

Aus der
Berufsgenossenschaftlichen Unfallklinik
Klinik für Unfall- und Wiederherstellungschirurgie
an der Universität Tübingen

**Reliabilität und Vergleichbarkeit moderner CT-basierter
Frakturklassifikationen am Tibiakopf – eine
retrospektive radiologische Analyse**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin**

**der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen**

vorgelegt von

Dr. med. dent. Eggensperger, Felix

2024

Dekan: Professor Dr. B. Pichler
1. Berichterstatter: Professor Dr. C. C. Ihle
2. Berichterstatter: Privatdozent Dr. S. Gassenmaier

Tag der Disputation: 18.12.2023

Widmung

*Dr. Dr. Andreas Henssler und Dr. Johannes Laval für die einzigartige,
kompromisslose Unterstützung während meines zweiten Studiums*

Inhaltsverzeichnis

| | | |
|---------|--|----|
| 1 | Einleitung | 10 |
| 1.1 | Hintergrund | 10 |
| 1.2 | Intention und Validierungsprozess eines Klassifikationssystems | 13 |
| 1.3 | Die Anatomie des Schienbeinkopfes | 15 |
| 1.4 | Die Rolle der modernen bildgebenden Verfahren | 17 |
| 1.5 | Klassifikationssysteme | 18 |
| 1.5.1 | Geschichte und klassische Klassifikationssysteme | 18 |
| 1.5.1.1 | Die AO/OTA-Klassifikation [45, 46] | 19 |
| 1.5.1.2 | Die klassische Schatzker-Klassifikation [24] | 19 |
| 1.5.2 | Wichtige moderne Klassifikationssysteme | 20 |
| 1.5.2.1 | Die Luo-Klassifikation „Three-Column-Concept“ [44] | 20 |
| 1.5.2.2 | Die „10-Segment-Klassifikation“ nach Krause et al. [8] | 21 |
| 1.5.2.3 | Die Klassifikation von Kfuri und Schatzker [27] | 22 |
| 1.6 | Problemstellung | 24 |
| 1.7 | Ziel und Hypothese | 27 |
| 2 | Material und Methoden | 28 |
| 2.1 | Studiendesign | 28 |
| 2.2 | Ethik | 28 |
| 2.3 | Patientenkollektiv | 28 |
| 2.4 | Begutachtung CT präoperativ und Klassifikation der Frakturen | 29 |
| 2.5 | Daten aus der Patientenakte und OP-Berichte | 31 |
| 2.6 | Definition von Subgruppen nach Frakturgröße | 32 |
| 2.7 | Begriffsdefinitionen | 33 |
| 2.8 | Statistische Auswertung | 34 |
| 2.8.1 | Intrarater-Reliabilität | 34 |
| 2.8.2 | Sicherheit, Erlernbarkeit, Präzision, Ableitung einer OP-Strategie | 35 |
| 3 | Ergebnisse | 36 |
| 3.1 | Patientenkollektiv | 36 |
| 3.2 | Ergebnisse zu dem „Three-Column-Concept“ von Luo et al. | 36 |
| 3.2.1 | Häufigkeiten - Deskriptive Statistik | 36 |
| 3.2.2 | Percent-agreement | 38 |
| 3.2.3 | Intrarater-Reliabilität gesamt | 39 |
| 3.2.4 | Intrarater-Reliabilität Subgruppen | 39 |

