

Aus der Kinderchirurgie
Dr. med. habil. Bertram Reingruber
Lehrstuhl für Chirurgie
Prof. Dr. med. Hans J. Schlitt
der medizinischen Fakultät
der Universität Regensburg

Klassifikation kindlicher Frakturen

Zuverlässigkeit und Anwenderfreundlichkeit der Systematik nach AO und LiLa am Beispiel der unteren Extremität

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Benjamin Haber

8. Juli 2015

Aus der Kinderchirurgie
Dr. med. habil. Bertram Reingruber
Lehrstuhl für Chirurgie
Prof. Dr. med. Hans J. Schlitt
der medizinischen Fakultät
der Universität Regensburg

Klassifikation kindlicher Frakturen

Zuverlässigkeit und Anwenderfreundlichkeit der Systematik nach AO und LiLa am Beispiel der unteren Extremität

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Benjamin Haber

8. Juli 2015

Dekan:	Prof. Dr. Dr. Torsten E. Reichert
1. Berichterstatter:	Dr. med. habil. Bertram Reingruber
2. Berichterstatter:	PD Dr. med. Jörg Marienhagen
Tag der mündlichen Prüfung:	16.10.2015/20.10.2015

Inhaltsverzeichnis

Abkürzungsverzeichnis	6
Zusammenfassung	7
1. Einleitung	9
1.1 Allgemeines	9
1.2 Geschichte der kindlichen Frakturklassifikation	10
1.3 Fragestellung	11
1.4 AO-Klassifikation	13
1.5 LiLa-Klassifikation	18
2. Material und Methoden	22
2.1 Zusammenstellung der Datensätze	22
2.2 Inhalt der Untersuchungs-CD	23
2.3 Anforderung an die Untersucher	24
2.4 Durchführung der Klassifikation	24
2.5 Statistische Auswertung	25
2.6 Dateneingabe	27
2.7 Interpretation der Werte	28
3. Ergebnisse	30
3.1 Interobserver Zuverlässigkeit Gruppe Studenten	30
3.2 Interobserver Zuverlässigkeit Gruppe Ärzte	31
3.3 Spezielle Aspekte der Auswertung	33
3.3.1 Vergleich Studenten und Ärzte	33
3.3.1.1 Femurfrakturen	33
3.3.1.2 Tibiafrakturen	35
3.3.2 Auswertung der 5. Stelle	37
4. Diskussion	40
Literaturverzeichnis	49
Abbildungsverzeichnis	55
Tabellenverzeichnis	55
Anhang	56
Danksagung	59

Abkürzungsverzeichnis

AO - Arbeitsgemeinschaft für Osteosynthesefragen

AOCID - AO Clinical Investigation and Documentation

a.p. - anterior posterior

bzw. - beziehungsweise

CD - Compact Disc

iAGKT - International Working Group for Paediatric Traumatology

ID - identification

LiLa e.V. - Licht und Lachen für Kinder in der Medizin, Effizienz in der Medizin e.V.

o.g. - oben genannte

PAEG - AO Paediatric Expert Group

PJ - Praktisches Jahr

SPSS – Statistical Package for the Social Sciences

Zusammenfassung

Zur Klassifikation von Frakturen der langen Röhrenknochen im Wachstumsalter existieren zwei Varianten einer universellen Systematik, die „AO-Klassifikation für Kinder“ und die „LiLa-Klassifikation“. Mit dieser Untersuchung soll zum einen die Anwenderfreundlichkeit beider Klassifikationssysteme erhoben werden, zum anderen deren Zuverlässigkeit. Für diesen Zweck wurden zwei Subgruppen von Probanden mit unterschiedlichem Erfahrungsgrad untersucht, um so die Bedeutung von Erfahrung auf die Klassifikationszuverlässigkeit zu überprüfen und deren Übereinstimmung mittels Fleiss-Kappa-Werten zu berechnen.

Beide Klassifikationssysteme unterscheiden die Fraktur anhand von sechs Stellen. So findet bei beiden Systematiken an Stelle 1 und 2 eine Einteilung bzgl. der Lokalisation im Skelett und im Knochen statt. Bei der weiteren Einteilung unterscheiden sich die Systeme. In der AO-Klassifikation wird eine weitere Einteilung nach Epiphyse/Metaphyse/Diaphyse an Stelle 3 und nach Schweregrad an Stelle 5 vorgenommen. Die LiLa-Klassifikation entscheidet hingegen an Stelle 3 nach artikulär/diaphysär und an Stelle 5 nach dem Dislokationsausmaß. Stelle 4 unterscheidet in beiden Systematiken nach besonderen Frakturtypen (Salter Harris 1-4, Two Plane Fraktur, etc.). Die 6. Stelle wird bei beiden Klassifikationssystemen als Ausnahmeziffer angegeben. So werden bei der AO-Klassifikation im Falle besonderer Frakturtypen weitere Details wie Angulation, Dislokation oder Verlauf der Frakturlinie klassifiziert. Bei der LiLa-Klassifikation wird im Falle von Frakturen paariger Knochen bei der Klassifizierung des nicht tragenden Knochens an 6. Stelle ein U für Ulna bzw. ein F für Fibula angegeben.

Von den Probanden wurden insgesamt 24 Röntgenbildersätze mit a.p. und seitlichen Aufnahmen der frakturierten langen Röhrenknochen von Kindern, die das 18. Lebensjahr noch nicht begonnen und offene Epiphysenfugen hatten, klassifiziert. Ohne ein vorangegangenes Training wurden die Bilder von jedem Probanden nach Lektüre der Klassifikationsstandards beurteilt. Die Probanden gehörten entweder der Subgruppe der Anfänger, bestehend aus Studenten im letzten klinischen Jahr, oder der Erfahrenen, bestehend aus kindertraumatologisch tätigen Ärzten, an. Zur Objektivierung der Übereinstimmung der Klassifikationscodes der teilnehmenden Probanden im Sinne der Interobserver-Reliabilität wurden die Fleiss-Kappa-Werte mit Hilfe eines SPSS-Makros berechnet. Der Wert 0 entspricht der Zufallswahrscheinlichkeit, der Wert + 1,0 der maximalen Übereinstimmung, ein Wert < 0 entspricht einer Übereinstimmung geringer als die Zufallswahrscheinlichkeit. Die beiden Hauptunterschiede der Klassifikationssysteme sind die Stellen 3 und 5. Diesen

wurde besondere Aufmerksamkeit geschenkt, um eventuelle Vorteile eines der beiden Systeme herauszuarbeiten.

Beim Vergleich der Klassifikationssysteme und der erzielten Kappa-Werte lässt sich in vier Unterpunkten der Auswertung eine signifikante Überlegenheit der AO-Klassifikation feststellen. Die Befragung der Probanden ergab jedoch eine höhere Anwenderfreundlichkeit für die LiLa-Klassifikation, da diese einfacher im Vorgehen des Klassifizierungsprozesses und nützlicher für den klinischen Alltag sei. Die Analyse der 3. Stelle zeigte einen signifikant höheren Kappa-Wert bei der AO-Klassifikation. Diese ist jedoch nicht mit der 3. Stelle der LiLa-Klassifikation vergleichbar, da diese einen unterschiedlichen Sachverhalt klassifiziert, der eine größere Erfahrung beim Anwender voraussetzt. Beim Vergleich der Anwendergruppen zeigten sich bei der AO-Klassifikation keine signifikanten Unterschiede, sodass eine Klassifizierung bis inklusive Stelle 3 mit validen Ergebnissen auch von traumatologisch unerfahrenen Probanden möglich ist. Die Auswertung der 5. Stelle erbrachte trotz einer subjektiv einzuschätzenden und somit erfahrungsgeprägten Dislokation zwei signifikant höhere Kappa-Werte für die LiLa-Klassifikation. Eine Erklärung hierfür können wir nicht abgeben, jedoch empfehlen wir aufgrund der Therapierelevanz die 5. Stelle der LiLa-Klassifikation erfahrenen Traumatologen zu überlassen. Die Bedeutung der Erfahrung der Anwender für die Klassifikationszuverlässigkeit konnte durch insgesamt acht signifikant höhere Kappa-Werte (vier innerhalb der LiLa- und vier innerhalb der AO-Klassifikation) in der Gruppe der Ärzte im Vergleich zur Studentengruppe nachgewiesen werden. Dieses Ergebnis ist kritisch zu werten, da eine Abhängigkeit der Zuverlässigkeit einer Klassifikation von der Erfahrung seiner Anwender die Objektivität eines Klassifikationssystems limitiert.

1 Einleitung

1.1 Allgemeines

“Perhaps the day is near, when surgeons will speak, or at least read a lingua franca of injury classification” [15]

Das Oxford English Dictionary definiert *to classify* als “arrange in classes according to shared characteristics”. Ins Deutsche lässt sich dies als „Anordnen in Klassen entsprechend gemeinsamer Merkmale“ übersetzen. Auch für die Klassifikation von Frakturen ist diese Definition zutreffend. So werden die Frakturen entsprechend ihrer Lokalisation, Morphologie und Schwere in verschiedene Klassen eingestuft und als alphanumerischer Code verschlüsselt.

Wenn es um die Frage des Nutzens der Klassifikationssysteme geht, können viele verschiedene Vorteile aufgeführt werden. So ist Rüedi 2008 [37] der Meinung, dass die Systematiken „die Kommunikation zwischen den Ärzten erleichtern und der Dokumentation und Forschung dienen“. Diese Dokumentation beinhaltet auch die Qualitätskontrolle und die Qualitätssicherung, die laut v. Laer [47] ein weiteres Argument für eine kindliche Frakturklassifikation darstellt. Audigé et al. 2005 [6] gehen weiter und stellen fest: „Das Klassifizieren einer Fraktur ist gleichbedeutend mit dem Fällen einer Entscheidung.“ Im Bereich der Forschung wäre besonders der Vergleich von Ergebnissen verschiedener Behandlungsansätze bei gleichen Frakturtypen von entscheidender Bedeutung [8] [45] [51], um durch deren Evaluierung allgemeingültige Richtlinien der Indikationsstellung zur Behandlung bestimmter Frakturtypen erstellen zu können. Um Internationalität zu gewährleisten, stellten Colton et al. 1991 [15] an Klassifikationssysteme die Anforderung, dass sie „ungeachtet der Muttersprache des Anwenders, international verständlich sein müssen“, denn nur dann könne man internationale Zusammenarbeit vorantreiben und vom Wissen anderer profitieren.

Frakturklassifikationssysteme sind in der Erwachsenen-Traumatologie fester Bestandteil der Befunddokumentation, -kommunikation und Grundlage der weiteren therapeutischen Vorgehensweise. So gibt es „wohl keine einzige Fraktur eines Knochens, zu der man nicht den Versuch einer Klassifikation unternommen hätte“ [32]. Da bei erwachsenen Patienten der Nutzen der Klassifikationen ohne Zweifel besteht, wurden auch für die Kindertraumatologie Klassifikationssysteme für bestimmte Knochen bzw. Regionen entwickelt, die jedoch

aufgrund uneinheitlicher Systematik zum Teil für Verwirrung sorgen. So gibt es allein für die Einteilung der suprakondylären Humerusfraktur bei Kindern sechs verschiedene Systematiken: nach Baumann, Felsenreich, Lubinus, Gartland, Holmberg und v. Laer [52]. Um diese große Zahl konkurrierender Klassifikationssysteme einzelner Knochen zu verringern, machten es sich zwei internationale Expertenteams zur Aufgabe, ein universelles System für sämtliche Frakturen an langen Röhrenknochen auch für das Wachstumsalter zu definieren.

1.2 Geschichte der kindlichen Frakturklassifikationen

Bereits im Jahr 1990 begann man in der Arbeitsgemeinschaft für Osteosynthesfragen (AO), sich Gedanken über die Entwicklung einer kinderspezifischen Frakturklassifikation zu machen und stellte noch im selben Jahr den sogenannten „AO Child Survey“ vor. Dieser orientierte sich an der AO-Klassifikation für Erwachsene und „enthielt 10 zusätzliche, hauptsächlich morphologische, kinderspezifische Elemente“ [47]. Dieses Klassifikationssystem konnte sich aber im klinischen Alltag nicht durchsetzen und wurde bald als unzureichend befunden, da es nicht adäquat auf die kindermorphologischen Besonderheiten einging. Die Begründer der beiden kinderspezifischen Frakturklassifikationssysteme, T. Slongo (AO-Klassifikation) und L. von Laer (LiLa-Klassifikation) waren sich einig, dass „der Heilungsprozess von kindlichen Frakturen bezüglich verschiedener Behandlungsstrategien Gegenstand prospektiver und multizentrischer Untersuchungen und Dokumentationen sein muss“ [40]. Daher sollte auch für Kinder ein universelles Frakturklassifikationssystem der langen Röhrenknochen eingeführt werden. Im Jahr 2001 begann die Entwicklung unter Mitwirkung von T. Slongo und der AO Paediatric Expert Group (PAEG) in Zusammenarbeit mit der AO Clinical Investigation and Documentation (AOCID) und der International Working Group for Paediatric Traumatology (iAGKT) auf Basis der AO-Klassifikation für lange Röhrenknochen für Erwachsene [43], da diese „sehr bekannt ist und auf breite Akzeptanz stößt“ [41]. Auch zu dieser Arbeit wurden von Slongo et al. im Jahr 2006 [41] erste Veröffentlichungen vorgenommen.

Laut Slongo et al. 2007 [42] habe die vorgeschlagene AO-Klassifikation für kindliche Frakturen bereits einen langen Validierungsprozess hinter sich. In einer ersten Phase wurden vier Pilotstudien durchgeführt, in deren Rahmen erfahrene Chirurgen kindliche Frakturen

klassifizierten, um die Klassifikation in Feinheiten anzupassen. In einer zweiten Phase nahmen 70 Chirurgen aus 15 Kliniken in fünf Ländern an einer Online-Studie teil. Hierbei wurden vielversprechende Ergebnisse erzielt, sodass eine weitere Evaluierung des Klassifikationssystems vorangetrieben werden konnte, wie es auch Slongo et al. 2007 [42] in ihrer Arbeit verlangen.

Der Verein „Licht und Lachen für Kinder in der Medizin, Effizienz in der Medizin e.V.“ (LiLa e.V.) hat es sich „zur Aufgabe gemacht, ein gemeinsames überdisziplinäres Qualitätsverständnis in der medizinischen Grundversorgung von Kindern zu formulieren“ [50], da „auch heute noch ca. 50% aller Kinder von Therapeuten behandelt werden, die für die Behandlung von Kindern nie aus- oder weitergebildet worden sind“ [28]. Um einen Beitrag zur Qualitätsverbesserung zu leisten, entwickelte der gemeinnützige Verein LiLa e.V. in Zusammenarbeit mit dem Institut für evaluative Forschung in orthopädischer Chirurgie der Universität Bern im Jahr 2000 die ersten Vorschläge für ein kindliches Frakturklassifikationssystem, welches die Besonderheiten des Skeletts im Wachstum und der kindlichen Frakturheilung mit einbezog [50]. Die entstandene LiLa-Klassifikation steht seit 2006 im Internet unter der Seite www.li-la.org der Öffentlichkeit zu Verfügung.

