Der GCS setzt sich jedoch wiederum aus drei Subskalen mit 120 möglichen Kombinationen zusammen. Zudem ist ein GCS von 9 allgemein als kritischer Grenzwert für eine Bewusstseinslage akzeptiert, der eine Intubation erforderlich macht.

Ob daher die unter Studienbedingungen gefällte Entscheidung zur Intubation einer sog. sozial erwünschten Antwort („intubiere ab einem GCS < 9“) oder einem verinnerlichten Handlungsmuster nach Erkennung eines Merkmals („intubiere den bewusstseinsgetrübten Schwerstverletzten“) entspringt, kann nicht abschließend beurteilt werden.

Es wurde kürzlich darauf hingewiesen, dass die beste motorische Reaktion nahezu vollständig den prädiktiven Wert des GCS erklärt und in weiteren wissenschaftlichen Untersuchungen den Globalscore ersetzen sollte.¹⁶ Diese Ergebnisse lagen zum Zeitpunkt der Durchführung des Experiments noch nicht vor, müssen jedoch bei einer Validierungsstudie berücksichtigt werden.

Kritisiert werden kann die noch unzureichende Wirklichkeitsnähe des hier vorgestellten Ansatzes.

Aus methodischen Gründen musste das dynamische Szenario der Notfall-Intubation auf einen statischen Fallbericht reduziert und artifiziell aufgeteilt werden. Statistische Modelle versuchen, möglichst viel der Varianz realer Beobachtungen durch möglichst wenige Einflussfaktoren zu erklären. Die Auswahl der angebotenen Variablen beruhte auf einem informellen Konsens zwischen Klinikern und Methodikern. Es erscheint angebracht, den notwendigen Datensatz für eine größere Untersuchung durch eine formale Konsensusprozedur zu generieren. Auch muss

¹⁶ Healey C, Osler TM, Rogers FB, Healey MA, Glance LG, Kilgo PD, Shackford SR, Meredith JW. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. *J Trauma* 2003;54:671-678.

zukünftig darauf geachtet werden, Testskalen mit maximalen alpha-Koeffizienten zu verwenden.

Die Ergebnisse belegen allenfalls eine Einschätzung des Notarztes, nicht die tatsächliche *Notwendigkeit* einer Intubation am Unfallort, für die bisher trotz Plausibilität keine eindeutigen wissenschaftlichen Daten existieren.¹⁷

Interessant sind in diesem Zusammenhang die durch Kaplan-Meier-Methode generierten Überlebenswahrscheinlichkeiten intubierter und nicht-intubierter Patienten (s. Abbildung 4).

Die Kurven lassen mehrere Interpretationen zu. Es ist möglich, dass die Studienärzte besser zwischen denjenigen Patienten diskriminierten, die von einer Intubation profitierten und denjenigen, bei denen die möglichen Komplikationen einen möglichen Vorteil überwiegen. Es ist jedoch auch möglich, dass in der Studiensituation lediglich Patienten mit unterschiedlicher Prognose (unabhängig von einem Effekt der Intubation) voneinander unterschieden wurden.

Es ist problematisch, im Intubationsszenario die Richtung der Kausalität zu erkennen. Die Intubation könnte mehr schaden als nutzen oder lediglich bei Patienten mit schlechterer Prognose zur Anwendung kommen.