Inhaltsverzeichnis

| | | |
|-------|--|----|
| 3.2.5 | Sicherheits-Ranking..... | 39 |
| 3.3 | Ergebnisse zur „10-Segment-Klassifikation“ von Krause et al..... | 40 |
| 3.3.1 | Häufigkeiten - Deskriptive Statistik..... | 40 |
| 3.3.2 | Percent-agreement..... | 41 |
| 3.3.3 | Intrarater-Reliabilität gesamt..... | 43 |
| 3.3.4 | Intrarater-Reliabilität Subgruppen..... | 43 |
| 3.3.5 | Sicherheits-Ranking..... | 43 |
| 3.4 | Ergebnisse zur Klassifikation von Kfuri und Schatzker..... | 44 |
| 3.4.1 | Häufigkeiten - Deskriptive Statistik..... | 44 |
| 3.4.2 | Percent-agreement..... | 46 |
| 3.4.3 | Intrarater-Reliabilität gesamt..... | 46 |
| 3.4.4 | Intrarater-Reliabilität Subgruppen..... | 47 |
| 3.4.5 | Häufige Fehler bzw. abweichende Bewertungen..... | 47 |
| 3.4.6 | Sicherheits-Ranking..... | 48 |
| 3.5 | Ergebnisse zur Klassifikation der AO/OTA..... | 49 |
| 3.5.1 | Häufigkeiten - Deskriptive Statistik..... | 49 |
| 3.5.2 | Percent-agreements..... | 50 |
| 3.5.3 | Intrarater-Reliabilität gesamt..... | 52 |
| 3.5.4 | Intrarater-Reliabilität Subgruppen..... | 52 |
| 3.5.5 | Häufige Fehler bzw. abweichende Bewertungen..... | 52 |
| 3.5.6 | Sicherheits-Ranking..... | 52 |
| 3.6 | Zusammenfassung der Ergebnisse..... | 53 |
| 3.6.1 | Intrarater-Reliabilitäten und percent-agreements..... | 53 |
| 3.6.2 | Sicherheits-Rankings..... | 54 |
| 3.6.3 | Statistische Werte der Klassifikationen und Vergleich nach Frakturgröße..... | 55 |
| 3.6.4 | Anzahl der unterschiedlich bewerteten Klassifikationen je Fallbewertung..... | 57 |
| 3.6.5 | Zusammenfassung der Ergebnisse aller vier Klassifikationen..... | 59 |
| 4 | Diskussion..... | 60 |
| 4.1 | Allgemein..... | 60 |
| 4.2 | Patientenkollektiv, Methoden und Limitationen..... | 61 |
| 4.3 | Reliabilität, Stärken und Schwächen der Klassifikationen..... | 65 |
| 4.3.1 | Luo et al..... | 65 |
| 4.3.2 | Krause et al..... | 68 |
| 4.3.3 | Kfuri und Schatzker..... | 71 |
| 4.3.4 | AO-Klassifikation..... | 75 |

Inhaltsverzeichnis

| | |
|----------------------------------|----|
| 4.3.5 Schlussfolgerungen..... | 79 |
| 4.4 Ausblick..... | 81 |
| 5 Zusammenfassung | 84 |
| 6 Literaturverzeichnis..... | 86 |
| 7 Erklärung zum Eigenanteil..... | 92 |
| 8 Danksagung..... | 92 |

Tabellen- und Abbildungsverzeichnis

| | |
|---|----|
| Tabelle 1: Intention und Validierungsprozess eines Klassifikationssystems nach Audigé et al. [13, 29] | 15 |
| Tabelle 2: Häufigkeiten, percent-agreements und kappa-Statistiken – Luo-Klassifikation | 37 |
| Tabelle 3: Sicherheits-Ranking – Luo-Klassifikation | 39 |
| Tabelle 4: Häufigkeiten der Segmente – Krause-Klassifikation | 40 |
| Tabelle 5: Häufigkeiten, percent-agreements und kappa-Statistiken – Krause-Klassifikation | 42 |
| Tabelle 6: Sicherheits-Ranking – Krause-Klassifikation | 44 |
| Tabelle 7: Häufigkeiten, percent-agreements und kappa-Statistiken - Kfuri/Schatzker-Klassifikation | 44 |
| Tabelle 8: Sicherheits-Ranking - Kfuri/Schatzker-Klassifikation | 49 |
| Tabelle 9: Häufigkeiten, percent-agreements und kappa-Statistiken – AO-Klassifikation | 49 |
| Tabelle 10: Sicherheits-Ranking – AO-Klassifikation | 53 |
| Tabelle 11: Intrarater-Reliabilitäten und percent-agreements der Klassifikationen im Vergleich | 53 |
| Tabelle 12: Vergleich der Klassifikationen nach Frakturgröße | 56 |
| Tabelle 13: Anzahl unterschiedlich bewerteter Klassifikationen je Fallbewertung | 57 |
| Tabelle 14: Zusammenfassung der Qualitäten aller vier Klassifikationen | 59 |
| | |
| Abbildung 1: Das "Three-Column-Concept" nach Luo et al. | 20 |
| Abbildung 2: Die "10-Segment-Klassifikation" nach Krause et al. | 22 |
| Abbildung 3: Die Klassifikation von Kfuri und Schatzker | 23 |
| Abbildung 4: Fallbeispiel Fraktur proximale Tibia mit Gelenkbeteiligung | 24 |
| Abbildung 5: Häufigkeiten der Säulen in der Klassifikation – Luo-Klassifikation | 37 |
| Abbildung 6: Häufigkeiten der Subtypen-Wechsel bei abweichender Bewertung - Kfuri/Schatzker | 48 |

Tabellen- und Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 7: Percent-agreements bei 41B- und 41C-Fraktursubtypen – AO-Klassifikation | 51 |
| Abbildung 8: Vergleich des Sicherheits-Rankings der Klassifikationen in Kat.1 "trifft sicher zu" | 54 |
| Abbildung 9: Bewertungsfehler je Fall nach Frakturgröße | 58 |
| Abbildung 10: Verschiedene Subtypen (3 von insg. 8) der C-Frakturen in der AO-Klassifikation | 77 |
| Abbildung 11: Häufigster Subtypenwechsel in der AO-Klassifikation | 77 |
| Abbildung 12: AO 41B3.2 und 41B3.3f ohne bzw. mit Einbeziehung der E. intercondylaris | 78 |