Auch die LiLa-Klassifikation wurde in mehreren Phasen von Pilotstudien auf ihre klinische Anwendbarkeit untersucht. So nahmen in einer ersten Pilotstudien-Phase von Januar bis April 2003 15 unfall- und kinderchirurgische Kliniken teil. Hier konnten bereits erste Mängel aufgezeigt und der Aufwand der Klassifikation ermittelt werden. Durch Einarbeitung der Ergebnisse der ersten Pilotstudie wurde das Klassifikationssystem vereinfacht und komprimiert. In einer dritten Phase der Evaluierung fand eine Prüfung auf „klinische Anwendbarkeit, Effektivität und Effizienz“ [49] statt. Nach diesem Prozess der Evaluierung und Testung konnte die endgültige Form des Klassifikationssystems Anfang 2006 für die allgemeine klinische Anwendung eingeführt werden.

1.3 Fragestellung

“What driving force is there to improve our classification system, if not our desire to consistently test and evaluate those systems that we use and to discard those that are flawed?”

Albert H. Burstein, Ph.D. in [9]

Beide Klassifikationssysteme zeigten in den ersten Phasen des Validierungsprozesses gute Ergebnisse. Unsere Gegenüberstellung der zwei Systematiken soll deren klinische Praktikabilität hinterfragen, wie es Lichtenhahn et al. 1992 [27] in ihrer Studie zur Benutzerfreundlichkeit der AO-Klassifikation darlegten. So wollen wir aufzeigen, welche Unterschiede diese beiden Systeme aufweisen und untersuchen, an welchen Stellen unterschiedliche Übereinstimmungswerte der Untersucher erzielt werden. Zu Beginn dieser Untersuchung lagen keine Studien vor, in welchen die beiden Klassifikationssysteme miteinander verglichen wurden.

Der zweite Teil unserer Studie befasst sich mit dem Thema der Erfahrung und deren Einfluss auf den Klassifikationsprozess. In beiden Klassifikationssystemen gehen an unterschiedlichen Positionen (siehe 1.4. und 1.5.) subjektive Einschätzungen in die Klassifikation mit ein. Ob genau an diesen Stellen die Erfahrung des Untersuchers einen entscheidenden Unterschied macht, soll untersucht werden, denn das Ziel einer Klassifikation sollten - um Objektivität und Zuverlässigkeit gewährleisten zu können - untersucherunabhängige Ergebnisse sein. Über die Bedeutung der Erfahrung bei der Klassifikation von Frakturen liegen verschiedene Arbeiten vor. So schreiben Wen Jie Jin et al. 2005 [53], dass die Erfahrung des Untersuchers ein entscheidender Faktor sein kann, der die Verlässlichkeit beeinflussen kann. Ähnliche Ergebnisse zeigen auch eine Reihe anderer Studien [25] [7] [30]. Gegensätzliche Ergebnisse brachte die Studie von Johnstone et al. 1993 [20] hervor: die Gruppe der Erfahrenen hatte genauso viele Fehler bei der Klassifizierung der Röntgenbilder gemacht wie die gänzlich Unerfahrenen. Einen sehr geringen bzw. gar keinen Zusammenhang zwischen der Erfahrung des Untersuchers und der Richtigkeit seiner Klassifizierung stellten auch andere Studien fest [39] [8] [13] [19] [24] [1] [2].

Zusätzlich zu der objektiven Auswertung der Klassifikation soll hier jedoch auch eine subjektive Einschätzung der Praktikabilität der zwei Klassifikationssysteme im Verständnis und der Umsetzung im klinischen Alltag abgegeben werden. Diese umfasst verschiedene Aspekte wie zum Beispiel den klinischen Nutzen und die Möglichkeit einer eventuellen Therapieentscheidung anhand des Klassifikations-Codes.

1.4 AO-Klassifikation für Frakturen im Kindesalter

Da die AO Klassifikation für Erwachsene weltweit in vielen Kliniken im täglichen Gebrauch eingesetzt wird, war es ein Gedanke der Entwickler, „an einigen Gemeinsamkeiten mit der AO-Klassifikation für Erwachsene festzuhalten“ [41]. Aus diesem Grund ist die AO Klassifikation für Frakturen im Kindesalter der für Frakturen bei Erwachsenen sehr ähnlich und erleichtert somit die Einführung in den klinischen Alltag. Um einen ersten Überblick über das Klassifikationssystem zu bekommen, ist unten eine schematische Übersicht abgebildet (Abb. 1). Die genaue Gliederung ist der Arbeit von Slongo 2007 [44] entnommen.

Hier wird ein kurzer Überblick über das Klassifikationssystem gegeben:

Position 1-3 stehen für die grobe Orientierung am frakturierten Knochen. Hier wird nach betroffenem Knochen, nach Lagebezeichnung und nach Knochenabschnitt unterschieden. Somit wird beim kurzen Blick auf den Verschlüsselungscode bereits ein Überblick über den Ort der Fraktur gegeben.

Position 4-6 stehen für die kindermorphologischen Besonderheiten. Hier wird in kindliche Frakturmuster, Frakturschwere und Ausnahmen bei besonderen Frakturen unterteilt.

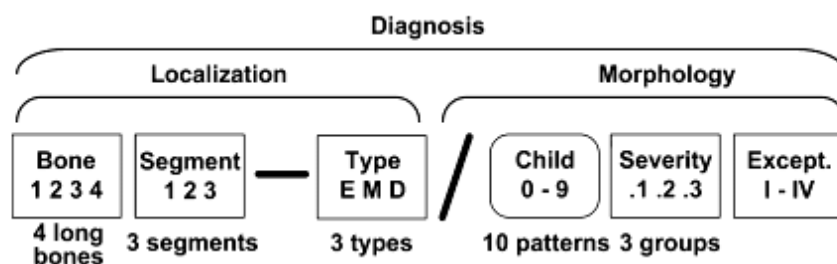


Abbildung 1: systematische Übersicht AO-Klassifikation für Frakturen im Kindesalter

An 1. Position werden die einzelnen Knochen entsprechend der Lage im Skelett nummeriert:

- 1 - Humerus
- 2 - Radius/Ulna,
- 3 - Femur
- 4 - Tibia/Fibula

Bei Verletzung paariger Knochen mit demselben Verletzungsmuster (Kindercode), welche mit dem gleichen Frakturcode klassifiziert werden, wird stets die schwerwiegendere Fraktur angegeben [31]. Sind beide Knochen beteiligt, jedoch von einem unterschiedlichen Verletzungsmuster gekennzeichnet, so müssen beide Frakturen separat mit einem eigenen Code versehen werden. Ist bei Ziffer 2 bzw. 4 nur ein Knochen beteiligt, so wird hinter dem Segmentcode eine zusätzliche Position (r, u, t, f) angegeben:

So steht zum Beispiel 43t für eine isolierte distale Tibiafraktur.

An 2. Position ist der Segmentcode aufgeführt.

1 - proximale Fraktur

2 - diaphysäre Fraktur

3 - distale Fraktur

Epiphysäre und metaphysäre Frakturen sind unter proximaler bzw. distaler Lokalisation zusammenzufassen.

Die 3. Position unterscheidet nach

E - Epiphyse

M - Metaphyse

D - Diaphyse

Hierbei wird die Metaphyse als Quadrat über der Epiphysenfuge definiert, in welchem das Zentrum der Fraktur liegen muss. Bei paarigen Knochen umfasst dieses Quadrat die Wachstumsfugen beider Knochen. Ausnahme bildet hier das proximale Femur, bei dem sich die Metaphyse als Bereich zwischen Femurkopf und der intertrochantären Linie des Femurs definiert.

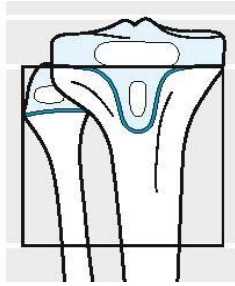


Abbildung 2: Darstellung des Metaphysenquadrates bei paarigen Knochen (Marzi I. [31])

Als 4. Position wurde der Kinder-Code eingeführt, in dem spezifische Verletzungsmuster der kindlichen Frakturen, entsprechend der Lokalisation nach E-M-D, eingearbeitet wurden.

So wird zwischen neun verschiedenen epiphysären Frakturen unterschieden:

E 1-4: Salter Harris 1-4

E 5: Two Plane Fraktur (Übergangsfaktur)

E 6: Tri Plane Fraktur (Übergangsfaktur)

E 7: Epiphysärer Bandausriss

E 8: Flake Fraktur

E 9: Sonstige epiphysäre Frakturen

Im Bereich der Metaphyse wiederum werden vier verschiedene Frakturarten unterschieden:

M 2: Grünholz-/ Wulst- und Spiralfaktur

M 3: Komplette Fraktur

M 7: osteo- oder muskuloligamentäre Avulsionen, einfache Avulsionsverletzungen

M 9: Sonstige metaphysäre Frakturen

Im diaphysären Bereich unterscheidet man neun verschiedene Frakturen:

D 1: Bowing Frakturen

D 2: Grünholz Frakturen

D 3: Toddler's Fraktur

D 4: komplette Querfraktur (Winkel zwischen Lot zu Kortikalis und Frakturspalt $< 30^\circ$)

D 5: Schräg- und Spiralfraktur (Winkel $> 30^\circ$)

D 6: Monteggia Fraktur

D 7: Galeazzi Fraktur

D 9: andere diaphysäre Frakturen

Als 5. Position wird die Frakturkomplexität kategorisiert.

.1: einfache Fraktur

.2: Keilfrakturen bzw. partiell instabile Frakturen mit drei Fragmenten einschließlich eines vollständig dislozierten Fragmentes

.3: vollständig instabile Fraktur mit mehr als drei Fragmenten

Die 6. Position wird als Ausnahmeziffer bezeichnet. Die Dislokation oder Rotation ist hier unmittelbare Grundlage der Therapieindikation. So gibt es vier besondere Frakturen, bei denen eine Ausnahmeziffer nötig ist.

Bei der suprakondylären Humerusfraktur (13-M/3) wird wie folgt nach Dislokationsausmaß und Rotation der Fragmente unterschieden:

I. Keine Dislokation

II. Dislokation in einer Ebene

- III. Rotation des distalen Fragmentes mit Dislokation in zwei Ebenen
- IV. Rotation und Dislokation in drei Ebenen oder vollständige Dislokation ohne jeglichen Knochenkontakt

Bei Radiushalsfrakturen (21-M/2 oder /3; 21-E/1 oder /2) unterscheidet man nach Dislokationsausmaß und Angulation:

- I. Keine Angulation oder Dislokation
- II. Angulation mit Dislokation weniger als eine halbe Schaftbreite
- III. Angulation mit Dislokation mit mehr als eine halbe Schaftbreite

Schenkelhalsfrakturen werden als metaphysär bezeichnet (31-M/3) und je nach Verlauf der Frakturlinie unterschieden:

- I. Transzervikal
- II. Basozervikal
- III. Pertrochantär

Bei Epiphysenlösungen des Schenkelhalses wird entsprechend der Einteilung nach Salter Harris unterschieden:

E/1 ohne metaphysärem Keil

E/2 mit metaphysärem Keil

Die Codierung eines Röntgenbildes mit Hilfe der AO-Klassifikation für Frakturen im Kindesalter beinhaltet im Regelfall fünf Positionen. In den vier oben genannten Ausnahmefällen würde als Ausnahme-/Zusatzcode eine sechste Position hinzukommen.

1.5 LiLa-Klassifikation

Auch die LiLa-Klassifikation ist im Aufbau an die AO-Klassifikation für Erwachsene angelehnt. Zusätzlich wird für jede Fraktur das Dislokationsausmaß bestimmt und eine Unterscheidung zwischen Gelenk- und Schaftfrakturen getroffen. Die Übersicht, die einen ersten Eindruck des Klassifikationssystems geben soll, und die genaue Gliederung sind der Arbeit von v. Laer 2007 [50] entnommen.

Die Systematik kurz zusammengefasst:

- 1.-2. Stelle stehen für Lokalisation im Skelett und am Knochen
- 3.-4. Stelle unterscheiden nach Gelenk und Schaft und deren speziellen Frakturen
5. Stelle unterscheidet drei Dislokationsausmaße
6. Stelle als Ausnahmeziffer bei paarigen Knochen

LiLa Klassifikation Version 2

1. Stelle Lokalisation im Skelett: (1-4)	2. Stelle Lokalisation im Knochen (Segment): (1-3)	3. Stelle Morphologie: – Gelenk (a) – Schaft (s)	4. Stelle Spezifizierungen Morphologie: – Gelenk (1-5) – Schaft (1-5)	5. Stelle Dislokationsausmass: – undisloziert (0) – tolerabel (1) – nicht tolerabel (2)	6. Stelle (Ausnahme) paariger Knochen nicht tragend – Ulna (U) – Fibula (F)
---	--	--	--	---	---

Abbildung 3: systematische Übersicht der LiLa-Klassifikation

Die 1. Stelle steht für die Lokalisation im Skelett des Kindes und wird entsprechend der AO-Klassifikation für Erwachsene vergeben

1 - Humerus

2 - Radius/Ulna

3 - Femur

4 - Tibia/Fibula

Die 2. Stelle steht für eine weitere Unterteilung des Knochens. Sie teilt diesen in drei Abschnitte auf. Hierbei wird die Grenze zur Mitte, am kaudalen und proximalen Ende des Knochens, durch das Quadrat über der Epiphysenfuge begrenzt.

1 - proximal

2 - Mitte

3 - distal

Die 3. Stelle steht für eine Unterscheidung zwischen Gelenk- und Schaftfrakturen, um die therapeutisch unterschiedlichen Ansätze bei den jeweiligen Frakturen zu berücksichtigen.

a- Gelenk

s- Schaft

Epiphysäre Frakturen zählen zu den Gelenkfrakturen, wohingegen Epiphysenlösungen zu den Schaftfrakturen gezählt werden, d.h. eine a-Fraktur liegt nur dann vor, wenn die Gelenkfläche des entsprechenden Knochens frakturiert ist.

Die 4. Stelle markiert die unterschiedlichen Frakturtypen, jeweils separat für Gelenk und Schaftfrakturen.

Artikuläre Frakturen

1 - Salter Harris 3 (offene Epiphysenfuge)

2 - Salter Harris 4 (offene Epiphysenfuge)

3 - Two Plane Fraktur (beginnender Epiphysenschluss)

4 - Tri Plane Fraktur (beginnender Epiphysenschluss)

5 - Andere

Schaftfrakturen

- 1 - Salter Harris 1/2
- 2 - Metaphysäre Grünholz- und Wulstfraktur, diaphysäre Grünholz- und Bowingfraktur
- 3 - Quer-, Schräg- und Torsionsfraktur
- 4 - Mehrfragmentfraktur
- 5 - Andere

Auch hier wird die Grenze zwischen Metaphyse und Schaft durch das Quadrat über der Epiphyse bestimmt.