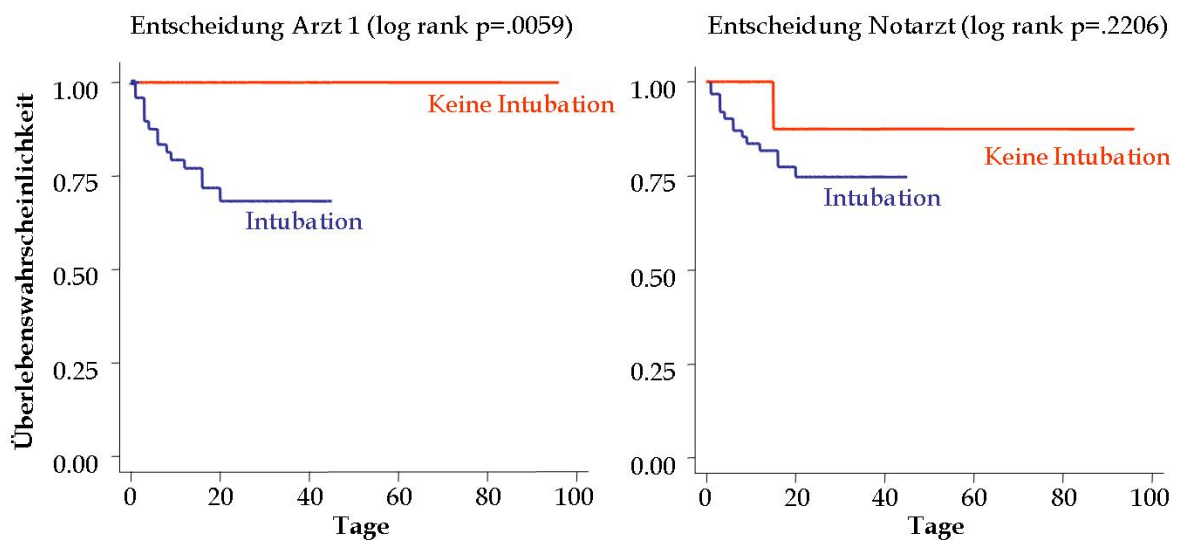


Abbildung 4 Überlebenswahrscheinlichkeit intubierter und nicht-intubierter Patienten in der artifiziiellen und der realen Vor-Ort-Situation.

¹⁷ Liberman M, Mulder D, Sampalis J. Advanced or basic life support for trauma: meta-analysis and critical review of the literature. *J Trauma* 2000;49:584-599.

Es könnte wertvoll sein, bei nachgewiesenen genauen Tests, die mittel- und langfristig keinen Nutzwert i.S. einer Verbesserung des Outcomes zeigen, den Einfluss auf ärztliche Handlungsmuster zu überprüfen. Auch der effektivste Test birgt keinen Nutzen, wenn er für die Entscheidung des Arztes im Vergleich zu anderen diagnostischen Strategien nur von untergeordneter Bedeutung ist.

Physiologische Scores wie der GCS werden regelhaft weiterentwickelt und validiert. Zudem liefern sie sowohl beim wachen als auch beim bewusstseinsgetrübten Patienten handlungsrelevante Informationen. Demgegenüber sind anatomische Scores wie z.B. die Ottawa Ankle¹⁸ bzw. Knee Rules¹⁹, die Canadian C-Spine Rule²⁰ und auf Beckenfrakturen hinweisende klinische Befunde²¹ nur beim wachen, kooperationsfähigen Patienten von diagnostischem Wert.

Die in Deutschland und den europäischen Nachbarstaaten nahezu ausnahmslos durch stumpfe Gewalt bedingten Verletzungen des Abdomens können durch eine körperliche Untersuchung allein nur schwer diagnostiziert und quantifiziert werden – dies gilt insbesondere für Patienten mit eingeschränkter Bewusstseinslage.^{22,23} Dennoch werden traditionell immer wieder sog. Indikatorverletzungen (Blutergüsse, Frakturen in benachbarten anatomischen Regionen u.ä.) als Schlüsselzeichen gelehrt und weithin als Voraussetzung für weiterführende bildgebende Diagnostik angesehen. Es wurden bisher nur wenige Untersuchungen angestrebt, um diese potenziell risikoreichen Aussagen anhand diagnostischer Referenzstandards kritisch zu hinterfragen und zu widerlegen.

¹⁸ Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ* 2003;326:417.

¹⁹ Bachmann LM, Haberzeth S, Steurer J, ter Riet G. The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. *Ann Intern Med* 2004;140:121-124.

²⁰ Stiell IG, Clement CM, McKnight RD, Brison R, Schull MJ, Rowe BH, Worthington JR, Eisenhauer MA, Cass D, Greenberg G, MacPhail I, Dreyer J, Lee JS, Bandiera G, Reardon M, Holroyd B, Lesiuk H, Wells GA. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. *N Engl J Med* 2003;349:2510-2518.