Abkürzungsverzeichnis

| | |
|--------|--|
| A. | Arteria (deutsch: Arterie) |
| AO/OTA | Arbeitsgemeinschaft für Osteosynthesefragen |
| AHRQ | Agency for Healthcare Research and Quality |
| CT | Computertomographie |
| et al. | et alii, et aliae, et alia (deutsch: und andere) |
| ICD-10 | 10. Version des International Statistical Classification of Diseases and Related Health Problems der WHO |
| k.A. | keine Angaben |
| Kat. | Kategorie |
| KI | künstliche Intelligenz |
| lat | hier: laterale Säule |
| Lig. | Ligamentum (deutsch: Band) |
| M. | Musculus (deutsch: Muskel) |
| med | hier: mediale Säule |
| Mm. | Musculi (deutsch: Muskeln) |
| MR | Magnetresonanz |
| MRT | Magnetresonanztomographie |
| N. | Nervus (deutsch: Nerv) |
| n.a. | nicht auswertbar |
| n.s. | nicht signifikant |
| OPS | Operations- und Prozedurenschlüssel |
| post | hier: posteriore Säule |
| uTCC | updated Three-Column-Concept |
| V. | Vena (deutsch: Vene) |
| VOI | Volume of Interest |
| VU | Verkehrsunfall |
| WHO | World Health Organization |
| 2D | zweidimensional |
| 3D | dreidimensional |
| κSD | kappa-Standard-Deviation (deutsch: Standardabweichung κ) |

1 Einleitung

1.1 Hintergrund

Tibiakopffrakturen sind mit weniger als 1% aller knöcherner Verletzungen eine relativ seltene Entität. Die Inzidenz liegt bei 10,3 Fällen pro 100.000 Einwohnern im Jahr. Der Häufigkeitsgipfel liegt geschlechtsunabhängig zwischen dem 40. und 60. Lebensjahr [1]. Frakturen der proximalen Tibia entstehen überwiegend infolge direkter Gewalteinwirkungen (Hochrasanztrauma). Weitere adäquate Mechanismen sind indirekte Gewalteinwirkung durch Varus- oder Valgusstress sowie eine Kombination aus axialer Kompression und Valgisierung/Varisierung bei Stürzen aus großer Höhe [2]. Ist die Gelenkfläche der proximalen Tibia beteiligt, besteht ein höheres Risiko einer bleibenden Dysfunktion des Kniegelenks. Wird nicht oder ungeeignet behandelt (beispielsweise ohne anatomische Reposition der Gelenkfläche) entwickelt sich in bis zu 45% der Fälle eine posttraumatische Gonarthrose innerhalb von 3 bis 13 Jahren [3, 4]. Ausgeprägte postoperative Impressionen der Gelenkfläche (>2,5mm) mit entsprechender Inkongruenz der Gelenkflächen führen zu signifikant schlechteren funktionellen Ergebnissen und erhöhten Bewegungseinschränkungen [5]. Patienten mit erlittener Fraktur der proximalen Tibia haben dementsprechend ein 3,5-fach erhöhtes Risiko einen endoprothetischen Kniegelenkersatz zu erhalten [6]. Begleitende Weichteiltraumata, meniskale- und ligamentäre Verletzungen sowie Nerven- und Gefäßbeteiligungen sind aufgrund der anatomischen Nachbarschaft häufig [3]. Schäden der A. und V. poplitea sowie des N. peroneus treten laut Literatur in bis zu 20% der Fälle, Bandverletzungen und Meniskuspathologien in bis zu 25% der Fälle auf [7]. Ein Verlust der Menisken und Schädigungen der Bänder erhöhen zusätzlich zur ossären Fehlstellung die Inzidenz sekundärer Gelenkdegenerationen [4]. Insgesamt stellen Frakturen der proximalen Tibia mit Gelenkbeteiligung somit eine erhebliche chirurgische Herausforderung dar.

Der Schlüssel zu einer erfolgreichen Therapieplanung komplexer Verletzungen ist eine präzise Befundaufnahme und Diagnostik. Vielfältige Unfallmechanismen und eine spezielle Knochenstruktur des Tibiakopfes führen zu einem breiten