Aufgrund der großen Frakturvariationen und der speziellen Anatomie des distalen Humerus wird dieser gesondert klassifiziert.

- 1 - Condylus radialis Fraktur
- 2 - Y- Fraktur
- 3 - Condylus ulnaris Fraktur
- 5 - Andere

Die 5. Stelle unterscheidet die Tolerabilität der Dislokation. Dies jedoch ohne konkrete Angaben, was als tolerabel oder nicht tolerabel disloziert zu gelten hat.

- 0 - nicht disloziert
- 1 - tolerabel disloziert
- 2 - nicht tolerabel disloziert

Die 6. Stelle steht für die Codierung von paarigen Knochen. Bei diesen wird prinzipiell der tragende Knochen codiert. Soll der nicht tragende Knochen klassifiziert werden, so erfolgt folgender Zusatz

F - Fibula

U - Ulna

Auch bei der Codierung von Röntgenbildern mithilfe der LiLa-Klassifikation benötigt man fünf Stellen, um die Fraktur genau zu beschreiben. Nur bei der Klassifikation von nicht tragenden Knochen bedarf es einer zusätzlichen Stelle.

2 Material und Methoden

2.1 Zusammenstellung der Datensätze

Für diese Arbeit wurden ausschließlich Röntgenbilder von Frakturen verwendet, die im Jahr 2007 in der Klinik für Kinderchirurgie, Klinik St. Hedwig, Krankenhaus Barmherzige Brüder, Regensburg und der Universitätsklinik Erlangen, Abteilung Kinderchirurgie, behandelt worden sind. Die Auswahl führten wir anhand der codierten DRGs und mit Hilfe des Programms SAP durch. Berücksichtigt wurden nur Bilder von Patienten, die das 18. Lebensjahr noch nicht begonnen hatten und bei denen offene Epiphysenfugen bestanden. Nach Literaturdurchsicht und Evaluierung, schlossen wir eine zweite Selektion der Bilder unter dem Kriterium der Bildqualität an. Genau diese Auswahl kritisieren Audigé et al. 2004 [4]: „Es ist angemessener alle Fälle mit einzubeziehen, unabhängig von der Bildqualität, um sich über die Qualitätskriterien einig zu werden und deren Einfluss auf die Ergebnisse auszuwerten“. In seiner Arbeit stellte er fest, dass in 39% der Studien die Röntgenbilder nach Qualitätsmerkmalen aussortiert wurden. So fand z.B. in den Studien von Sidor et al. 1993 [39], Andersen et al. 1996 [3], Illaramendi et al. 1998 [19] und Bernstein et al. 1996 [10] eine Aussortierung statt. Da es bei uns jedoch primär nicht um die korrekte Einteilung der Frakturen in die richtige Codierung geht, sondern um den Umgang und den Vergleich der Klassifikationssysteme, spielt unserer Meinung nach die Auswahl der Bilder nach Qualitätskriterien keine entscheidende Rolle. Des Weiteren wurden jene Bilderreihen, auf denen die a.p. und seitlichen Aufnahmen nur in einer schrägen und nicht standardmäßigen Position zu sehen waren, ausgeschlossen. Dazu Slongo et al. 2006 [41]: “Beim Anwenden der Quadrate sollten sich die Chirurgen bewusst sein, dass im Falle einer nicht exakten a.p. Aufnahme, z.B. einer nur leicht rotierten Aufnahme, das verwendete Quadrat kleiner als erwartet sein wird und somit ein erhöhtes Risiko der Fehlklassifizierung nach sich zieht.“

Nach der Selektion der Bilder nach den o.g. drei Kriterien waren insgesamt 24 Bildersätze mit jeweils a.p. und seitlichen Aufnahmen in unserer Studie zu klassifizieren. Die Bilder wurden als Power Point Präsentation, getrennt nach den vier Hauptlokalisationen, angeordnet. Alle personenbezogenen Daten wurden ausgeblendet, jedoch zur Orientierung für die Autoren mit einer ID versehen. Somit wurde sichergestellt, dass für die Probanden keinerlei Verbindung zu den Patienten hergestellt werden konnte.

Die Präsentation bestand aus jeweils einer a.p. und einer seitlichen Aufnahme der zu klassifizierenden Fraktur. Unter den jeweiligen Bildern wurde das Alter des Patienten angegeben und die Patienten ID, um bei möglicherweise auftretenden Problemen weitere Nachforschungen bezüglich des Patienten anstellen zu können. Als Eingabemöglichkeit für den Verschlüsselungscode platzierten wir unter jedes Bild eine Textbox, in die die entsprechenden Ziffern, jeweils für die zwei Klassifikationssysteme, eingegeben werden konnten. Des Weiteren war in Form einer Kommentar-Zeile Platz gegeben, um direkte Einwände bzw. Probleme mit dem vorliegenden Bild zu äußern, wie es auch Slongo et al. 2007 [42] in ihrer Online-Studie ermöglichten (siehe Anhang 3).

Diese Präsentation wurde mit unten aufgelisteten Zusätzen auf eine CD gebrannt und den Untersuchern ausgehändigt.

2.2 Inhalt der Untersuchungs-CD

Die Untersucher erhielten von uns jeweils eine CD mit folgenden Dateien:

1 - Anleitung zur Durchführung der Klassifikation

Hier erläuterten wir kurz das Ziel unserer Arbeit, gaben eine kurze Anleitung zum Vorgehen bei der Klassifizierung unserer Röntgenbilder und einen Verweis auf den auszufüllenden Fragebogen (siehe Anhang 1).

2/3 - Anleitung AO-/ LiLa-Klassifikation

Die Anleitung entnahmen wir dem Buch „Kindertraumatologie“ von Ingo Marzi [31].

4 - Tipps zum Umgang mit Power Point

Hier erklärten wir den Umgang mit Power Point und gaben Hilfestellung, um die Röntgenaufnahmen mit Hilfe des Programms in Helligkeit und Kontrast zu verändern, sodass eine gute Beurteilbarkeit der Bilder erreicht werden konnte.

5 - Fragebogen

Der Fragebogen wurde mit Hilfe von Microsoft Word erstellt und dient der Erfragung einzelner Daten der Untersucher. Des Weiteren soll der Umgang mit den zwei Klassifikationssystemen mit Hilfe einer Notenskala von 1-6 bewertet werden (siehe Anhang 2).

6 - Powerpoint Präsentation der acht Frakturen des Femurs

7 - Powerpoint Präsentation der 16 Frakturen der Tibia und der Fibula

Beispielfolie siehe Anhang 3

2.3 Anforderungen an den Untersucher

In einem ersten Durchgang der Studie wurden die Röntgenbilder an sechs Studenten in den letzten zwei klinischen Semestern - als Gruppe der Unerfahrenen - ausgehändigt. Wie unter 1.3. bereits erwähnt, bestehen unterschiedliche Meinungsbilder bzgl. des Einflusses der Erfahrung auf die Interobserver-Reliabilität. Um diese Ergebnisse vorangegangener Studien genauer zu untersuchen, wurden die Bildersätze in einem zweiten Durchgang an sieben traumatologisch tätige Kinderchirurgen unterschiedlichen Erfahrungsniveaus ausgehändigt. Hierunter befanden sich zwei Oberärzte und fünf Assistenzärzte. Bei einer Anzahl von insgesamt 13 Untersuchern ergeben sich somit 676 einzelne Verschlüsselungen von Frakturen. Somit sind zwei Gruppen sehr weit auseinander liegender Erfahrungslevel an der Studie beteiligt, um der Forderung von Slongo et al. 2006 [41] nachzukommen. Dieser ist der Meinung, dass Studien mit „Chirurgen verschiedener Erfahrungslevel“ durchgeführt werden müssen, „bevor endgültige Resultate dargestellt werden können.“ Da in einem Krankenhaus Ärzte vieler verschiedener Erfahrungsstufen miteinander arbeiten und untereinander kommunizieren müssen, entspricht diese Einschätzung auch unserer eigenen Auffassung des Validierungsprozesses eines klinischen Klassifikationssystems, das zum Ziel hat, Qualität im Sinne der optimalen Behandlung kindlicher Frakturen zu gewährleisten.

2.4 Durchführung der Klassifikationen

Den Studienteilnehmern wurde die CD mit o.g. Inhalt ausgehändigt und zur eigenständigen Durchsicht mit nach Hause gegeben. Zur Einführung in die Klassifikationssysteme dienten nur ausgehändigte Kopien der Systematiken aus dem Buch „Kindertraumatologie“ von Ingo Marzi [31]. Auch Gehrchen et al. 1997 [18] verzichteten auf ein explizites Training und händigten ihren Untersuchern eine schriftliche Einführung aus. Brorson et al. 2008 [13]

untersuchten den Einfluss von Training auf die Interobserver-Variation und stellten eine deutliche Besserung eben dieser fest. Da aber der klinische Alltag nicht immer Zeit lässt für eine tiefgehende Einführung in neue Methoden, wählten auch wir die kurze schriftliche Aufklärung [3] [45] [17], da uns diese einer im Alltag realistischen Lektüre von Fachliteratur am ähnlichsten erschien. Zwar waren Rasmussen et al. 1993 [36] der Meinung, dass Training das Outcome der Studie verbessere, jedoch zeigten Audigé et al. 2004 [4] in einer Metastudie, dass nur zehn von 44 Studien ein explizites Training mit ihren Studienteilnehmern vornahmen. Der Klassifikationsprozess wurde von jedem Studienteilnehmer alleine und ohne Austauschmöglichkeiten mit weiteren Studienteilnehmern durchgeführt. Des Weiteren bestand keine Möglichkeit, zu bestimmten Frakturen bzw. Unsicherheiten bezüglich der eindeutigen Einteilung einer Fraktur Fragen an die Studienleiter zu stellen. Für die Durchführung der Klassifikation setzten wir kein Zeitlimit, da nach Audigé et al. 2004 [4] nur zehn von 44 Studien den Klassifikationsprozess unter Aufsicht und Zeitlimit durchführten.

Der von uns mitgesandte Fragebogen sollte zusätzlich von den Teilnehmern ausgefüllt werden, um uns einen Überblick darüber zu verschaffen, wie sich der Umgang mit den zwei Klassifikationssystemen gestaltete. Zusätzlich sollten Angaben gemacht werden, welche der zwei Systematiken nach Meinung der Untersucher besser für den täglichen Gebrauch im klinischen Alltag zu verwenden sei.

2.5 Statistische Auswertung

Der Hauptaspekt unserer Auswertung beinhaltete die Übereinstimmung der Klassifikationscodes der einzelnen Studienteilnehmer; hierbei sollte gemessen werden, wie viele Studienteilnehmer einer Fraktur den jeweils selben Code geben. Diese Übereinstimmung entspricht der Interrater-Reliabilität.

Zur Berechnung eben dieser wird seit Cohen 1960 [14] der Kappa-Wert verwendet. “ Der Kappa-Koeffizient ist die Quote der Fälle, in welcher die Untersucher standard- und chancenkorrigiert übereinstimmen”[23]. Als Voraussetzung zur Anwendung der Kappa Berechnung geben Brennan et al. 1981 [12] folgende drei Punkte an: „a) die kategorisierten Objekte N sind unabhängig, b) die Untersucher arbeiten unabhängig, c) die Kategorien sind unabhängig, gegenseitig überschneidungsfrei und umfassend.“ Im Falle unserer Studie treffen alle Forderungen zu.

Bei Kappa handelt es sich um einen zufallskorrigierten Wert. Kappa = 0 entspricht der zufälligen Übereinstimmung. Bei größtmöglicher Übereinstimmung beträgt Kappa +1. Ist die Übereinstimmung geringer als die Zufallswahrscheinlichkeit, wird Kappa negativ. „Dieser Fall könnte z.B. eintreten, wenn die Urteiler sich bewusst darum bemühen, keine übereinstimmenden Urteile abzugeben“ [11]. Der maximal negative Wert ist -1, ist jedoch in der Ausdehnung der Negativität abhängig von der Randverteilung der Messwerte.

Zur Berechnung des Kappa-Wertes wird folgende Formel von Cohen 1960 [14] verwendet:

$$K = \frac{p(a) - p(e)}{1 - p(e)}$$

$p(a)$ = Gemessene Übereinstimmung der zwei Beobachter

$p(e)$ = Zufällig erwartete Übereinstimmung

Da man mit Hilfe des Cohens Kappa nur die Übereinstimmung zwischen zwei Beobachtern berechnen kann, jedoch häufig mehr als nur zwei Beobachter an Studien teilnehmen, machte es sich Fleiss zur Aufgabe, eine Generalisierung für mehr als zwei Beobachter zu erstellen. „Die Fleiss Erweiterung wurde über die letzten 30 Jahre zur Haupt-Maßzahl für Untersucher-Übereinstimmung“ [21]. Daher zogen auch wir diese Kappa Erweiterung zur Berechnung der Interrater-Reliabilität heran, um die Übereinstimmung der Klassifizierung unserer Studiengruppen genauer untersuchen zu können.

Um die errechneten Kappa-Werte interpretieren zu können, ist es notwendig, sie anhand von vorhandenen Vergleichstabellen, in Kategorien einzuteilen. Die meist verbreitete Interpretations-Tabelle stammt von Landis und Koch 1977 [26] und sieht folgende Einteilung vor:

Kappa-Werte	Übereinstimmung
<0	schlecht
0,0 bis 0,20	gering
0,21 bis 0,40	ordentlich
0,41 bis 0,60	mäßig
0,61 bis 0,80	gut
0,81 bis 1,00	exzellent

Tabelle 1: Interpretation der Kappa-Werte nach Landis und Koch 1977 [26]

Mit Hilfe dieser Tabelle lassen sich Vergleiche zu anderen Studien ziehen, um die erzielten Ergebnisse dieser Arbeit mit bereits vorhandenen Studien zu vergleichen.

2.6 Dateneingabe

Wir erhielten die Klassifikations-Ergebnisse der Probanden zusammen mit der von uns angefertigten Power Point Präsentation. Diese fassten wir in einer Excel-Tabelle zusammen, in welcher wir sämtliche Verschlüsselungscodes der Röntgenbilder sortiert nach Untersucher und Bild übertrugen.

Als zweiten Schritt führten wir nun eine Neucodierung der vergebenen Klassifikationscodes durch. Hierbei wiesen wir jedem Code, der mindestens einmal vergeben wurde eine neue Zahl zu, um die Anzahl der Codes auf die Verggebenen zu reduzieren. Diese Reduzierung führten wir jeweils für die Stellen 345, 34, die 3. und die 5. Stelle durch. Somit konnten wir jedem vergebenen Code eine Ziffer zuteilen, die wir dann im Weiteren verwendeten, um die Übereinstimmung der Untersucher zu beurteilen.