²¹ Sauerland S, Bouillon B, Rixen D, Raum MR, Koy T, Neugebauer EA. The reliability of clinical examination in detecting pelvic fractures in blunt trauma patients: a meta-analysis. *Arch Orthop Trauma Surg* 2004;124:123-128.

²² Poletti PA, Mirvis SE, Shanmuganathan K, Takada T, Killeen KL, Perlmutter D, Hahn J, Mermillod B. Blunt abdominal trauma patients: can organ injury be excluded without performing computed tomography? *J Trauma* 2004;57:1072-1081.

²³ Kirkpatrick AW, Brenneman FD, McLean RF, Rapanos T, Boulanger BR. Is clinical examination an accurate indicator of raised intra-abdominal pressure in critically injured patients? *Can J Surg* 2000;43:207-211.

2.3. Einfluss klinischer Indikatoren auf die Vortest-Wahrscheinlichkeit

2.3.1. Das Problem des stumpfen Bauchtraumas

Es existieren zwei klinische Entscheidungsregeln, die bei negativem Ergebnis die Vortest-Wahrscheinlichkeit für eine Abdominalverletzung erniedrigen sollen (klinische Entscheidungsregeln sollten eine Sensitivität nahe 100% aufweisen, die Spezifität ist zumeist von untergeordnetem Interesse).^{24,25} Das Fehlen von Thorax- und Beckenverletzungen soll hiernach die klinisch-diagnostische Sicherheit in einen unauffälligen Abdominalbefund erhöhen.

Jede Entscheidungsregel ist nur so gut wie die zur Messung der Exposition (also z.B. einer Beckenverletzung) und des Merkmals (also einer intraabdominellen Organschädigung) verwendeten Methoden. Etablierte Algorithmen haben den Nachteil, dass nicht alle Patienten unabhängig von anderen Befunden auch alle klinischen und apparativen diagnostischen Tests durchlaufen- es besteht die Gefahr eines Bestätigungs-Bias, wenn nur positive Befunde durch einen zweiten Test gesichert werden. Zu fordern ist daher eine konsekutive Patientenpopulation mit einheitlicher Dokumentation der genannten Kriterien, um tatsächlich die Aussage treffen zu können, dass z.B. bei unauffälligem klinischem Befund auch keine weitere Bildgebung erforderlich ist. Dies stellt unter den Prämissen der Nichtinvasivität, Strahlenhygiene und Kosten-Nutzen-Rechnung die Idealsituation einer Schockraumversorgung dar- realistisch ist sie hingegen nicht. Übersehene Verletzungen (insb. Gefäß- und Parenchymrisse nach Kreislaufstabilisierung) können beim Polytraumatisierten innerhalb kürzester Zeit zur Dekompensation führen, die bei Unkenntnis des kompletten Verletzungsmusters kaum noch therapeutisch zu beherrschen ist. Es ist schwierig, in dieser Situation triviale von nichttrivialen Befunden zu trennen.

2.3.2. Leberverletzungen

Lebertraumen stellen unter allen Höhlenverletzungen eine besondere Entität dar. Innerhalb der letzten Jahre hat sich bei kreislaufstabilen Patienten ein klarer Wandel

²⁴ Mackersie RC, Tiwary AD, Shackford SR, Hoyt DB. Intra-abdominal injury following blunt trauma. Identifying the high-risk patient using objective risk factors. *Arch Surg* 1989;124:809-813.

²⁵ Grieshop NA, Jacobson LE, Gomez GA, Thompson CT, Solotkin KC. Selective use of computed tomography and diagnostic peritoneal lavage in blunt abdominal trauma. *J Trauma* 1995;38:727-731.

in den Behandlungsoptionen zugunsten einer nichtoperativen Therapie unter intensivmedizinischer Beobachtung vollzogen.

Dies setzt jedoch eine unmissverständliche Klassifikation der Verletzungsschwere (nach den Empfehlungen der American Association for the Surgery of Trauma, AAST, s. Abbildung 5)²⁶ voraus, um im Falle einer erforderlichen Notfall-Laparotomie die operative Taktik zu optimieren und ggf. bereits Gefäß- und Viszeralchirurgen sowie interventionelle Radiologen hinzuzuziehen.