Spektrum an Frakturmustern [8]. Die vollständige Identifikation des Verlaufs und der Charakteristik der Frakturen ist das Schlüsselement im Entscheidungsprozess zur Findung der geeigneten Behandlungsmethode. Generell unterscheidet man zwischen Plateau-, Trümmer- und Luxationsfrakturen. Die häufig begleitenden meniskalen und ligamentären Pathologien müssen für eine korrekte Versorgung ebenfalls erkannt werden. Nicht zuletzt sollten die Lebensumstände, Vor- und Begleiterkrankungen und der Anspruch und die Wünsche des Patienten bei der Therapieplanung berücksichtigt werden. Der Tibiakopf ist eine komplexe, dreidimensionale Struktur mit einer großen kranialen Gelenkfläche. Die Möglichkeiten, die realen, häufig mit Hochrasanz-Traumen assoziierten Frakturlinien und Fragmente in der Frontalebene mittels nativer Röntgendiagnostik-Technik zu diagnostizieren, sind limitiert. Frakturen im posterioren Bereich des Tibiaplateaus werden hierbei oftmals unterschätzt, obwohl sie maßgeblich die Therapie und die Prognose beeinflussen. Heutzutage hat sich die Durchführung einer Computertomographie mit zumeist multiplanarer 3D-Rekonstruktion etabliert. Sie wird ergänzend zu einer konventionellen Röntgenaufnahme durchgeführt [9-12].

Für eine effiziente Beschreibung der Lokalisation von Frakturverläufen, Fragmenten und des allgemeinen Charakters der Frakturmuster ist neben einer dreidimensionalen Bildgebung ein qualitativ hochwertiges Klassifikationssystem unabdingbar. Nur mittels weit verbreiteter, sinnvoller und anerkannter Systeme mit hoher Validität und Reliabilität können Diagnosen kommuniziert und dokumentiert sowie standardisierte Behandlungsprotokolle etabliert werden. [13] Das große Spektrum an unterschiedlichsten Frakturmustern am Tibiakopf geht einher mit einer Vielzahl an operativen Zugangsmöglichkeiten, sich stets weiter entwickelnden Behandlungsansätzen und zahlreichen Klassifikationssystemen [14-20]. Millar et al. konnten 2018 in einer systematischen Übersichtsarbeit insgesamt 38 Klassifikationen identifizieren [21]. Die ossär fokussierten Systeme von Schatzker (1979) und das etwas später veröffentlichte Klassifikationssystem der AO/OTA werden weltweit am häufigsten eingesetzt [22-24]. Ein weiteres, gängiges System von Moore et al. konzentriert sich auf Dislokationen und assoziierte Weichteilverletzungen [25]. Die Klassifikationen zeigten in der auf

Röntgenbildern basierenden Frakturanalyse eine moderate Reliabilität [26]. Die Computertomographie ermöglicht mittlerweile eine dreidimensionale Analyse des Tibiakopfes. Das diagnostische Verständnis der komplexen Frakturentität hat sich durch die Darstellungsmöglichkeiten in der sagittalen, koronaren und axialen Ebene verbessert.

Millar et. al konnten bei über 15 eingeschlossenen Reliabilitätsstudien eine Überlegenheit der CT-basierten Diagnostik im Vergleich zur konventionellen Röntgentechnik u.a. bezüglich der Inter- und Intraobserver-Reliabilität nachweisen [21]. Kfuri/Schatzker publizierten im Jahre 2018 eine dreidimensionale, CT-basierte, reformierte Version der vor 40 Jahren veröffentlichten klassischen 2D-Systematik von Schatzker et al. [27].

Die Vielfalt an vorhandenen Klassifikationssystemen ist also groß. Die Studienlage erlaubt es richtungsweisende klinische Aussagen zu treffen, beinhaltet jedoch keine abschließende und generell gültige Reliabilitätsfassung.

Ein allgemein akzeptierter und wissenschaftlich belegter Goldstandard hinsichtlich einer einheitlichen, international anerkannten Klassifikation und einem klassifikationsbasierten Behandlungsalgorithmus konnte im Bereich der Frakturen der proximalen Tibia mit Gelenkbeteiligung letztendlich noch nicht gefunden werden.

Wünschenswert wäre ein verständliches Klassifikationssystem mit hoher Validität und Reliabilität sowie detaillierten Informationen bezüglich computertomographischer zwei- und dreidimensionaler Frakturmusterbewertung. Diese Informationen sollten verschiedene Bruch-Charakteristika beinhalten. Neben anatomischen Aspekten (mediales/laterales Plateau, Säulen, Quadranten, Segmente), Fraktureigenschaften (Spalt, Impression, Depression, Luxation), Unfallmechanismen (Varusstress, Valgusstress, axiale Kompression, Rotation, Extension/Flexion) sollten auch Begleitverletzungen erfassbar sein [28]. Ein konsekutiver Algorithmus bezüglich medizinischer Frakturversorgung, chirurgischem Zugang und Art der Frakturstabilisierung sowie einem Ausblick auf das Outcome und die Prognose könnten eine entsprechende Klassifikation optimieren [21].

Die Erarbeitung einer Anwendungsempfehlung und viel mehr noch die Definition eines Goldstandards können nicht durch eine singuläre wissenschaftliche Studie dargestellt werden. Vielmehr ist die Exploration verbunden mit einem erheblichen Aufwand an wissenschaftlicher Arbeit und einer Reihe an hochwertigen Studien [13].