Im dritten Schritt gaben wir die erhobenen Ziffern in ein SPSS Makro des Baylor College of Medicine in Houston, Texas ein, welches für die Berechnung des Fleiss Generalized Kappa für mehrere Beobachter programmiert war. Um uns von der Validität des Makros zu überzeugen, nahmen wir Kontakt zu dessen Autor Jason King Ph.D., Associate Director

Office of CME (Evaluation) auf, welcher uns per E-Mail bestätigte, dass es sich um ein mehrfach validiertes und auf Kongressen bereits vorgestelltes Programm handele.

Dieses Makro berechnete die Kappa-Werte für die entsprechenden Kategorien und erstellte eine Ergebnisliste wie unten beispielhaft dargestellt:

Basic Information					
		z	P	Lower 95% Confidence Limit	Upper 95% Confidence Limit
		17,61734	,00000	,54312	,67910

Empirical Confidence Limits - Overall kappa					
kappa	Standard Error	z	p	Lower 95% Confidence Limit	Upper 95% Confidence Limit
,61111	,03469	17,61734	,00000	,54312	,67910

Tabelle 2: Ergebnisse des SPSS Makros (Femur Ärzte AO Stelle 3+4)

Bei dem oben aufgezeigten Kappa-Wert handelt es sich um den Fleiss-Kappa-Wert für mehrere Beobachter, welcher ebenfalls nach Tab. 1 interpretiert werden kann und somit einen guten Anhaltspunkt für die Übereinstimmung einer Bewertung durch mehrere Beobachter darstellt.

2.7 Interpretation der Werte

Mithilfe der durch das Makro berechneten Konfidenzintervalle, Lower 95% Confidence Limit und Upper 95% Confidence Limit, lässt sich eine Aussage über die Signifikanz der einzelnen Kappa-Werte machen. Die statistische Signifikanz der Kappa-Werte geht mit ausschließlich positiven Werten der Konfidenzintervalle einher. „Wenn also ein Vertrauensbereich einer

Differenz die Null einschließt, so kann der Effekt in beide Richtungen gehen... Nur wenn der Effekt in eine Richtung geht, vermuten wir, dass ein signifikanter Zusammenhang besteht“ [34]. In diesem Falle liegt der tatsächliche Kappa-Wert mit einer Wahrscheinlichkeit von 95 Prozent im positiven Wertebereich und ist somit größer als die zufällig erzielte Übereinstimmung der Beobachter. Somit erfüllt dieser Wert die Ansprüche für eine statistische Signifikanz.

Um eine Aussage über den Vergleich bzw. die statistische Signifikanz des Unterschiedes zweier Werte zu machen, ist ein Vergleich der Konfidenzintervalle der jeweiligen Werte nötig. „...ob sich der vorgegebene Wert innerhalb des Konfidenzintervalles des geschätzten Wertes zum Niveau alpha befindet. Liegt er innerhalb so lautet das Test-Ergebnis: Kein Unterschied; liegt er außerhalb so lautet es: es besteht ein Unterschied. Man sagt dann auch, der Unterschied ist signifikant zum Niveau alpha“ [33]. Auch dies sei hier an einem Beispiel erläutert:

Kappa 1: Lower CI 0,56 ; Upper CI 0,78

Kappa 2 : Lower CI 0,03 ; Upper CI 0,43

Hier ist folgende Aussage möglich: Da es nicht zu einer Überschreitung der Konfidenzintervalle kommt und somit der gesamte Wertebereich, in dem Kappa 1 liegen könnte, höher ist und daher außerhalb des Konfidenzintervalles von Kappa 2 liegt, ist ein statistisch signifikanter Unterschied der Kappa-Werte festzustellen. Diese Überlegung ermöglicht beim Vergleich der zwei Klassifikationssysteme eine Aussage über die Signifikanz der Kappa-Wert-Unterschiede und somit eine Einschätzung der Überlegenheit eines der Systeme.

3 Ergebnisse

In den folgenden Abschnitten der Ergebnisdarstellung wurde eine Farbkodierung vorgenommen, um die Visualisierung der verschiedenen Stellenanzahlen zu erleichtern. Hierbei stehen die roten Ziffern für die errechneten Fleiss-Kappa-Werte der Stellen 3, 4 und 5, die blauen Ziffern für die Werte der Stellen 3 und 4 und die grünen Ziffern für die Werte berechnet aus der Stelle 3 der jeweiligen Klassifikationssysteme.

Aufgrund der Besonderheit der 5. Stelle werden die Ergebnisse der Berechnung der Kappa-Werte für Stelle 5 gesondert in Kapitel 3.3.2 dargestellt und erläutert.

Die Ergebnisse sind jeweils in Form von Tabellen dargestellt, in denen sich auf der linken Seite, in Form der Zeilen, der zu klassifizierende Knochen befindet und oben, in Form der Spalten, die jeweils angewendete Klassifikation. Zur besseren Veranschaulichung der Ergebnisse, sind die Ergebnisse jeweils noch in Form eines Balkendiagramms dargestellt.

3.1 Interobserver Zuverlässigkeit Gruppe Studenten

In Tab. 3 sind die Ergebnisse der Auswertung der Studentengruppe dargestellt, welche zwischen 0,049 und 0,598 liegen und somit eine weite Streuung aufweisen. Nach Landis und Koch 1977 [26] ergibt dies eine Übereinstimmung zwischen gering und mäßig.

	AO-Klassifikation	LiLa-Klassifikation
Femur	0,188	0,208
	0,457	0,125
	0,598	0,238
Tibia	0,16	0,08
	0,34	0,049
	0,526	0,241

Tabelle 3: Übersicht Fleiss-Kappa-Werte der Studenten Studiengruppe

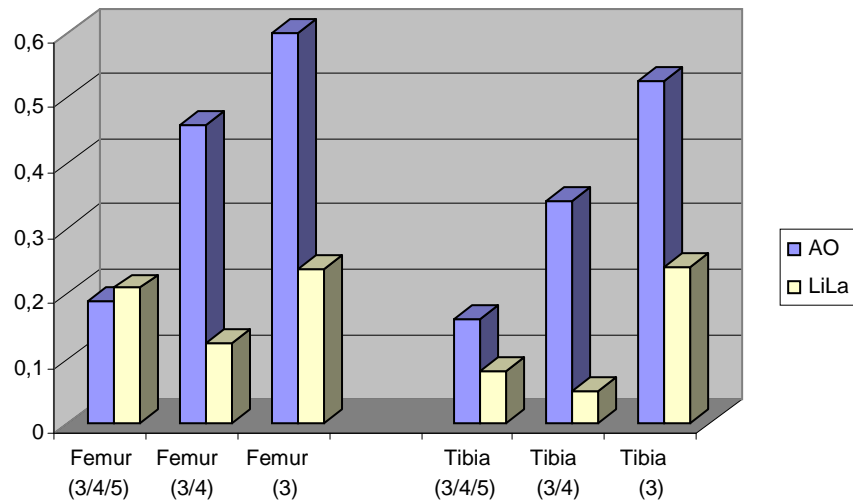


Abbildung 4: Übersicht Fleiss-Kappa-Werte der Studenten Studiengruppe

Wie aus Abb.3 zu entnehmen ist, erzielte die AO-Klassifikation in der Gruppe der Studenten deutlich höhere Werte als die LiLa-Klassifikation. Insgesamt zeigt sich eine große Diskrepanz zwischen den Kappa-Werten der zwei Klassifikationssysteme. Auch ein Einkürzen der zu klassifizierenden Stellen änderte nichts an den höheren Kappa-Werten der AO-Klassifikation.

Beim Vergleich der Werte von Femur und Tibia fallen die durchwegs höheren Werte bei der Klassifikation des Femurs auf. Des Weiteren ergeben sich bei der Kürzung der Stellen von drei auf zwei in der LiLa-Klassifikation jeweils niedrigere Kappa-Werte. Die genaue Analyse dieser drei Feststellungen wird später noch erörtert.

3.2 Interobserver Zuverlässigkeit Gruppe Ärzte

Tab. 4 zeigt eine Zusammenstellung der Kappa-Werte der Ärzte-Gruppe, welche bei der Klassifizierung der Bilder mit dem jeweiligen Klassifikationssystem erzielt wurden. Hierbei bewegen sich die Werte zwischen 0,115 und 0,673, was nach Landis und Koch 1977 [26] einer geringen bis guten Übereinstimmung entspricht.

	AO-Klassifikation	LiLa-Klassifikation
Femur	0,497	0,481
	0,611	0,532
	0,673	0,115
Tibia	0,254	0,239
	0,389	0,263
	0,621	0,368

Tabelle 4: Übersicht Fleiss-Kappa-Werte der Ärzte Studiengruppe

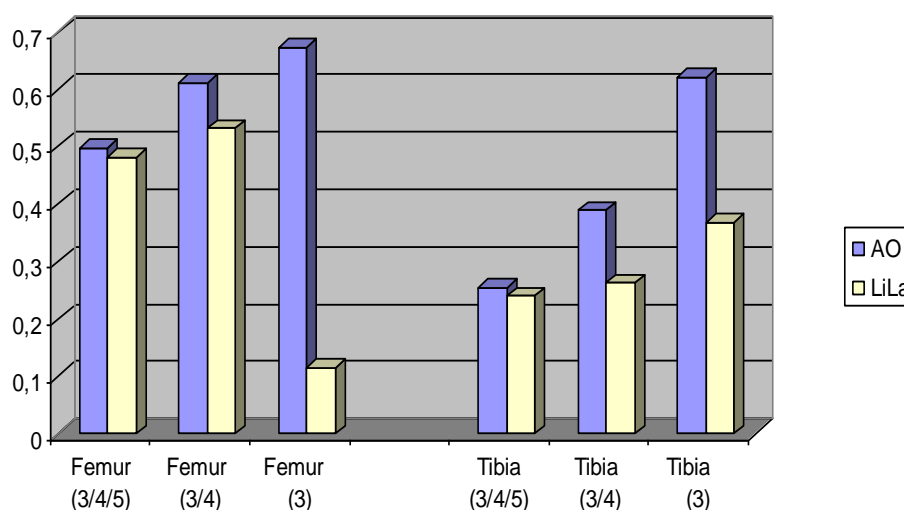


Abbildung 5: Übersicht Fleiss-Kappa-Werte der Ärzte Studiengruppe

Insgesamt erreichte die Gruppe der Ärzte deutlich höhere Kappa-Werte als die Gruppe der Studenten. Auch in der Gruppe der Ärzte erzielte die AO-Klassifikation durchgehend höhere Kappa-Werte, jedoch ist die Diskrepanz der Werte zwischen AO- und LiLa-Klassifikation nicht so deutlich ausgeprägt wie in der studentischen Vergleichsgruppe.

Beim Vergleich der Ergebnisse von Femur und Tibia fallen auch in der Gruppe der Ärzte die höheren Werte bei der Klassifizierung der Femur-Frakturen auf. Jedoch steigen im Unterschied zur Gruppe der Studenten die Kappa-Werte bei der Klassifizierung nach LiLa, bei Einkürzung der Stellen auf zwei, weiterhin an.

3.3 Spezielle Aspekte der Auswertung

3.3.1 Vergleich Studenten und Ärzte

In diesem Abschnitt sollen die Ergebnisse durch Vergleich der erreichten Kappa-Werte in der Gruppe der Studenten und Ärzte dargestellt werden. Hiermit wird auf den oben genannten Punkt der sehr häufig und kontrovers diskutierten Erfahrung im Umgang mit Klassifikationssystemen und der Traumatologie eingegangen. [53] [25] [7] [30] [39] [8] [13] [19] [24] [1] [2] [35]

Um eine Aussage über die Signifikanz des Unterschiedes zwischen Studenten und Ärzten zu ermöglichen, sind hier die jeweiligen Konfidenzintervalle angegeben. Wie in Kap. 2.7. erläutert, ist mit deren Hilfe eine Einschätzung der Signifikanz des Unterschiedes zweier Werte möglich.

3.3.1.1 Femurfrakturen

Die einzelnen Tabellen stellen jeweils die Werte mit den dazugehörigen Konfidenzintervallen der verschiedenen Stellen dar.

	Studenten	Ärzte
AO-Klassifikation	0,18 (KI 0,13-0,24)	0,49 (KI 0,44-0,55)
LiLa-Klassifikation	0,20 (KI 0,14-0,27)	0,48 (KI 0,41-0,54)

Tabelle 5: Fleiss-Kappa-Werte der Klassifikation des Femurs (Stellen 3+4+5)

Bei der Auswertung der drei Stellen zeigten sich insgesamt niedrige Kappa-Werte, die auf eine nicht sehr große Übereinstimmung der Untersucher bei der Klassifizierung der Röntgenbilder hindeuten. Nach Landis und Koch 1977 [26] erzielten die Studenten eine geringe Übereinstimmung in beiden Klassifikationssystemen. In der Gruppe der Ärzte ergibt sich aus den Werten eine mäßige Übereinstimmung.

Bei genauer Betrachtung der Konfidenzintervalle lässt sich feststellen, dass bei der Klassifizierung mittels beider Klassifikationssysteme ein signifikanter Unterschied zwischen den zwei Gruppen zu erkennen ist. Somit kann hier ein Rückschluss auf die Thematik der Erfahrung im Beruf gezogen werden. Ein statistisch signifikanter Unterschied innerhalb der einzelnen Untersuchergruppen bezüglich der zwei Klassifikationssysteme ist jedoch nicht nachweisbar.

	Studenten	Ärzte
AO-Klassifikation	0,45 (KI 0,36-0,54)	0,61 (KI 0,54-0,67)
LiLa-Klassifikation	0,12 (KI 0,02-0,22)	0,53 (KI 0,44-0,61)

Tabelle 6: Fleiss-Kappa-Werte der Klassifikation des Femurs (Stellen 3+4)

Bei der nun vorgenommenen Reduzierung der Stellen auf insgesamt zwei, wie in Tab. 6 dargestellt, zeigten sich ansteigende Kappa-Werte für beide Untersucher Gruppen. Als einziger Wert fällt der studentische Wert der LiLa-Klassifikation heraus.

In dieser Gruppe liegen die erzielten Werte bei den Studenten im geringen Bereich für die LiLa-Klassifikation und im mäßigen Bereich für die AO-Klassifikation. In der Gruppe der Ärzte wird für die AO-Klassifikation eine gute Übereinstimmung erzielt, jedoch nur eine mäßige für die LiLa-Klassifikation.

Beim Vergleich der zwei Untersuchergruppen lässt sich auch hier ein statistisch signifikanter Unterschied innerhalb des jeweiligen Klassifikationssystems feststellen. Zusätzlich zeigt sich innerhalb der Gruppe der Studenten ein statistisch signifikanter Unterschied zwischen den zwei Klassifikationssystemen mit einer Überlegenheit der AO-Klassifikation.