Abbildung 5 Kontrastmittelverstärkte Spiral-CT des Abdomens bei Leberverletzungen unterschiedlicher Schweregrade. Von links nach rechts: Moore II, III und IV Verletzungen.

In eine retrospektive Kohortenstudie wurde eine Fall-Kontroll-Studie eingebettet, um die These der ausreichenden Aussagekraft fehlender Begleitverletzungen wie Thoraxtraumen oder Beckenfrakturen zu überprüfen.

2.3.3. Bedingtes Risikos für eine Leberverletzung

2.3.3.1. Methoden

Die Studie schloss alle zwischen September 1997 und Januar 2001 unter der klinischen Diagnose einer schweren Verletzung bzw. eines Polytraumas (Injury Severity Score > 16) oder wegen eines Hochrasanztraumas (Verkehrsunfall, Sturz aus großer Höhe) mittels Ganzkörper-Spiral-CT untersuchten Patienten ein. Diese Patienten wurden im Rahmen der deutschlandweiten Erfassung Schwerstverletzter durch das Traumaregister der Deutschen Gesellschaft für Unfallchirurgie (DGU) e.V.

²⁶ Moore EE, Cogbill TH, Jurkovich GJ, et al. Organ injury scaling: spleen and liver (1994 revision). *J Trauma* 1995;38:323-324.

mit Hilfe eines standardisierten Erhebungsinstrumentes dokumentiert.²⁷ Zudem wurden klinikintern die Ergebnisse aller Ganzkörper-CT-Untersuchungen in einer separaten Datenbank festgehalten.

Nach erster klinischer Sichtung und abdomineller Ultraschalluntersuchung wurde bei allen Patienten eine native CT-Untersuchung des Schädels (ggf. einschließlich Mittelgesicht) und der HWS, im Anschluss eine Iohexol-verstärkte Spiral-CT von Thorax, Abdomen und Becken durchgeführt. Nach Abschluss der CT-Untersuchung wurde bei Extremitätenverletzungen die Diagnostik durch konventionelle Röntgen-Aufnahmen komplettiert und die Patienten entweder zur chirurgischen Therapie in den Operationssaal oder auf die Intensivstation transportiert.

Neben den axialen Geometrien wurden durch die diensthabenden Radiologen routinemäßig multiplanare Rekonstruktionen angefertigt.

Um die statistische Effizienz zu erhöhen (und Nullbesetzungen in Vierfeldertafeln zu vermeiden) wurden Fälle (Polytraumatisierte mit Leberverletzung) mit Kontrollen (Polytraumatisierte ohne Leberverletzung) individuell für die Kriterien Geschlecht, Alter und ISS gepaart.

Zur Erhöhung der statistischen Power war zunächst eine 1:2 Paarung vorgesehen; im Datensatz konnten jedoch nicht für alle Fälle passende Kontrollen identifiziert werden. Um ein Ausscheiden von Fällen zu vermeiden (die in Fall-Kontroll-Studien führend die statistische Mächtigkeit beeinflussen), wurde eine 1:1 Paarung akzeptiert, die unter den im weiteren dargestellten Stichprobengrößen den Nachweis eines Odds Ratio von 2,0 bei einer angenommenen Prävalenz von interessierenden Indikatoren bei Kontrollen um 30%, einer Power von 80% und einem zweiseitigen Testniveau von 5% erlaubte.

Die statistische Auswertung erfolgte mittels univariater, bei entsprechenden Verteilungsvoraussetzungen auch multivariater konditionaler logistischer Regression. Im Falle eines hohen Grades an unerklärter Varianz kam ein Random-effects-Modell zur Anwendung. Die Modellierung erfolgte im Backward-Selektionsverfahren sowohl unter statistischen als auch klinischen Kriterien. Variablen wurden bei einem univariaten p-Wert von 0,25 eingeschlossen und einem Grenzwert von 0,1 verworfen. Die Modellgüte wurde durch Wald-Statistik und ROC-Analyse bestimmt.