Welche der bestehenden, etablierten Systeme für die Erarbeitung einer Anwendungsempfehlung in Frage kommen könnten, soll Teil dieser Arbeit sein. Die Aufarbeitung und der Vergleich der Stärken und Schwächen der Klassifikationssysteme kann als Basis für die Weiterentwicklung bestehender Klassifikationen dienen. Die vorliegende Studie kann überdies zur Entwicklung eines neuen Systems beitragen, das, im besten Falle, die Stärken der bestehenden Systeme vereint und erkannte Schwächen minimiert.

Von besonderem Interesse ist hierbei die Reliabilität der Klassifikationen, der Informationsgehalt und die Erkundung möglicher Algorithmen zur operativen Versorgung.

1.2 Intention und Validierungsprozess eines Klassifikationssystems

Audigé et al. listeten Kriterien auf, die den Zweck eines guten Klassifikationssystem beschreiben. Die aufgeführten Anforderungen sollten im Zentrum eines Validierungsverfahren sowie der Evidenzentwicklung u.a. im Bereich der Tibiakopffrakturen stehen. Die vorliegende Arbeit orientierte sich generell an den von ihm aufgeführten Qualitätsanforderungen (vgl. Tabelle 1). Audigé et al. beschreiben einerseits Techniken zur Entwicklung neuer Systeme sowie andererseits Methoden zur Validierung bereits bestehender Klassifikationen [13]. Zum guten wissenschaftlichen Standard empirischer Untersuchungen gehören heutzutage klassischerweise die drei Gütekriterien Objektivität, Reliabilität sowie Validität.

Klassifikationssysteme sind valide, wenn die Messung von Qualitätsparametern wie klinisch relevanten diagnostischen Elementen, Genauigkeit und Zuverlässigkeit, objektiv sind. Die Genauigkeit misst, wie gut sich die

beschriebenen Typen in reale und in der Praxis häufig anzutreffende Fälle übersetzen lassen. Explizit sollte eine Fraktur-Kategorie demnach mit einer entsprechenden Therapie-Strategie und einer vorhersagbaren Prognose reproduzierbar in Zusammenhang stehen.

Klassifikationssysteme sind reliabel, wenn die wiederholte Anwendung der Klassifikation für einen spezifischen Frakturtyp konsistent ist. Hierbei kommt es zu einer Übereinstimmung der Klassifizierung entweder durch denselben Beobachter (Intraobserver-Reliabilität) oder durch verschiedene Beobachter (Interobserver-Reliabilität). Hohe Reliabilitäten fördern eine für unterschiedliche Behandler nachvollziehbare Dokumentation und die Möglichkeit einer adäquaten Kommunikation.

Eckpunkte der Intention eines Klassifikationssystems sind die Verbesserung des Verständnisses von Pathologien, die Vermittlung einer verständlichen und standardisierten Kommunikations- und Dokumentationsmöglichkeit, die Leistung einer Hilfestellung bei der Vergleichbarkeit von Studien sowie, im besten Fall, die Etablierung einer einheitlichen Verknüpfung von Diagnose und abgeleiteter Behandlungsstrategie bzw. Prognose. Die Klassifizierung einer Pathologie darf natürlich nicht in einer definierten und unveränderbaren Beziehung zur Therapieentscheidung stehen. Vielmehr sollte sie, ergänzend zu anderen beeinflussenden Faktoren wie Patientencharakteristika, Patientenwünschen, medizinischen Komorbiditäten sowie anfallenden Kosten, einen maßgeblichen Einfluss auf die therapeutische Strategie haben [13, 29].

Viele heutige Klassifikationssysteme scheinen diese Attribute zu besitzen. Wenige davon wurden jedoch vor ihrer Veröffentlichung wissenschaftlich korrekt und ausreichend umfangreich validiert [30-32]. Des Weiteren sollte bei der Anwendung von Klassifikationssystemen jederzeit beachtet werden, dass bei der Klassifizierung kontinuierliche Variablen in kategoriale Variablen transferiert werden. Übergänge sind hingegen häufig fließend und sind dementsprechend nicht darstellbar. Beispielsweise kann eine klare Abgrenzung zwischen den Kategorien „3 Frakturfragmente \approx Anwendung einer chirurgischen Therapie“ und „2 Frakturfragmente \approx Anwendung einer konservativen Therapie“ nicht zielführend sein. In seltenen Fällen könnte die Verwendung kontinuierlicher

Variablen wie der Frakturlänge oder der Anzahl der Frakturfragmente die bessere Wahl für die Therapieentscheidung sein.