	Studenten	Ärzte
AO-Klassifikation	0,59 (KI 0,44-0,75)	0,67 (KI 0,56-0,78)
LiLa-Klassifikation	0,23 (KI -0,16-0,63)	0,11 (KI -0,01-0,24)

Tabelle 7: Fleiss-Kappa-Werte der Klassifikation des Femurs (3. Stelle)

Beim Vergleich der 3. Stelle zeigten sich erneut leicht ansteigende Werte in beiden Gruppen, und wiederum ein sinkender Wert, der nicht erklärt werden kann. Somit ergeben sich für die studentische Gruppe gute Übereinstimmungen für die AO-Klassifikation und mäßige für die LiLa-Klassifikation. In der Gruppe der Ärzte stieg der Kappa-Wert der AO-Klassifikation an, zeigte jedoch immer noch eine gute Übereinstimmung. Der Wert für die LiLa-Klassifikation zeigte eine nur geringe Übereinstimmung an.

Hier stellte sich bei der Signifikanzprüfung kein signifikanter Unterschied der zwei Untersuchergruppen innerhalb der AO-Klassifikation ein. Bei der LiLa-Klassifikation werden Kappa-Werte erzielt, die das Kriterium der statistischen Signifikanz aufgrund negativer Konfidenzintervalle nicht erfüllen. Daher scheint es für beide Untersuchergruppen nicht einfach zu sein, die Frakturen nach Gelenk- und Schaftfrakturen, wie in der LiLa-Klassifikation gefordert, zu unterscheiden.

Zusammenfassend lässt sich sagen, dass bei der Klassifikation von Femur-Frakturen mittels der zwei von uns untersuchten Klassifikationssysteme die Gruppe der Ärzte im direkten Vergleich mit den Studenten - bei drei bzw. zwei untersuchten Stellen - statistisch signifikant höhere Kappa-Werte erzielte. Dies können wir als wichtigen Anhaltspunkt für die Bedeutung der Erfahrung bei der Klassifizierung von Frakturen heranziehen. Als zweite Schlussfolgerung ist auf die Problematik der nicht signifikanten Kappa-Werte und der damit verbundenen schlechten Übereinstimmung der dritten Stelle der LiLa-Klassifikation hinzuweisen.

3.3.1.2 Tibiafrakturen

Zur Verdeutlichung der statistischen Signifikanz sind die Konfidenzintervalle der einzelnen Kappa-Werte angegeben.

	Studenten	Ärzte
AO-Klassifikation	0,16 (KI 0,11-0,2)	0,25 (KI 0,22-0,28)
LiLa-Klassifikation	0,08 (KI 0,04-0,12)	0,23 (KI 0,20-0,27)

Tabelle 8: Fleiss-Kappa-Werte der Klassifikation der Tibia (Stellen 3+4+5)

In Tab. 8 sind die erzielten Kappa-Werte der Auswertung aller drei Stellen tabellarisch dargestellt. Hier wurden von beiden Untersuchergruppen nur sehr geringe Kappa-Werte erreicht. So erzielte die Gruppe der Studenten bei beiden Klassifikationssystemen nach Landis und Koch 1977 [26] nur geringe Übereinstimmung. Die erzielten Werte der Ärzte spiegeln sich in einer ordentlichen Übereinstimmung wider.

Statistisch signifikante Unterschiede lassen sich auch hier zwischen Ärzten und Studenten in den jeweiligen Klassifikationssystemen nachweisen. Zusätzlich lässt sich ein schwach signifikanter Unterschied innerhalb der studentischen Gruppe mit Vorteilen für die AO-Klassifikation feststellen – jedoch insgesamt auf einem sehr niedrigen Übereinstimmungsniveau.

	Studenten	Ärzte
AO-Klassifikation	0,34 (KI 0,27-0,39)	0,38 (KI 0,35-0,42)
LiLa-Klassifikation	0,049 (KI -0,07-0,10)	0,26 (KI 0,21-0,31)

Tabelle 9: Fleiss-Kappa-Werte der Klassifikation der Tibia (Stellen 3+4)

Die hier dargestellte Übersicht zeigt die Reduzierung der Stellenanzahl auf zwei Stellen. Die erzielten Kappa-Werte sind im Vergleich mit der vorangehenden Tabelle angestiegen. Der studentische Wert der LiLa-Klassifikation fällt deutlich aus der Reihe. Mögliche Ursachen für diese Werte werden im Diskussionsteil dieser Arbeit erörtert.

Die Gruppe der Studenten erzielte für die AO-Klassifikation eine ordentliche, für die LiLa-Klassifikation jedoch nur eine geringe Übereinstimmung. Bei der Gruppe der Ärzte erzielten beide Klassifikationssysteme eine ordentliche Übereinstimmung.

Beim Vergleich der Konfidenzintervalle lässt sich nur eine schwache statistische Signifikanz des Unterschiedes zwischen der studentischen und der ärztlichen Gruppe bei der Bewertung nach AO feststellen. Das Konfidenzintervall der studentischen Gruppe der LiLa-Klassifikation zeigt negative Werte und ist somit Indikator eines nicht statistisch signifikanten Kappa-Wertes. Daher ist auch der Vergleich mit der ärztlichen Gruppe nicht möglich.

	Studenten	Ärzte
AO-Klassifikation	0,52 (KI 0,32-0,73)	0,62 (KI 0,54-0,69)
LiLa-Klassifikation	0,24 (KI -0,04-0,52)	0,36 (KI 0,26-0,47)

Tabelle 10: Fleiss-Kappa-Werte der Klassifikation der Tibia (3.Stelle)

Als letzte sei nun eine Übersicht über die erzielten Werte bei nur einer untersuchten Stelle gegeben. Auch hier zeigt sich ein Abfall des Kappa-Wertes bei den Studenten. So zeigt sich bei der LiLa-Klassifikation ein geringerer Wert als in der vorangehenden Tabelle, obwohl die Auswahl an möglichen Codes aufgrund der reduzierten Stellenzahl deutlich eingeschränkt wurde. Die Studenten erzielten für die LiLa-Klassifikation nur eine ordentliche Übereinstimmung, für die AO-Klassifikation jedoch eine mäßige. Die Gruppe der Ärzte erzielte auch hier insgesamt höhere Kappa-Werte, aber nach Landis und Koch 1977 [26] für die LiLa-Klassifikation auch nur eine ordentliche und für die AO-Klassifikation eine gute Übereinstimmung.

Bei Betrachtung der Werte lässt sich innerhalb der Gruppe der Ärzte ein signifikanter Unterschied beim Vergleich der LiLa- mit der AO-Klassifikation feststellen. Zusätzlich zeigte sich auch hier bei der Einteilung der Frakturen nach artikulär und diaphysär ein nicht signifikanter Kappa-Wert in der Gruppe der Studenten.

Als Schlussfolgerung aus diesem Abschnitt lässt sich feststellen, dass auch bei der Klassifizierung von Tibiafrakturen die Ärzte bessere Kappa-Werte erzielen. Diese weisen jedoch nur bei der Auswertung von drei Stellen einen signifikanten Unterschied zur Gruppe der Studenten auf. Auch die niedrigen bzw. nicht signifikanten Kappa-Werte der dritten Stelle der LiLa-Klassifikation innerhalb beider Untersuchergruppen sind hier aufgetreten.

3.3.2 Auswertung der 5. Stelle

Bei der fünften Stelle handelt es sich in beiden Klassifikationssystemen jeweils um eine Differenzierung nach Schweregrad. Diese wird in beiden Systemen jedoch nach unterschiedlichen Kriterien vorgenommen. Daher erfolgt die Auswertung der Ergebnisse für die 5. Stelle getrennt. Dies insbesondere zur Frage der für die Klassifizierung erforderlichen Erfahrung und der Praktikabilität in der täglichen klinischen Routine.

So wird, wie bereits in Kapitel 1.4 erläutert, in der AO-Klassifikation nach einfacher, Keil- oder komplexer Fraktur, also nach Schwere der Fraktur, differenziert. Somit bestehen 3 Auswahlmöglichkeiten, zwischen denen der Untersucher wählen muss. Auch bei der LiLa-Klassifikation (siehe Kap.1.5) bestehen drei Wahlmöglichkeiten, jedoch eine andere Art der Unterscheidung. So wird hier unterschieden in undisloziert, tolerable Dislokation und nicht tolerable Dislokation.

Die Ergebnisse werden in Form von Tabellen getrennt nach Knochen präsentiert und mit graphischen Darstellungen veranschaulicht.

	Studenten	Ärzte
AO-Klassifikation	0,26 (KI 0,12-0,39)	0,56 (KI 0,45-0,67)
LiLa-Klassifikation	0,60 (KI 0,47-0,73)	0,47 (KI 0,37-0,57)

Tabelle 11: Fleiss-Kappa-Werte der Klassifikation des Femurs (5. Stelle)

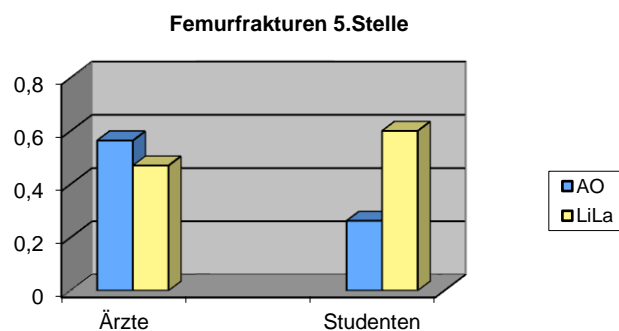


Abbildung 6: Übersicht Kappa-Werte 5. Stelle Femurfrakturen

Bei Betrachtung der Ergebnisse zeigt sich in der Gruppe der Ärzte kein signifikanter Unterschied zwischen AO- und LiLa-Klassifikation. Signifikant unterschiedlich sind die Werte in der Gruppe der Studenten. So erweist sich die LiLa-Klassifikation als signifikant überlegen gegenüber der AO-Klassifikation. Auf mögliche Ursachen und Interpretationen wird später eingegangen.

Als weiteres Ergebnis lässt sich aus obigen Werten ein signifikanter Unterschied zugunsten der Ärzte bei der Klassifikation nach AO nachweisen.

	Studenten	Ärzte
AO-Klassifikation	0,039 (KI -0,90-0,16)	0,185 (KI 0,08-0,28)
LiLa-Klassifikation	0,195 (KI 0,1-0,28)	0,415 (KI 0,33-0,49)

Tabelle 12: Fleiss-Kappa-Werte der Klassifikation der Tibia (5. Stelle)

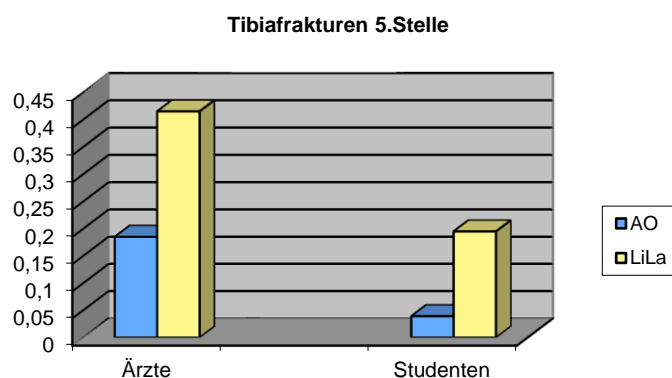


Abbildung 7: Übersicht Kappa-Werte 5. Stelle Tibiafrakturen

Auch hier zeigt sich ein signifikanter Unterschied in der Gruppe der Ärzte zu Gunsten der LiLa-Klassifikation. Zusätzlich stellen wir einen nicht statistisch signifikanten Kappa-Wert für die AO-Klassifikation der studentischen Gruppe fest.

Als weiteres Ergebnis lässt sich aus den Daten eine signifikante Überlegenheit der Ärzte gegenüber den Studenten bezüglich der LiLa-Klassifikation erkennen.

4 Diskussion

Im Rahmen dieser Arbeit soll ein Vergleich der beiden bestehenden Klassifikationssysteme für Frakturen im Wachstumsalter gezogen werden. Bisher ist in der Literatur kein Vergleich dieser zwei Systematiken beschrieben. Daher ist es uns nicht möglich, unsere Ergebnisse mit denen anderer Studien zu vergleichen und eine Bewertung der erzielten Werte zu geben. Einzig die Pilotstudien aus der Entstehungsphase der zwei Klassifikationssysteme stehen vergleichend zur Verfügung. Zusätzlich existiert eine Vielzahl an Klassifikationssystemen aus anderen Bereichen der Frakturlehre, die im Rahmen einer Metastudie von Audigé et al. 2004 [4] zusammengestellt wurden. Hier fanden sich im Schnitt fünf Untersucher mit je 50 zu klassifizierenden Bildern. Die Auswertung fand in 39 von 44 Studien mit Hilfe des Kappa-Koeffizienten statt. Somit lag der Umfang unserer Studie über dem durchschnittlichen Umfang an Untersuchern jedoch mit einer etwas geringeren Anzahl an zu klassifizierenden Frakturen. Eine entsprechende Studie der oberen Extremitäten wurde simultan am Klinikum der Universität Erlangen durchgeführt, deren Ergebnisse weiter unten zusammenfassend aufgezeigt werden [51].

Beim ersten Vergleich der zwei Untersuchergruppen fielen die höheren Kappa-Werte in der Gruppe der Ärzte in beiden Klassifikationssystemen auf. Sie lagen bei der Klassifizierung aller drei untersuchten Stellen im Bereich von 0,5 für den Femur und 0,25 für die Tibia. In der Gruppe der Studenten jedoch bei 0,2 für den Femur bzw. 0,1 für die Tibia. Somit lagen die berechneten Kappa-Werte der Untersuchergruppe der Ärzte, unabhängig vom Klassifikationssystem, deutlich über denen der Untersuchergruppe der Studenten und zeigten bei der Klassifizierung des Femurs eine mäßige und der Tibia eine ordentliche Übereinstimmung nach Landis und Koch 1977 [26]. Aufgrund dieser Werte soll hier bereits eine zentrale Frage unserer Arbeit diskutiert werden: Spielt die Erfahrung der Untersucher eine entscheidende Rolle bei der Klassifizierung der Frakturen? Unsere Untersuchungen haben gezeigt, dass in der Gruppe der Ärzte, unabhängig von der Zahl der zu klassifizierenden Stellen, durchweg höhere Kappa-Werte erzielt wurden. Bei der Signifikanzprüfung zeigten sich signifikante Unterschiede zugunsten der Untersuchergruppe der Ärzte beim Vergleich der erzielten Kappa-Werte bei der Klassifikation der Stellen 3/4/5 und 3/4 des Femurs und 3/4/5 der Tibia bei beiden Klassifikationssystemen. Auch bei Auswertung der 5. Stelle zeigte sich bei der LiLa-Klassifikation der Tibiafrakturen eine signifikante Überlegenheit der Ärztegruppe. Dasselbe Ergebnis zeigte sich bei der

Klassifizierung von Femurfrakturen mithilfe der AO- Klassifikation. In der Literatur sind gleiche Ergebnisse bei der Auswertung anderer Klassifikationssysteme beschrieben [53] [25] [7] [30]. So fanden Kristiansen et al. 1998 [25] bei der Untersuchung von proximalen Humerusfrakturen mittels der Neer-Klassifikation jeweils in der Gruppe, in der sich der Orthopäde mit der geringsten Erfahrung befand, Werte zwischen 0,07 und 0,24, welche die niedrigsten Kappa-Werte der gesamten Studie darstellten. Da wir mit unserer Gruppenauswahl im Hinblick auf die Erfahrung zwei weit auseinander liegende Gruppen gewählt haben, fallen bei uns die Unterschiede umso deutlicher aus. Daher sehen auch wir den Einfluss der Erfahrung auf die Untersucherübereinstimmung bei der Klassifizierung von Frakturen als bestätigt. Dies spricht jedoch gegen das Qualitätsmerkmal „Objektivität“ eines Klassifikationssystems, da untersucherabhängige Ergebnisse schlecht reproduzierbar sind und somit als Hilfe im klinischen Alltag von limitiertem Wert sind.