²⁷ Pape HC, Oestern HJ, Leenen L, Yates DW, Stalp M, Grimmer K, Tscherner H, Krettek CH. Documentation of blunt trauma in Europe. Survey of the current status of documentation and appraisal of the value of standardization. *Eur J Trauma* 2000; 26:233-247.

2.3.3.2. Ergebnisse

Im genannten Zeitraum wurden insgesamt 734 Patienten mittels Spiral-CT untersucht, von denen 218 (29,7%) die Eingangskriterien erfüllten. Die demografischen Eckdaten sind in Tabelle 3 dargestellt. Mit 55 Leberverletzten lag die Prävalenz bei 25,2% (95% KI 19,6 - 31,5%).

	Fälle n=55	Gesamte übrige Kohorte n=163	Nur Kontrollen n=55
Alter in Jahren	32,1 (28,0 - 36,1)	36,2 (33,3 - 39,1)	32,5 (28,5 - 36,6)
Anteil Männer	58,2% (44,1 - 71,3%)	70,9% (63,3 - 77,8%)	60,0% (45,9 - 72,9%)
ISS	36,5 (34,3 - 38,6)	34,5 (32,9 - 36,0)	36,1 (33,9 - 38,2)

Tabelle 3 Patientenprofil. Werte in Klammern entsprechen 95% Konfidenzintervallen.

Es wurden 99 Parenchymkontusionen bzw. subkapsuläre Pathologien, 15 Kapsleinrisse und zwei Leberzerreißen nachgewiesen. Patienten wiesen im Mittel 1,8 (95% KI 1,5 - 2,1) Läsionen mit einem mittleren Volumen von 68,5 ml (95% KI 26,4 - 110,6 ml) auf. Elf Patienten boten isolierte Verletzungen des linken Leberlappens, bei jeweils 22 fand sich eine rechts- oder bilobäre Beteiligung. Bei sechs Patienten wurde das Lebertrauma als führende Verletzung klassifiziert. Eine Übersicht über den Segmentbezug gibt die Abbildung 6.

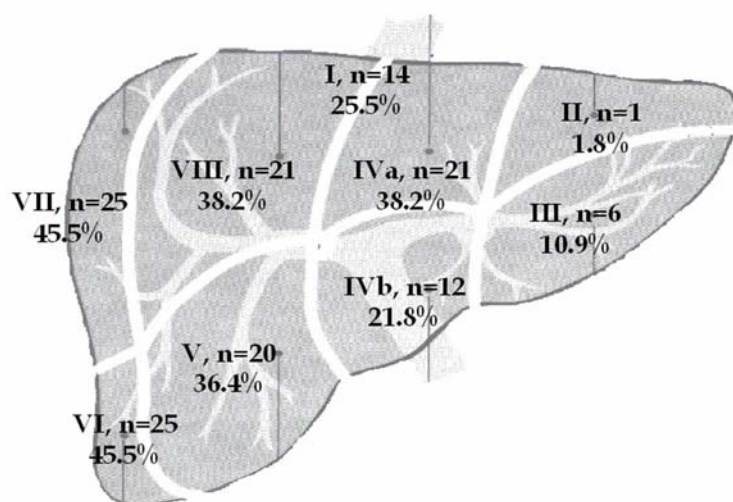


Abbildung 6 Beobachtete Verteilung der Verletzungen in den anatomischen Lebersegmenten.

Die Letalität Leberverletzter betrug 16,4% (95% KI 7,8 - 28,8%). In der Cox-Regression hatte die Leberverletzung im Vergleich zur übrigen Kohorte aus 163 Patienten keinen Einfluss auf die Überlebenswahrscheinlichkeit (Hazard Ratio 0,82, 95% KI 0,33 - 2,04, $p=0,674$).

Erwartungsgemäß erwies sich der ISS als robuste und einzige unabhängige Modellvariable (OR 1,22, 95% KI 1,08 - 1,38). Aufgrund der Stichprobengröße wurden keine Modelle mit mehr als zwei Variablen generiert.