Die meisten Klassifikationssysteme basieren auf radiologischen Auswertungen. Die Korrektheit (engl. accuracy) als wichtiges Qualitätskriterium wissenschaftlichen Handelns kann mit dieser „unzureichend perfekten“ diagnostischen Technik nur eingeschränkt eingehalten werden [33].

Tabelle 1: Intention und Validierungsprozess eines Klassifikationssystems nach Audigé et al. [13, 29]

| Intention |
|---|
| Verbesserung des Verständnisses von Pathologien |
| Verständliche und standardisierte Kommunikations- und Dokumentationsmöglichkeit |
| Optimierung der Vergleichbarkeit von Studien |
| Einheitliche Verknüpfung von Diagnose und abgeleiteter Behandlungsstrategie |
| Abschätzung der Prognose der Pathologie für den Patienten |

| Validierungskriterien im Prozess |
|---|
| 1 Klare und ausführliche Beschreibung des Klassifikationssystems |
| 2 Eindeutige und dargestellte Einschluss- und Ausschlusskriterien |
| 3 Ausgewählte Fälle sind repräsentativ für das Studienkollektiv |
| 4 Die Menge der bearbeiteten Fälle ist statistisch gerechtfertigt |
| 5 Die Gruppe der Rater ist repräsentativ bezüglich der klinischen Anwender |
| 6 Die Anzahl der Rater ist angemessen |
| 7 Die Rater klassifizieren voneinander unabhängig |
| 8 Die Rater sind verblindet bezüglich klinischer Informationen der Patienten |
| 9 Ein „Gold-Standard“ wird zur Einschätzung der Richtigkeit der Klassifizierungen definiert |
| 10 Angewandte statistische Methoden passen zum Studienziel |

1.3 Die Anatomie des Schienbeinkopfes

Die Anatomie des Tibiakopfes bedingt Frakturmuster. Anatomische Landmarken dienen zur Einteilung der Gelenkfläche in den jeweiligen Klassifikationen. Die wichtigsten anatomischen Charakteristika der proximalen Tibia sind in den folgenden Absätzen dargestellt.

Der Tibiakopf stellt die proximale Epiphyse des Schienbeins dar. Funktionell bildet er den distalen Anteil des Kniegelenks. Er besteht aus zwei Säulen (Condylus lateralis et medialis tibiae), die auf ihrer Gelenkfläche (Facies articularis superior tibiae) überknorpelt sind und die entsprechende Femurkondyle tragen. Die mediale Gelenkfläche ist mit ca. 4,5 cm² etwas größer als ihr laterales Gegenstück mit 3 cm². Die beiden Säulen unterscheiden sich bedeutend in ihrer Architektur. Die mediale Kondyle weist eine höhere Knochendichte auf und ist widerstandsfähiger gegenüber einwirkenden Kräften. Knöcherne Verletzungen des medialen Tibiaplateaus sind zumeist mit dem Auftreten von hohen Energien assoziiert.

Ungefähr die Hälfte der tibialen Tragfläche wird von den beiden Menisken (Meniscus medialis et lateralis) bedeckt. Die beiden Kondylen sind mittig vollständig durch die Eminentia intercondylaris und die Areae intercondylaris anterior bzw. posterior getrennt. Die drei Strukturen bilden zusammen mit den Facies articulares das Tibiaplateau. An den beiden Tuberkeln der Eminentia intercondylaris (Tuberculum intercondylaris anterior et posterior) setzen die jeweiligen Kreuzbänder an. Der laterale Meniskus ist an der Eminentia intercondylaris befestigt. Der mediale Meniskus ist hingegen dorsal in der Area intercondylaris und ventral an der Vorderseite der Tibia verankert.

Posterolateral artikuliert der Tibiakopf mit dem Fibulakopf (Caput fibulae). An dessen lateralem Tuberkel setzt das laterale Kollateralband des Knies an (Ligamentum collaterale laterale), das zum Epicondylus lateralis des Femurknochens zieht. Eine Verbindung zum Meniscus lateralis besteht im Gegensatz zum medialen Kollateralband (Ligamentum collaterale mediale) nicht. Das mediale Kollateralband setzt mittig (anterior-posterior), leicht unterhalb des medialen Kondylus, unmittelbar am Tibiakopf an und zieht zum Epicondylus medialis des Femurknochens.

Die bedeutendsten Nervstrukturen im Bereich des Kniegelenks sind der Nervus tibialis und der Nervus fibularis communis. Letzterer verläuft im Oberschenkel lateralseitig des Nervus tibialis unter dem Musculus biceps femoris weiter in die Fossa poplitea medial der Bizepssehne. Er umschlingt beim Austritt aus dem Kniegelenk das Collum fibulae und tritt in die Peroneusloge ein. Eine Schädigung

des Nervus fibularis communis geht einher mit einer Schwäche oder einem Ausfall der Fußheber (sog. Steppergang). Der Nervus tibialis verläuft im distalen Oberschenkel medial des Nervus fibularis communis und zieht in der Nähe der Vasa poplitea relativ oberflächlich durch die Fossa poplitea weiter in die Loge der tiefen Flexoren. Ein Ausfall des Nervus tibialis äußert sich in einer Lähmung der Supinatoren und Flexoren des Unterschenkels (sog. Hackenfuß).