Des Weiteren stiegen in der Gruppe der Ärzte bei einer Verringerung der Stellenanzahl erwartungsgemäß die Kappa-Werte an, so wie es auch Kreder et al. 1996 [24] in ihrer Studie erwarteten. Eine Reduzierung der Auswahlmöglichkeiten vermindert ebenfalls die Möglichkeit der Fehlklassifizierung und ergibt somit höhere Übereinstimmungswerte. Dazu Marsh et al. 2007 [29]: “Mit steigender Komplexität sinkt die Untersucher Reliabilität.“ So gilt es einen Mittelweg zu finden zwischen einerseits höchst möglicher Genauigkeit bei der Beschreibung einer Fraktur im akademischen Sinne, und andererseits einer übersichtlichen und einfachen Frakturbeschreibung, wie sie für den täglichen klinischen Gebrauch notwendig ist. Auch die Klassifizierung einer Fraktur mittels zweier Systeme im Sinne beider oben genannter Ziele, wie von Bernstein et al. 1996 [10] vorgeschlagen, könnte die Problematik der Komplexität im klinischen Alltag lösen.

Beim Vergleich der zwei Klassifikationssysteme miteinander lässt sich feststellen, dass die AO-Klassifikation höhere Kappa-Werte erzielte. So zeigte sich bei der Klassifizierung der Stellen 3/4/5 der Tibia in der Gruppe der Studenten ein signifikanter Unterschied zu Gunsten der AO-Klassifikation. Des Weiteren lassen sich signifikante Differenzen bei der Bewertung der Stellen 3/4 des Femurs in der Gruppe der Studenten und der Tibia in der Gruppe der Ärzte feststellen. Zusätzlich ist ein signifikanter Unterschied bei Stelle 3 der Tibia in der Gruppe der Ärzte festzustellen. Somit ergab sich bei vergleichender Auswertung der Kappa-Werte beider Systeme auf vier Positionen eine signifikante Überlegenheit der AO-Klassifikation.

Die erhobene Umfrage sollte uns neben den errechneten Kappa-Werten Aufschluss darüber geben, welche der Klassifikationssysteme in der Anwendung leichter erschien und welche als

Standard für den klinischen Alltag zu bevorzugen ist. Die Befragung ergab das Gegenteil dessen, was die Auswertung mittels der Kappa-Werte ergab. Die Mehrheit in beiden Untersucher Gruppen war der Meinung, dass die LiLa-Klassifikation zu bevorzugen sei. Das am häufigsten vorgebrachte Argument in der Gruppe der Ärzte war der große Nutzen des Dislokationsausmaßes für den Fortgang der Therapie. Ähnlich äußert sich auch eine Reihe von Autoren zur Therapie von Schenkelhalsfrakturen [16] [38] [54]. Aber auch in der kindlichen Traumatologie spielt das Dislokationsausmaß eine entscheidende Rolle [31]. Des Weiteren wurde von beiden Untersuchergruppen die Einfachheit der LiLa-Klassifikation gelobt. So äußerte sich ein Teilnehmer über die LiLa-Klassifikation: „An sich einfacheres System mit klarer Aussage gegenüber AO (5. Stelle tolerabel oder nicht). Im Praxisgebrauch etwas weniger komplex. Jedoch ist für eine klare Entscheidung nur das Fachwissen entscheidend. Eine Klassifikation kann hier nur ergänzen oder zur Auswertung dienen. Ein unerfahrener Traumatologe kann auch durch diese Klassifikationen nicht sicherer in seiner Indikationsstellung werden.“ Diese Aussage steht zusammenfassend für die Meinung aller Teilnehmer. So waren sie der Ansicht, dass die Entscheidung für eine Therapie, ausschließlich auf Basis der Klassifizierung einer Fraktur nicht möglich sei und sich der Therapieerfolg auch nur schwer abschätzen ließe. Denn sowohl in die Therapieentscheidung als auch in den Therapieerfolg fließen Faktoren mit ein, welche sich nicht vorhersehen lassen [6] [24].

So stellt sich die Therapieentscheidung vielseitig beeinflussbar dar. Hier spielt die Compliance des Patienten in der Heilungsphase und die Kommunikation mit den Eltern eine entscheidende Rolle. Heutzutage werden bei Kindern und Jugendlichen „90% aller Frakturen konservativ, und nur 10% operativ behandelt“ [48], was sich bei radiologisch und klinisch erkennbaren Dislokationen häufig nur schwer mit den therapeutischen Vorstellungen der Eltern vereinbaren lässt. Auch wenn Koletzko et al. 2007 [22] schreiben, dass „Fehlstellungen ad axim, ad latum und ad longitudinem ... durch Wiederaufrichtung der Epiphysen und vermehrtes Längenwachstum teilweise oder ganz ausgeglichen werden...“ können, ist noch lange nicht das Verständnis der Eltern gegeben, welche dann möglicherweise eine operative Korrektur auch innerhalb der spontanen Korrekturgrenzen erzwingen und von unerfahrenen Therapeuten häufig erhalten.

Zur Betrachtung der Kappa-Werte, verglichen nach Lokalisation, wird hier eine Übersicht über die prozentuale Verteilung der Frakturen an den unteren Extremitäten gegeben.

	Lokalisation innerhalb des Knochens	Prozent an Gesamtheit der Frakturen
FEMUR	Proximal	0,5
	Diaphysär	1
	Distal	0,5
TIBIA	Proximal	1
	Diaphysär	15
	Distal	5

Tabelle 13: Übersicht der Frakturverteilung im Wachstumsalter [20]

Tab. 13 zeigt, dass Frakturen am Unterschenkel häufiger auftreten. Das Femur als stärkster Knochen des menschlichen Körpers [31] hat ein geringeres Frakturrisiko. Des Weiteren sind die Verletzungen der Tibia und ihrer umgebenden Weichteile, vor allem im Jugendalter, durch erhöhte sportliche Aktivität [31] deutlich häufiger zu diagnostizieren, da Verletzungen beim Sport eher mit Frakturen bzw. Traumata des Unterschenkels - distal und proximal - als mit Oberschenkelfrakturen einhergehen. Somit erwarten wir durch größere Routine im Umgang mit Tibia-Frakturen bessere Übereinstimmungswerte. Die Ergebnisse unserer Studie waren jedoch gegenteilig. So erzielte die Gruppe der Ärzte einen Kappa-Wert von 0,49 für den Femur und 0,25 für die Tibia bei drei zu klassifizierenden Stellen der AO-Klassifikation. Auch die Gruppe der Studenten erreichte in der LiLa-Klassifikation Werte von 0,2 für den Femur und 0,08 für die Tibia. Somit ergeben sich doppelt so hohe Übereinstimmungswerte bei der Klassifizierung von Femur-Frakturen. Ursächlich dafür sehen wir die einfachere Klassifizierung des alleinigen Femur-Knochens gegenüber den häufig komplizierteren Verletzungsmustern der Tibia, in einzelnen Fällen kombiniert mit Verletzungen der Fibula.

Parallel zu unserer Studie wurde an der Universität Erlangen die entsprechende Untersuchung für die Klassifikationssysteme der oberen Extremitäten durchgeführt [51]. Sie konnte teilweise die von uns erzielten Ergebnisse bestätigen. So erzielte die Gruppe der Ärzte bei der Auswertung der oberen Extremität ebenfalls höhere Kappa-Werte. Anders als bei uns ergaben sich bei der Auswertung der paarigen Knochen, Radius und Ulna, höhere Werte, welche die Autoren durch das häufigere Auftreten dieser Frakturen im klinischen Alltag erklärten. Eine Überlegenheit eines der beiden Klassifikationssysteme bei der Auswertung der Frakturen der oberen Extremitäten konnte jedoch nicht festgestellt werden.

Als letzter wesentlicher Punkt sollen nun die Ergebnisse der einzelnen Stellenreduzierungen beider Klassifikationssysteme diskutiert werden. Zu diesem Zwecke wurden die

Röntgenbilder nicht erneut klassifiziert, sondern die Klassifikationscodes bei reduzierter Stellenzahl erneut ausgewertet. Der von uns verfolgte Gedanke hinter diesem Schritt war die Herausarbeitung von Schwachstellen innerhalb der Klassifikationssysteme, welche sich bei der Auswertung der reduzierten Stellen zeigen würden. Unsere größte Aufmerksamkeit schenken wir der 5. Stelle beider Klassifikationssysteme, welche sich mit dem Schweregrad bzw. der Dislokation befasst. So stellt diese Stelle einen der größten Unterschiede zwischen den beiden Systemen dar und ist schon alleine wegen ihrer subjektiven Beurteilung in einer als objektives Instrument anzusehenden Klassifikation ein diskutierbarer Unterpunkt. Durch die Stellenreduzierung konnten wir zusätzlich die dritte Stelle als zweiten, deutlichen Unterschied der zwei Klassifikationssysteme genauer betrachten und kritisch hinterfragen.

Betrachtet man die 3. Stelle isoliert, so zeigt sich lediglich bei der LiLa-Klassifikation ein signifikanter Kappa-Wert, der eine Aussage bzgl. der Übereinstimmung der Untersucher erlaubt. Dieser wurde in der Gruppe der Ärzte bei der Klassifizierung der Tibia-Frakturen erreicht. In den anderen Gruppen ergaben sich schwach negative Konfidenzintervalle, weswegen die erzielten Kappa-Werte, wie in Kapitel 2.7. erläutert, als nicht signifikant zu interpretieren sind, und daher nicht in die Auswertung einbezogen werden können. Auffallend jedoch sind die durchwegs hohen Kappa-Werte der AO-Klassifikation, die nach Landis und Koch 1977 [26] eine mäßige bis gute Übereinstimmung zeigten. Somit ergibt sich beim Vergleich der zwei Klassifikationssysteme ein signifikanter Unterschied zu Gunsten der AO-Klassifikation innerhalb der Gruppe der Ärzte.

Bei der Untersuchung der 3. Stelle für die zwei Klassifikationssysteme der oberen Extremitäten in Erlangen [51] ergab sich keinerlei signifikante Überlegenheit für AO- oder LiLa-Klassifikation.

Was sind nun die Unterschiede der zwei Klassifikationssysteme an dritter Stelle?

Die AO-Klassifikation gibt die Lokalisation der Fraktur in Epiphyse - Diaphyse - Metaphyse an. Die Metaphyse des jeweiligen Knochens wird durch ein Quadrat über der Epiphysenfuge definiert. So muss bei paarigen Knochen zur genauen Bestimmung der metaphysären Ausdehnung der zweite Knochen in das Quadrat mit eingeschlossen werden [41]. In den anfänglichen Studien zur Validierung der AO-Klassifikation wurde das einfache Quadrat über dem haupttragenden Knochen angewandt. Jedoch führte eine Ausdehnung des Quadrates über beide Epiphysenfugen zu einem deutlichen Anstieg der erreichten Kappa-Werte [4] und somit legte man diese Modalität als Standard fest. In unserer Studie jedoch wurden keine Hilfsmittel wie Folienquadrate an die Teilnehmer ausgehändigt. Selbstverständlich kann jeder Arzt im

Rahmen der täglichen Klassifizierung derartige Quadrate zur Hilfe nehmen, jedoch ist deren Anwendung im Klinikalltag unrealistisch. Zusätzlich muss zur Unterscheidung zwischen Quer- und Schrägfraktur der Frakturwinkel gemessen werden. So liegt ab einem Winkel von 30° eine Querfraktur vor [44] [41]. Die Aufnahmemodalität der Röntgenbilder kann Einfluss auf die Qualität der Klassifizierung haben. So kann „bei Anwendung der Quadratdefinition ... eine Fehlklassifikation durch nicht korrekte a.p. Aufnahmen bzw. durch Angulation des Knochens in der frontalen Ebene entstehen.“ [31] [44] Auch die Aufnahmequalität der Bilder kann eine Rolle spielen, weswegen in einigen Studien eine Aussortierung der Röntgenbilder nach Qualitätsaspekten stattfand [39] [3] [10] [19]. Aussagen bezüglich des tatsächlichen Einflusses der Bildqualität auf die Ergebnisse wurden in diesen Studien nicht erörtert. Die genannten Fehlerquellen sind im klinischen Routineablauf nicht gänzlich auszuschließen und so können „gute Ergebnisse, obgleich sehr viel versprechend bei derart idealen Voraussetzungen, ... nicht auf die alltägliche klinische Situation übertragen werden.“ [42].

Die LiLa-Klassifikation unterteilt die Frakturen in Stelle 3 nach Schaft- und Gelenkfrakturen. „Sämtliche Frakturen, die die Gelenkfläche nicht tangieren, ... sind als Schaftfrakturen zu sehen“ [50]. Diese Einteilung ist von therapeutischer Konsequenz, da a-Frakturen eine operative anatomische Rekonstruktion verlangen, wohingegen bei s-Frakturen eine eventuelle Spontankorrektur mit bedacht werden muss [31]. Auf den ersten Blick erscheint die Einteilung der LiLa-Klassifikation einfach und unproblematisch. Jedoch zeigten die erzielten Kappa-Werte das genaue Gegenteil: nur einen signifikanten Kappa-Wert in der Gruppe der Ärzte. Bei Betrachtung der möglichen Ursachen der durchwegs niedrigen Kappa-Werte fallen oben genannte Einflüsse weg, denn auch auf einem nicht korrekt in a.p. Richtung getroffenen Röntgenbild lässt sich ein Frakturspalt in der Gelenkfläche, und somit eine Gelenkbeteiligung, ausmachen. Daher lässt sich über die tatsächliche Ursache für das schlechte Abschneiden der LiLa-Klassifikation in Stelle 3 einzig die Neuheit der Einteilung nach Schaft- und Gelenkfrakturen und daher geringerer Erfahrung in eben dieser Einteilung als eventuelle Ursache anführen. Den Vergleich von Stelle 3 abschließend lässt sich folgendes feststellen: Die Klassifizierung von Frakturen mittels AO-Klassifikation ist auch von unerfahrenen Traumatologen bis inklusive Stelle 3 mit validen Ergebnissen möglich; bei der Klassifizierung nach LiLa zeigen sich sowohl bei erfahrenen, als auch unerfahrenen Traumatologen nur unzureichende Ergebnisse.