Eine Laparotomie wurde bei 15 Patienten erforderlich (27,3%, 95% KI 16,1 - 40,9%). Patienten mit Kapselverletzungen hatten ein zweifach höheres relatives Risiko für einen operationspflichtigen Verlauf.

Als beste prädiktive Variable für die Notwendigkeit einer operativen Therapie erwies sich die Moore-Klassifikation (OR 3,91, 95% KI 1,59 - 9,61). Es bestand eine enge Korrelation zwischen Moore-Stadium und der Menge freier intraabdomineller Flüssigkeit (Spearman's rho 0,46, $p=0,0004$).

Zu den 55 Fällen mit Leberverletzungen wurden 55 Kontrollen gruppiert. Das Risikoprofil beider Gruppen ist in Tabelle 4 zusammengefasst.

Verletzungsmechanismus	Fälle		Kontrollen	
	n	%	n	%
Autounfall, Fahrer	23	41,8 (28,7 - 55,9)	21	38,2 (25,4 - 52,2)
Autounfall, Beifahrer	3	5,5 (1,1 - 15,1)	8	14,5 (6,5 - 26,7)
Autounfall, Rückinsassen	6	10,9 (4,1 - 22,2)	0	0,0 (0,0 - 6,5)
Frontalaufprall	31	56,4 (42,3 - 69,7)	23	41,8 (28,7 - 55,9)
Einklemmung	12	21,8 (11,8 - 35,0)	16	29,1 (17,6 - 42,9)
Herausschleuderung aus dem Fahrzeug	2	3,6 (0,4 - 12,5)	3	5,5 (1,1 - 15,1)
LKW-Fahrer	3	5,5 (1,1 - 15,1)	1	1,8 (0,1 - 9,7)
Fußgänger	2	3,6 (0,4 - 12,5)	6	10,9 (4,1 - 22,2)
Sturz aus großer Höhe	12	21,8 (11,8 - 35,0)	8	14,5 (6,5 - 26,7)
Überfahrene	3	5,5 (1,1 - 15,1)	4	7,3 (2,0 - 17,6)
Zweiradunfall	3	5,5 (1,1 - 15,1)	7	12,7 (5,3 - 24,5)

Tabelle 4 Risikoprofil, eingebettete Fall-Kontroll-Studie. Werte in Klammern entsprechen 95% Konfidenzintervallen.

Es ließen sich keine überzeugenden Modelle mit Verletzungsmechanismen oder Begleitverletzungen generieren, die *hinweisend* auf eine Leberbeteiligung (OR >1) gewesen wären.

Auch das *Fehlen* viel zitierter Indikatoren wie rechtsseitige Thorax- oder Beckenverletzungen bot keine verwertbare Information (d.h., OR <1). Im vorliegenden Datensatz war das relative Risiko für eine Leberverletzung nach Zweiradunfällen (OR 0,75, 95% KI 0,57 - 0,98, p=0,038) und *linksseitigen* Rippenserienfrakturen (OR 0,78, 95% KI 0,64 - 0,95, p=0,015) erniedrigt; dieser statistische Zusammenhang wurde jedoch als klinisch irreführend bzw. unbedeutend gewertet.

Abbildung 7 fasst die Richtung der Effekte für vermutete Expositionen zusammen. Insbesondere die Punktschätzer der aufgrund hoher unerklärter Varianz durchgeführten Random-effects Regression weisen mit ihrer Nähe zum Nulleffekt auf die zweifelhafte Bedeutung sog. Indikatorverletzungen hin.