Etwas tiefer als der Nervus tibialis gelegen verlaufen in der Fossa poplitea mit der Arteria und der Vena poplitea zwei große Gefäße zum Unterschenkel [34-36].

Die Vielzahl der Ansätze von Bändern und Menisken an der proximalen Tibia und die Nähe von großen Nerven und Gefäßen bedingen komplexe Frakturen mit vielfältigsten Weichteilschäden sowie begrenzte chirurgische Zugangsmöglichkeiten.

1.4 Die Rolle der modernen bildgebenden Verfahren

Mit Einführung der Computertomographie wurde ein neuer Standard in der bildgebenden Gelenkdiagnostik etabliert. Nach der Jahrtausendwende beschrieben erste Artikel die Bedeutung und das Potential eines dreidimensionalen Verständnisses von Gelenkpathologien mittels Computertomographie [37-39]. Hauptbrüchebenen können nun im Vergleich zur klassischen Röntgenprojektionstechnik besser charakterisiert werden. Die Reliabilitäten (Intra- und Inter-) der am häufigsten verwendeten Klassifikationssysteme von Schatzker et al. und AO/OTA konnten nachweislich gesteigert werden [40, 41]. Vergleiche der dreidimensionalen Computertomographie (visualisierte 3D-Rekonstruktion) mit der zweidimensionalen Technik zeigten hingegen keine höheren Reliabilitäten [10]. Die Magnetresonanztomographie erweiterte und komplettierte die Beurteilungsmöglichkeiten speziell in Bezug auf begleitende Weichgewebsverletzungen. Die ossäre Diagnostik wurde nicht positiv beeinflusst [42, 43]. Die nun mit Hilfe der Computertomographie mögliche Detektion von Frakturen in der Frontalebene bietet entscheidende Vorteile hinsichtlich der Beurteilung von

Verletzungen der proximalen Tibia. Den lange unterschätzten, aber zunehmend diagnostisch, therapeutisch und prognostisch wichtiger werdenden Frakturen der posterioren Plateauanteile wurde durch die CT-Bildgebung überhaupt erst Aufmerksamkeit zuteil [9, 11, 12, 21].

Nicht zuletzt führte die Einführung der dreidimensionalen Bildgebung zur Entwicklung neuer bzw. zur Überarbeitung bestehender Klassifikationssysteme. Die in dieser Arbeit enthaltenen Klassifikationen von Luo et al. (2010) [44] bzw. Krause et al. (2016) [8] wurden speziell für die CT entwickelt. Schatzkers traditionell etablierte Klassifikation aus den 1970er Jahren wurde von Kfuri/Schatzker 2018 überarbeitet, erweitert und an die moderne Bildgebung angepasst publiziert [24, 27]. Die vorliegende Studie beinhaltet des Weiteren die auf der klassischen Röntgentechnik basierende AO-/OTA-Systematik aus dem Jahre 1981, die 2018 für computertomographische Befunde aktualisiert wurde. [45, 46].

Die vier aufgeführten Klassifikationssysteme (Luo et al., Krause et al., Kfuri/Schatzker, AO) werden im nächsten Kapitel ausführlicher beschrieben.

1.5 Klassifikationssysteme

1.5.1 Geschichte und klassische Klassifikationssysteme

Marchant et al. beschrieben 1939 als erste Autoren Tibiakopffrakturen und teilten diese in drei Gruppen ein [47]. Duparc et al. veröffentlichten 1960 eine dem späteren Schatzker-System ähnliche Differenzierung [48]. Im deutschsprachigen Raum etablierte sich für die klassische, röntgenmorphologische Frakturanalyse die Klassifikation von Tscherne und Lobenhoffer (1993) [49]. Schatzkers Klassifikation (1979) [24], die AO-Klassifikation (1981) [45], angelehnt an Tscherne und initial beruhend auf der Einteilung von Müller und Allgöwer [50] sowie die Moore-Klassifikation (1981) [25] für Luxationsfrakturen, setzten sich im angloamerikanischen Raum durch. Zentrale Schwachstelle der klassischen Systeme ist die insuffiziente, für die Therapie (inklusive chirurgischer Reposition) und Prognose aber entscheidende, exakte Beschreibung der Frakturmorphologie, insbesondere in der Frontalebene.