Als letzter Punkt soll die 5. Stelle beider Klassifikationssysteme diskutiert werden. Wie bereits weiter oben ausführlicher erläutert, unterscheidet die AO-Klassifikation an 5. Stelle zwischen einfachen Frakturen, Keilfrakturen mit weniger als drei Teilen und

Mehrfragmentfrakturen mit mehr als drei Teilen. Auch die LiLa-Klassifikation unterscheidet drei Gruppen, jedoch nach gänzlich anderen Gesichtspunkten. So sind undislozierte, tolerabel dislozierte und nicht tolerabel dislozierte Frakturen zu unterscheiden. Schon bei der Betrachtung der einzelnen Definitionen fällt auf, dass die LiLa-Klassifikation keine genau definierte Begrenzung der einzelnen Gruppen bietet, obwohl ein Klassifikationssystem per Definition „ein Klassensystem, mit eigenen definierten Klassengrenzen“ [55] ist. Derart äußern sich auch Slongo et al. 2007 [44]: „Kategorien innerhalb eines Klassifikationssystems sollten so eindeutig wie möglich formuliert sein, da weniger eindeutige Kategorien zu großen Grauzonen führen und so die Untersucher Nichtübereinstimmung steigern.“ Inwieweit „tolerabel“ und „nicht tolerabel“ definierte und objektive Klassengrenzen darstellen, ist jedoch fraglich, insbesondere auch deshalb, weil die Toleranzen per se Gegenstand laufender Studien sind.

Bei der Auswertung der 5. Stelle erzielte die LiLa-Klassifikation in der Gruppe der Studenten bei der Klassifizierung der Femur-Frakturen und in der Gruppe der Ärzte bei der Klassifizierung der Tibia-Frakturen eine signifikante Überlegenheit gegenüber der AO-Klassifikation. Die Kappa-Werte in der Untersuchergruppe der Studenten lagen mit 0,60 nach Landis und Koch 1977 [26] im mäßigen Bereich und die der Ärzte mit 0,41 ebenfalls im mäßigen Bereich. In ihrer Veröffentlichung zur Validierung der AO-Klassifikation erzielten Slongo et al. 2007 [43] bei der alleinigen Auswertung der 5. Stelle Kappa-Werte zwischen 0,13 und 0,82. In ihrer Studie erzielte die Gruppe mit Ärzten verschiedener Erfahrungsstufen trotz vorangegangenen Training im Klassifizieren von Frakturen nur maximale Kappa-Werte von 0,61 und einen Durchschnittswert über alle Frakturen von 0,52. Somit liegt unser maximaler Kappa-Wert der LiLa-Klassifikation von 0,60 und einem Durchschnitt der zwei Gruppen von 0,42 nach Landis und Koch 1977 [26] im selben Bereich wie die von Slongo et al. 2007 [43] erzielten Ergebnisse, obwohl in unserer Studie kein Training vorangegangen war. Die Werte der AO-Klassifikation in unserer Studie mit maximal 0,56 und einem Durchschnitt von 0,26 fallen jedoch deutlich von den Vergleichswerten Slongos ab.

Auch in der Parallelstudie in Erlangen [51] schnitt die LiLa-Klassifikation bei der Auswertung der 5. Stelle besser ab, dieser Unterschied war jedoch nicht signifikant.

Zur 5. Stelle der LiLa-Klassifikation äußern sich v. Laer et al. 2007 [50]: „Vorderhand ist es der Erfahrung des behandelnden Kliniklers überlassen, ob er die Fraktur als undisloziert, tolerabel disloziert oder intolerabel disloziert einstuft und daraus die jeweilige therapeutische Konsequenz zieht.“ Somit spricht v. Laer genau jenen unserer Kritikpunkte an, der sich bei

der Betrachtung der 5. Stelle ergibt: der Erfahrung des Untersuchers ist es überlassen, das Röntgenbild zu kategorisieren. Erfahrung aber stellt kein objektives Kriterium dar, wie weiter oben in diesem Diskussionsteil bereits besprochen wurde. Neben der Spontankorrektur als beeinflussenden Faktor für die Therapieentscheidung [47] spielt auch die aktuelle Studienlage, die Angaben in der Fachliteratur und selbst die für die traumatologische Fachabteilung typische Vorgehensweise eine entscheidende Rolle, da in jeder Klinik die Abgrenzung der einzelnen Frakturtypen [41] und die Korrekturgrenzen innerhalb einer gewissen Bandbreite individuell definiert werden. Diese beiden Punkte spiegeln sich in Stelle 5 der LiLa-Klassifikation wider und stehen nicht für eindeutig definierte Gruppengrenzen innerhalb des Klassifikationssystems. Um den Einfluss dieser Kritikpunkte zu untersuchen, wäre es interessant, eine vergleichende Studie an unterschiedlichen Kliniken unter vergleichbaren Bedingungen durchzuführen. So geschehen in zwei der Pilotstudien zur Validierung der AO-Klassifikation, bei welcher sechs Ärzte weltweit (Session 3) die Bilder klassifizierten und im Vergleich dazu fünf Ärzte aus einer Klinik (Session 4) [5]. In Session 3 wurden Median-Kappa-Werte von 0,85 und in Session 4 von 0,99 erzielt. In der weltweit durchgeführten Studie wurden jedoch zusätzlich andere Hilfsmittel wie Folienquadrate zur Bemessung der Metaphyse und ein vorangegangenes Training erlaubt. Somit wurden in Session 4 trotz weniger Hilfsmittel bessere Werte erzielt, was die Theorie des hausinternen Vorgehens bestätigt. Letztendlich lässt sich für das bessere Abschneiden der LiLa-Klassifikation trotz der subjektiven Einschätzung mittels der 5. Stelle keine Erklärung finden, jedoch lässt sich bei einer signifikanten Überlegenheit der Gruppe der erfahrenen Traumatologen bei der Klassifizierung von Tibiafrakturen folgender Schluss ziehen: die LiLa-Klassifikation sollte aufgrund ihrer therapierelevanten Aussage in Stelle 5 ausschließlich von erfahrenen Traumatologen und an die aktuelle Studienlage zu den Korrekturgrenzen angepasst angewendet werden, um die Qualität der resultierenden therapeutischen Entscheidungen sicherzustellen.

Zusammenfassend lassen sich folgende Aussagen machen:

1. Die Erfahrung der Untersucher im Gebiet der Kindertraumatologie hat Einfluss auf die Übereinstimmung beim Klassifizieren von Frakturen in der LiLa- wie auch in der AO-Klassifikation.

2. Eine Stellenreduzierung der zu klassifizierenden Charakteristika erhöht die Übereinstimmung der Untersucher.
3. Die AO-Klassifikation erzielt bei vier Unterpunkten der Untersuchung signifikant höhere Übereinstimmungswerte.
4. Subjektiv wird die LiLa-Klassifikation als einfacher und praktikabler im Vergleich zur AO-Klassifikation angegeben und somit bevorzugt.
5. Bei der Klassifizierung von Femurfrakturen werden höhere Übereinstimmungswerte erzielt.
6. Bei Untersuchung der dritten Stelle zeigt sich eine signifikante Überlegenheit der AO- gegenüber der LiLa-Klassifikation auf einer Position. Die Klassifizierung nach AO ist auch von unerfahrenen Traumatologen mit validen Ergebnissen bis inklusive der dritten Stelle möglich.
7. Bei Untersuchung der fünften Stelle zeigt sich eine signifikante Überlegenheit der LiLa- gegenüber der AO-Klassifikation auf zwei Positionen; jedoch nur, wenn klassifiziert von erfahrenen Untersuchern. Daher setzt die Therapieentscheidung basierend auf der LiLa-Klassifikation eine entsprechende Erfahrung des Untersuchers voraus.

Aufgrund signifikant höherer Kappa-Werte bei der Auswertung lässt sich eine objektive Überlegenheit der AO-Klassifikation feststellen. Bei Betrachtung der für uns wichtigen 5. Stelle zeigten sich bei der LiLa-Klassifikation höhere Kappa-Werte. Auch subjektiv wurde die LiLa-Klassifikation als einfacher und praktikabler für den klinischen Alltag gesehen. Insgesamt kann eine eindeutige Empfehlung für eines der beiden Klassifikationssysteme nicht gegeben werden. Als bestätigt sehen wir den Einfluss der Erfahrung des Untersuchenden auf die Übereinstimmung bei der Klassifizierung von Frakturen. Dies wurde bereits in vielen anderen Studien zu verschiedensten Klassifikationssystemen festgestellt [53] [25] [7] [30]. Genau diese Tatsache ist jedoch kritisch zu betrachten, da eine hohe Untersucherabhängigkeit gegen die Objektivität eines Klassifikationssystems spricht und somit den Zielgedanken einer Klassifikation verfehlt.

Literaturverzeichnis

1. **Andersen E., Jorgensen L.G., Hededam L.T. (1990)**, Evans' classification of trochanteric fractures: an assessment of the interobserver and intraobserver reliability. *Injury* 21: 377-378.
2. **Andersen G.R., Rasmussen J.B., Dahl B., Solgaard S. (1991)**, Older's classification of Colles' fractures. *Acta Orthop Scand* 62 (5): 463-464.
3. **Andersen D.J., Blair W.F., Steyer C.M., Adams B.D. (1996)**, Classification of Distal Radius Fractures: An Analysis of Interobserver Reliability and intraobserver Reproducibility. *Journal of Hand Surgery* 21A: 574-582.
4. **Audigé L., Bhandari M., Kellam J. (2004)**, How reliable are reliability studies of fracture classifications? *Acta Orthop Scand* 75 (2): 184–194.
5. **Audigé L., Hunter J., Weinberg A.M., Magidson J., Slongo T. (2004)**, Development and Evaluation Process of a Pediatric Long-Bone Fracture Classification Proposal. *Eur J Trauma* 30: 248–254.
6. **Audigé L., Bhandari M., Hanson B., Kellam J. (2005)**, A Concept for the Validation of Fracture Classifications. *Journal of Orthopedic Trauma* 19 (6): 404-409.
7. **Beaule P.E., Dorey F.J., Matta J.M. (2003)**, Letournel Classification of Acetabular Fractures: Assessment of Interobserver and Intraobserver Reliability. *Journal of Bone and Joint Surgery Am.* 85-A: 1704-1709.
8. **Belotti J.C., Tamaoki M.J.S., da Silveira Franciozi C.E., dos Santos J.B., Balbachevsky D., Chap Chap E., Albertoni W. M., Faloppa F. (2008)**, Are distal radius fracture classifications reproducible? Intra and interobserver agreement. *Sao Paulo Medical Journal* 126(3):180-185.

9. **Bernstein J. (1994)**, Fracture classification systems: do they work and are they usefull? *Journal of Bone and Joint Surgery* 76:792-793.
10. **Bernstein J., Adler L.M., Blank J.E., Dalsey R.M., Williams G.R., Iannotti J.P. (1996)**, Evaluation of the Neer System of Classification of Proximal Humeral Fractures with Computerized Tomographic Scans and Plain Radiographs. *Journal of Bone and Joint Surgery* 78 A (9):1996-1998.
11. **Bortz J., Lienert G.A. (2008)**, *Kurzgefasste Statistik für die klinische Forschung*, 3. Auflage, Springer Verlag, Heidelberg.
12. **Brennan R.L., Prediger D.J. (1981)**, Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*. S. 687-699.
13. **Brorson S., Hróbjartsson A. (2008)**, Training improves agreement among doctors using the Neer system for proximal humeral fractures in a systemic review. *Journal of Clinical Epidemiology* 61: 7-16.
14. **Cohen, J. (1960)**, A coefficient of agreement for nominal scales. 20, S. 37-46.
15. **Colton C.L., (1991)**, Telling the bones. *Journal of Joint and Bone Surgery* 73-B:362-364.
16. **Elmerson S., Andersson G.B.J., Irstam L., Zetterberg C. (1988)**, Internal Fixation of femoral neck fracture. *Acta Orthop Scand* 59 (4):372-376.
17. **Flinkkilä T., Raatikainen T., Hämäläinen M. (1998)**, AO and Frykman's classifications of Colles' fracture. *Acta Orthop Scand* 69 (1):77-81.
18. **Gehrchen P.M., Nielsen J.O., Olesen B., Andresen B.K. (1997)**, Seinsheimer's classification of subtrochanteric fractures. *Acta Orthopaedica Scandinavia* 68(6):524-526.

19. **Illaramendi A., Gonzalez della Valle A., Segal E., de Carli P., Maignon G., Galluci G. (1998)**, Evaluation of simplified Frykman and AO classifications of fractures of the distal radius. *International Orthopaedics (SICOT)* 22:111-115

20. **Johnstone D.J., Radford W.J.P., Parnell E.J. (1993)**, Interobserver Variation using the AO/ASIF classification of long bone fractures. *Injury* 24(3):163-165.

21. **King J.E., Dates B.G. (2008)**, *SPSS Algorithms for Bootstrapping and Jackknifing Generalized Measures of Agreement*, Southwest Educational Research Association, New Orleans, LA.

22. **Koletzko B., v. Harnack G.A. (2007)**, *Kinder und Jugendmedizin*, 13. Auflage, Springer Verlag, Berlin Heidelberg New York.

23. **Krauth J. (1984)**, A Modification of Kappa for Interobserver Bias. *Biometrical Journal* 26 (4):435-445.

24. **Kreder H.J., Hanel D.P., Mc Kee M., Jupiter J., Mc Gillivray G., Swiontkowski M.F. (1996)**, Consistency of AO Fracture Classification for the distal radius. *Journal of Bone and Joint Surgery Br.* 78-B:726-31.

25. **Kristiansen B., Andersen U.L.S., Olsen C.A., Varmarken J.E. (1988)**, The Neer classification of the proximal Humerus An Assessment of interobserver variation. *Skeletal Radiology* 17:420-422.

26. **Landis J.R., Koch G.G. (1977)**, An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* Jun, 33 (2):363-374.

27. **Lichtenhahn P., Fernandez D.L., Schatzker J. (1992)**, Analysis of the „user friendliness of the AO Classification of fractures. *Helvetica chirurgica acta* 58 (6):953-958.