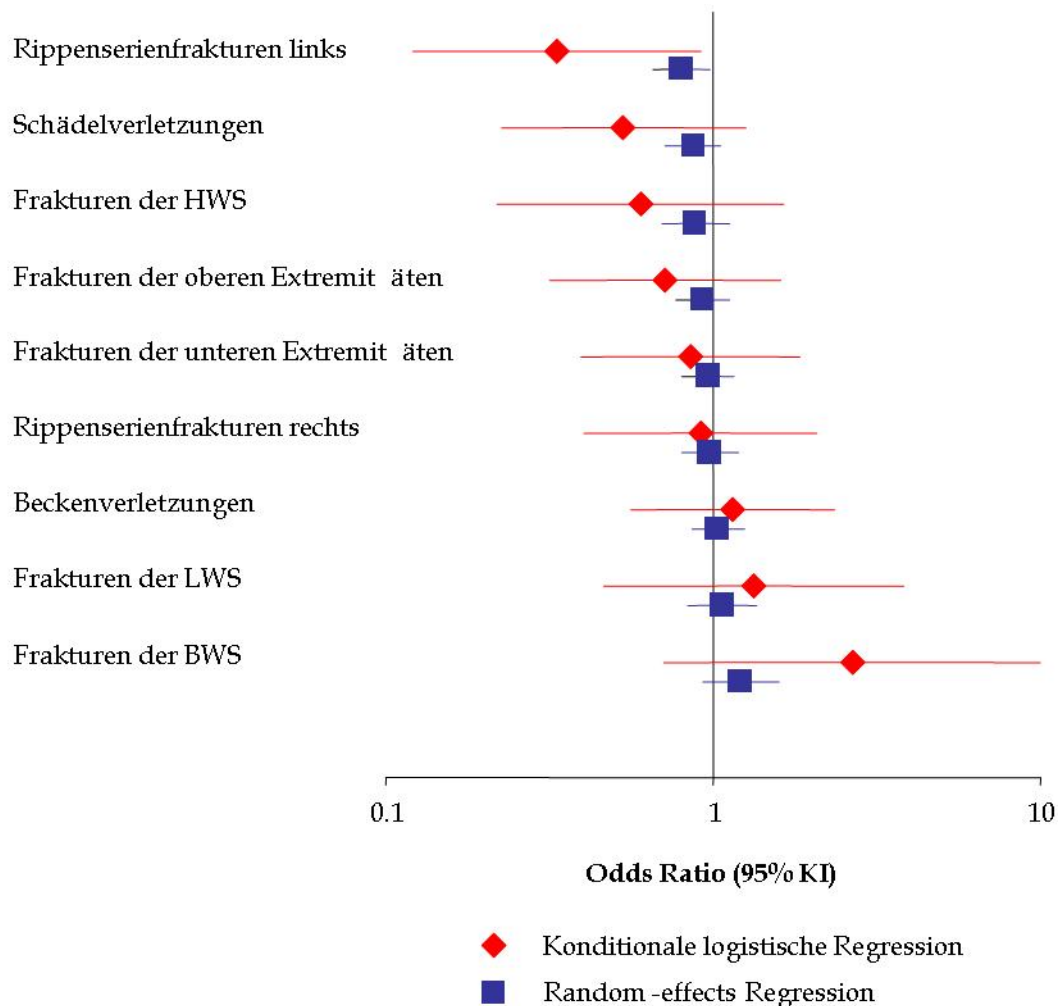


Abbildung 7 Einfluss propagierter Indikatorverletzungen auf das Risiko für eine Leberbeteiligung bei Polytrauma.

2.3.3.3. Wissenschaftlicher Erkenntnisbeitrag und Einschränkungen der Interpretation

Traditionell erlernte, gelehrte und akzeptierte klinische Indikatoren für Gesundheitsstörungen müssen sich einer formalen Überprüfung anhand unabhängig angewandter diagnostischer Referenzstandards stellen. Ohne in jedem Fall apparative Diagnostik zu fordern, muss sich der Arzt der Grenzen bestimmter Zeichen und Symptome bewusst sein.

Da diagnostische Algorithmen immer mit einer klinischen Einschätzung beginnen, ist es wichtig, die Grenzwerte dieser Entscheidungsknoten neu festzulegen. Zu unterstreichen ist die Sensitivität: lässt das Fehlen eines klinischen Indikators mit hinreichender Sicherheit den Schluss zu, dass die vermutete Gesundheitsstörung unwahrscheinlich ist? Erst wenn diese Frage bejaht wird, wird ein stufenweises diagnostisches Vorgehen im Ausschlussverfahren ermöglicht.

Die beobachtete Prävalenz der Leberverletzungen bei Polytrauma von 25% ist hoch und liegt oberhalb der in der Literatur berichteten Häufigkeit zwischen 4 und 10%.
28,29,30

Tatsächlich umfasste die hier untersuchte Stichprobe auch leichte Verletzungen der Moore-Stadien I und II; nur jeder zehnte Patient bot eine höhergradige Leberpathologie. Ob die Prädiktion ausschließlich chirurgisch bedeutsamer Verletzungen besser gelingt, kann mit den präsentierten Daten nicht beurteilt werden.

Zweifellos werden bei konsequenter Durchführung der CT auch Verletzungen aufgedeckt, deren klinische Relevanz fragwürdig erscheint. Diese von Protagonisten klinischer Entscheidungsregeln häufig vorgebrachte Argumentation wird auch im nächsten Abschnitt dieser Arbeit bedeutsam sein- sie hat jedoch sowohl aus wissenschaftlicher als auch aus praktischer Sicht Mängel.

Für die interne Validität einer Diagnosestudie ist die klinische Relevanz einer mit dem gültigen Goldstandard nachgewiesenen Pathologie unerheblich- sie ist oder sie ist nicht vorhanden. Relevanz ist ein klinischer Begriff, der einer erheblichen

²⁸ Clancy TV, Maxwell JG, Covington DL, Brinker CC, Blackman D. A statewide analysis of level I and II trauma centers for patients with major injuries. *J Trauma* 2001;51:346-351.

²⁹ Rutledge R, Hunt JP, Lentz CW, Fakhry SM, Meyer AA, Baker CC, Sheldon GF. A statewide, population-based time-series analysis of the increasing frequency of nonoperative management of abdominal solid organ injury. *Ann Surg* 1995;222:311-326.

³⁰ Bardenheuer M, Obertacke U, Waydhas C, Nast-Kolb D. Epidemiologie des Schwerstverletzten: eine prospektive Erfassung der präklinischen und klinischen Versorgung. *Unfallchirurg* 2000;103:355-363.

Verzerrung durch Erfahrungswerte und lokale Vorgaben unterliegt. Die diagnostische Genauigkeit einer experimentellen Methode kann und darf nur an reproduzierbaren Messungen festgemacht werden.

Das Polytrauma und seine Folgen werden zunehmend als Systemerkrankung mit häufig überraschender und unvorhersehbarer Dynamik verstanden. Ein bei der Erstuntersuchung kleines subkapsuläres Leberhämatom kann nach Kreislaufstabilisierung rasch zu einer hämodynamisch kritischen intraabdominellen Blutung führen- eine irrelevante Höhlenverletzung kann es beim Polytrauma nicht geben.

Trotz der aufgrund der Stichprobengröße eingeschränkten allgemeinen Übertragbarkeit und der Design-bedingt fehlenden Indikatoren der diagnostischen Genauigkeit lassen die Effektschätzer vermuten, dass die Ermittlung der Vortest-Wahrscheinlichkeit für eine Leberverletzung allein aufgrund körperlicher Untersuchungsbefunde einem Münzwurf gleichkommt.

Umso größere Bedeutung kommt daher der adäquaten bildgebenden Diagnostik zu. Die Abkehr von invasiven Prozeduren (d.h., Peritoneallavage oder explorative Laparotomie) bei vermutetem Abdominaltrauma zugunsten der Ultraschalluntersuchung wurde zu recht als Meilenstein in der Schockraumbehandlung angesehen. Dies verselbständigte allerdings sich so sehr, dass der Ultraschall bald als das „EKG des Chirurgen“ angesehen wurde. Die breite Einführung der CT für die Traumadiagnostik erhöhte sogar die Bedeutung der Sonografie- sie wurde jetzt zur Filterstation für verletzte Patienten; ein negatives Sonogramm galt und gilt als Befund, der einen Verzicht auf weiterführende Diagnostik erlaubt.

Ob der bewusste Nichteinsatz eines Referenzverfahrens mit einer geringeren diagnostischen Genauigkeit erkaufte wird, wurde seit der Etablierung der Ultraschalluntersuchung nicht stringent überprüft.