28. **Linhart W.E., von Laer L. (2005)**, Allgemeine Gesichtspunkte zur kinderorientierten Behandlung von Verletzungen. *Der Orthopäde* 34: 1169-1186.

29. **Marsh J.L., Slongo T., Agel J., Broderick J.S., Creevey W., De Coster T., Prokuski L., Sirkin M.S., Ziran B., Henley B., Audigé L. (2007)**, Fracture and Dislocation Classification Compendium. *J Orthop Trauma* 21(Suppl.): S1-S133.

30. **Martin J.S., Marsh J.L., Bonar S. K., De Coster T.A., Found E.M., Brandser E.A. (1997)**, Assessment of the AO/ASIF Fracture Classification for the Distal Tibia. *Journal of Orthopedic Trauma* 11 (7):477-483.

31. **Marzi I. (2006)**, *Kindertraumatologie*. 1. Auflage, Steinkopff Verlag, Darmstadt.

32. **Mc Rae R. (1995)**, *Praxis der Frakturbehandlung*, 3. Auflage, Gustav Fischer Verlag, Stuttgart Jena New York.

33. **Müller-Benedict V. (2006)**, *Grundkurs Statistik in den Sozialwissenschaften*, 3. Auflage, Verlag für Sozialwissenschaften/GWV Fachverlage GmbH, Wiesbaden.

34. **Müllner M. (2005)**, *Erfolgreich wissenschaftlich arbeiten in der Klinik*, 2. Auflage, Springer Verlag, Wien.

35. **Podeszwa D.A., Stanitski C.L., Stanitski D.F., Woo R., Mendelow M.J. (2000)**, The Effect of Pediatric Orthopaedic Experience on Interobserver and Intraobserver Reliability of the Herring Lateral Pillar Classification of Perthes Disease. *Journal of Pediatric Orthopaedics* 20:562-565.

36. **Rasmussen S., Madsen P.V., Bennicke K. (1993)**, Observer variation in the Lauge-Hansen classification of ankle fractures. *Acta Orthop Scand* 64 (6):693-694.

37. **Rüedi T.P., Buckley R.E., Moran C.G. (2008)**, *AO-Prinzipien des Frakturmanagements* (1), 2. Auflage, Thieme Verlag, Stuttgart-New York.

38. **Sernbo I., Johnell O., Baath L., Nilsson J.A. (1990)**, Internal fixation of 410 cervical hip fractures. *Acta Orthop Scand* 61 (5):411-414.

39. **Sidor M.L., Zuckerman J.D., Lyon T., Koval K., Cuomo F., Schoenberg N. (1993)**, The Neer classification system for proximal humeral fractures. An assessment of interobserver reliability and intraobserver reproducibility. *The Journal of Bone and Joint Surgery* 75:1745-1750.
40. **Slongo, T. (2005)**, Editorial. *Injury - International Journal of the Care of the Injured* 36 (1) Supplement 1:1.
41. **Slongo T., Audigé L., Schlickewei W., Clavert J.-M., Hunter J. (2006)**, Development and Validation of the AO Pediatric Comprehensive Classification of Long Bone Fractures by the Pediatric Expert Group of the AO Foundation in Collaboration with AO Clinical Investigation and Documentation and the International Association for Pediatric Traumatology. *Journal of Pediatric Orthopaedics* 26 (1): 43–49.
42. **Slongo T., Audigé L., Clavert J.M., Lutz N., Frick S., Hunter J. (2007)**, The AO Comprehensive Classification of Pediatric Long-bone Fractures. *Journal of Pediatric Orthopaedics* 27:17.
43. **Slongo T., Audigé L., Lutz N., Frick S., Schmittenbecher P., Hunter J., Clavert J.-M. (2007)**, Documentation of fracture severity with the AO classification of pediatric long-bone fractures. *Acta Orthopaedica* 78(2):247-253.
44. **Slongo T., Audigé L. (2007)**, Fracture and Dislocation Classification Compendium for Children. *J Orthop Trauma* 21(Suppl.): S135-S160.
45. **Thomsen N.O.B., Overgaard S., Olsen L.H., Hansen H., Nielsen S.T. (1991)**, Observer Variation in the radiographic classification of ankle fractures. *Journal of Bone and Joint Surgery* 73-B:676-678 Editorial.
46. **Vogel L. (2012)**, Frakturen im Wachstumsalter – Vergleich der LiLa-Klassifikation nach von-Laer mit der AO-Klassifikation nach Slongo. Unveröffentlichte Dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg.

47. **von Laer L., Gruber R., Dallek M., Dietz H.-G., Kurz W., Linhart W., Marzi I., Schmittenbecher P., Slongo T., Weinberg A., Wessel L. (2000)**, Classification and Documentation of Children's Fractures. *European Journal of Trauma* 26:2-14.
48. **von Laer L. (2005)**, Frakturen im Wachstumsalter. *Trauma Berufskrankh* 7:6–10.
49. **von Laer L. (2005)**, Qualitätssicherung: Traumatologie im Wachstumsalter. *Deutsches Ärzteblatt* 102 (37):A2434.
50. **von Laer L., Kraus R. (2007)**, Die konservative Behandlung von Frakturen der langen Röhrenknochen im Wachstumsalter. *Der Unfallchirurg* 110:811-823.
51. **Walton N.P., Harish S., Roberts C., Blundell C. (2003)**, AO or Schatzker? How reliable is classification of tibial plateau fractures? *Archive of Orthopedic Trauma Surgery* 123:396-398.
52. **Weinberg A.-M., Marzi I., Günter S.M., Wessel L., Riedel J., v. Laer L. (2002)**, Die suprakondyläre Oberarmfraktur im Kindesalter- eine Effizienzstudie. *Der Unfallchirurg* 105:208-216.
53. **Wen-Jie J., Li-Yang D., Yi-Min C., Qin Z., Lei-Sheng J., Hua L. (2005)**, Reliability of classification systems for intertrochanteric fractures of the proximal femur in experience orthopaedic surgeons. S. 858-861.
54. **Wihlborg O. (1990)**, Fixation of femoral neck fractures. *Acta Orthop Scand* 61 (5): 415-418.
55. <http://www.itwissen.info/definition/lexikon/Klassifizierung-classification.html>
(14.6.2011)

Abbildungsverzeichnis

Abbildung 1: systematische Übersicht AO-Klassifikation für Frakturen im Kindesalter	13
Abbildung 2: Darstellung des Metaphysenquadrates bei paarigen Knochen.....	15
Abbildung 3: systematische Übersicht der LiLa-Klassifikation	18
Abbildung 4: Übersicht Fleiss-Kappa-Werte der Studenten Studiengruppe.....	31
Abbildung 5: Übersicht Fleiss-Kappa-Werte der Ärzte Studiengruppe.....	32
Abbildung 6: Übersicht Kappa-Werte 5. Stelle Femurfrakturen	38
Abbildung 7: Übersicht Kappa-Werte 5. Stelle Tibiafrakturen	39

Tabellenverzeichnis

Tabelle 1: Interpretation der Kappa-Werte nach Landis und Koch 1977	27
Tabelle 2: Ergebnisse des SPSS Makros (Femur Ärzte AO Stelle 3+4).....	28
Tabelle 3: Übersicht Fleiss-Kappa-Werte der Studenten Studiengruppe.....	30
Tabelle 4: Übersicht Fleiss-Kappa-Werte der Ärzte Studiengruppe.....	32
Tabelle 5: Fleiss-Kappa-Werte der Klassifikation des Femurs (Stellen 3+4+5)	33
Tabelle 6: Fleiss-Kappa-Werte der Klassifikation des Femurs (Stellen 3+4).....	34
Tabelle 7: Fleiss-Kappa-Werte der Klassifikation des Femurs (3. Stelle).....	34
Tabelle 8: Fleiss-Kappa-Werte der Klassifikation der Tibia (Stellen 3+4+5)	35
Tabelle 9: Fleiss-Kappa-Werte der Klassifikation der Tibia (Stellen 3+4)	36
Tabelle 10: Fleiss-Kappa-Werte der Klassifikation der Tibia (3.Stelle).....	37
Tabelle 11: Fleiss-Kappa-Werte der Klassifikation des Femurs (5. Stelle).....	38
Tabelle 12: Fleiss-Kappa-Werte der Klassifikation der Tibia (5. Stelle).....	39
Tabelle 13: Übersicht der Frakturverteilung im Wachstumsalter	43

Anhang

Anhang 1:

Anleitung zur Durchführung der Klassifikation

Thema unserer Doktorarbeit ist es, den Nutzen einer Klassifikation für kindliche Frakturen zu untersuchen. Um den Umgang mit der AO- bzw. der LiLa-Klassifikation besser beurteilen zu können, haben wir diese CD erstellt, auf der Sie Bilder aller 4 Extremitäten finden. Zu jeder Gruppe finden sie jeweils eine Powerpoint Präsentation, in die sie die entsprechenden Ziffern der Klassifikation eintragen können.

In jeder Präsentation finden sie einmal die Nummer des Patienten, die wir für die Auswertung der Arbeit benötigen, sowie, in einer Textbox unter den Bildern, eine Möglichkeit die Ziffern für die AO- bzw. für die LiLa-Klassifikation einzutragen. Als letzte finden Sie eine Kommentar Zeile, in die Sie sämtliche auftretende Unregelmäßigkeiten bei der Klassifikation des jeweiligen Bildes eintragen können, wie z.B. eine nicht gesehene Fraktur, Probleme mit der jeweiligen Klassifikation oder Unsicherheiten bezüglich der Einordnung der Fraktur (entscheiden Sie sich bitte dennoch für eine Verschlüsselung).

Als Hilfe haben wir eine Anleitung aus dem Buch „Kindertraumatologie“ von Ingo Marzi (Steinkopff Verlag Darmstadt 2006) beigelegt. Beginnen Sie mit der Lektüre der AO-Klassifikation, welcher die LiLa-Klassifikation sehr ähnelt.

Abschließend möchten wir Sie noch bitten, den von uns mit gesendeten Auswertungsbogen zu beantworten. Dazu genügt es, an den vorgesehenen Stellen ein Zeichen in das Dokument einzufügen und den Freitext entsprechend zu ergänzen.

Wir wären Ihnen sehr dankbar, wenn Sie uns die Ergebnisse der Klassifikation und den ausgefüllten Fragebogen so schnell, wie es Ihnen möglich ist, an die unten stehenden E-Mail Adressen zusenden könnten.

Bei weiteren Fragen stehen wir Ihnen gerne unter
benjamin.haber@gmail.com oder Lutzvogel@gmx.de zur Verfügung.

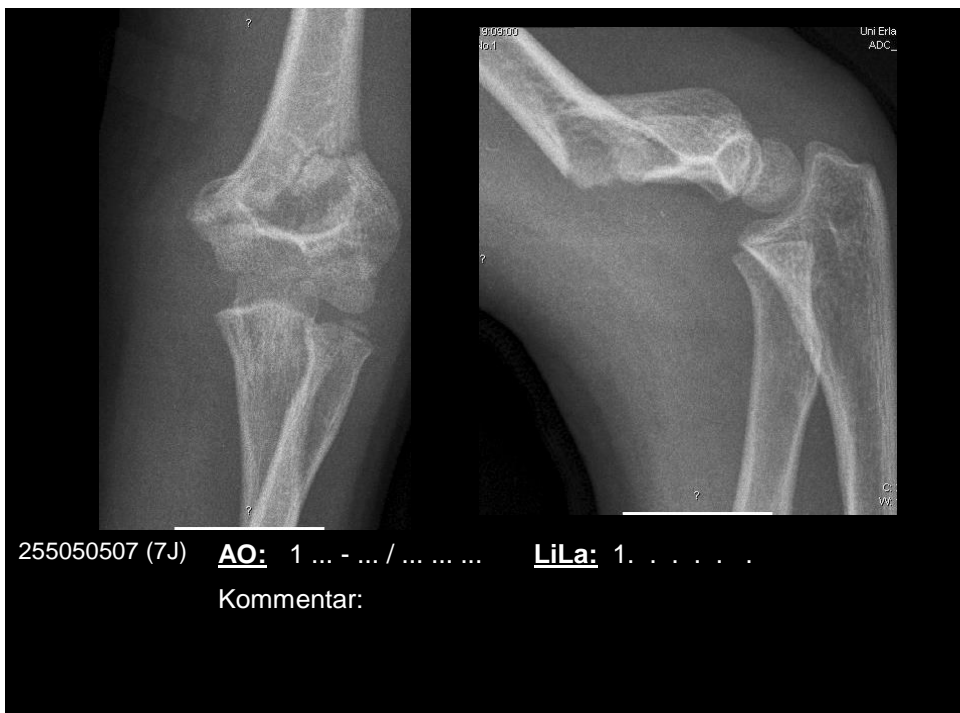
Wir bedanken uns ganz herzlich für Ihre Mitarbeit und verbleiben mit freundlichen Grüßen
Benjamin Haber und Lutz Vogel

- Welche Klassifikation wäre in Ihren Augen sinnvoller?
 AO LiLa keine
- Gibt es etwas, was Ihnen an der jeweiligen Klassifikation nicht gefällt? (Freitext)
 AO:
 LiLa:
- Was finden Sie an der Klassifikation als besonders gut/hilfreich? (Freitext)
 AO:
 LiLa:

Vielen Dank für Ihre Mitarbeit!

Anhang 3:

Powerpoint Folie einer zu klassifizierenden Fraktur



Danksagung

Zuerst möchte ich meinem Betreuer Dr. med. Michael Kertai und meinem Doktorvater Dr. med. habil. Bertram Reingruber für die Unterstützung während der Zeit der Datensammlung, Auswertung und Erstellung der Arbeit bedanken. Dr. Michael Kertai stand mir jederzeit für Fragen zur Gestaltung der Arbeit und deren Umsetzung zur Verfügung.

Der weitere Dank gebührt den Freiwilligen, die sich neben Studium bzw. Arbeit für unsere Studie als Teilnehmer zur Verfügung stellten, und nach langem Hin und Her auch Ergebnisse bei uns ablieferten, mit denen wir diese Arbeit erstellen konnten.

Für die Hilfe bei der statistischen Auswertung der Daten, die mich vor ein großes Rätsel stellte, möchte ich mich bei PD Dr. med. Jörg Marienhagen von der Universitätsklinik Regensburg bedanken, der mich völlig unkompliziert und spontan in die Tiefen der Statistik einführte.

Mein größter Dank gilt meinen Eltern, die es mir erst ermöglicht haben, dass ich den oben Genannten überhaupt für etwas danken kann. Ohne Zeitdruck konnte ich mein Studium und diese Arbeit angehen und mich auf meine wissenschaftliche Arbeit konzentrieren. Meinem Bruder danke ich für die Hilfe bei der Übersetzung der Originalzitate aus dem Englischen und Charlotte Hentze für die Korrektur nach hochdeutschem "Schtandard".