1.5.1.1 Die AO/OTA-Klassifikation [45, 46]

Die Klassifikation der Arbeitsgemeinschaft für Osteosynthese (AO) orientiert sich an der röntgenologischen Frakturbeschaffenheit. Extraartikuläre Brüche und knöcherne Ausrisse der Bänder werden der Gruppe der A-Frakturen (extraartikulär) zugeordnet. B-Frakturen (monokondylär) sind definitionsgemäß unikondyläre und intraartikuläre Spalt-, Impressions- und Spalt- Impressionsfrakturen. Der Gruppe C (bikondylär) sind bikondyläre Mehrfragment- und Trümmerfrakturen zugewiesen. Die mit den Ziffern 1.1 bis 3.3 gekennzeichneten Subgruppen der A-, B- und C-Frakturen beschreiben den Grad bzw. die Morphologie der Fragmentierung (Frakturcharakteristik). Insgesamt werden 28 Frakturtypen differenziert. 2018 wurde das klassische AO/OTA-System aktualisiert und beinhaltet nun eine 4-Quadranteneinteilung des Tibiaplateaus mit deren Hilfe auch computertomographische Befunde verschlüsselt werden können. Die komplexe Verschlüsselungs-Systematik der Aktualisierung wird aufgrund des enormen technischen und zeitlichen Aufwandes im klinischen Alltag nicht routinemäßig eingesetzt.

1.5.1.2 Die klassische Schatzker-Klassifikation [24]

Schatzker et al. differenzierten in ihrer 1979 veröffentlichten Klassifikation sechs Frakturtypen:

- Typ I: Laterale Spaltfrakturen
- Typ II: Laterale Spalt- Impressionsfrakturen
- Typ III: Impressionsfrakturen des lateralen Tibiaplateaus
- Typ IV: Mediale Plateaufrakturen
- Typ V: Bikondyläre Frakturen
- Typ VI: Frakturen mit vollständiger Unterbrechung der Epi- und Metaphyse

Wie in der AO/OTA-Klassifikation steht die röntgenmorphologische Frakturanalyse (2D) mit den korrelierenden Schwachpunkten (vgl. Abschnitt 1.4 und 1.5.1) im Zentrum des Systems.

1.5.2 Wichtige moderne Klassifikationssysteme

Die, neben der AO-Klassifikation, drei bedeutsamsten Systeme, die spezifisch für die CT-basierte Frakturanalyse entwickelt wurden, sind in zeitlich geordneter Reihenfolge das „Three-Column-Concept“ von Luo et al. [44], die „10-Segment-Klassifikation“ nach Krause et al. [8] und die von Schatzker und Kfuri et al. 2018 modifizierte Variante der klassischen Schatzker-Publikation von 1979 (vgl. 1.5.1.2) [27].

1.5.2.1 Die Luo-Klassifikation „Three-Column-Concept“ [44]

Luo et al. orientierten sich 2010 als erste Autoren an der chirurgischen Strategie und unterteilten das Tibiaplateau in drei „Säulen“ (lateral, medial und posterior). Bei 29 komplexen Frakturen (Schatzker V und VI) wurden die betroffenen Säulen anschließend mit guten klinischen Resultaten reponiert. 2019 erweiterten sie ihr evaluiertes Konzept um den Unfallmechanismus (updated Three-Column-Concept, uTCC) und evaluierten prospektiv eine hohe Intra- und Interrater-Reliabilität an weiteren 90 Fällen [51] (vgl. Abbildung 1).

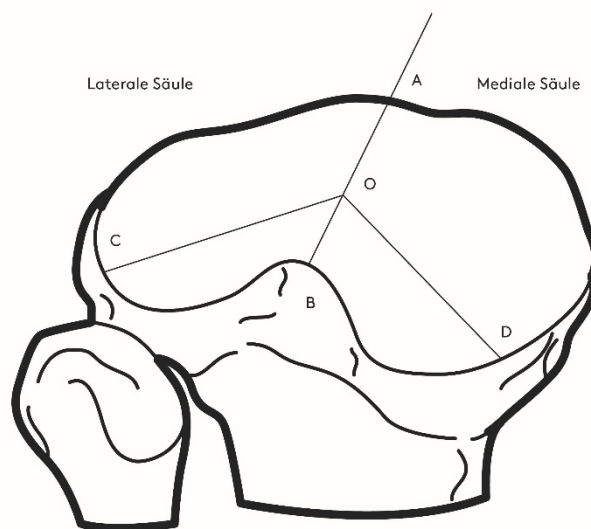


Abbildung 1: Das "Three-Column-Concept" nach Luo et al.

Die Einteilung des Plateaus erfolgt sternförmig mit Hilfe der anatomischen Trennlinien („connecting lines“) OA, OC und OD (Abb. 1). Die Landmarke O ist definiert als „Zentrum des Knies“, dem Mittelpunkt der Eminentia sowohl in medio-lateraler als auch in antero-posteriorer Richtung. Die anterior gelegene Tuberositas tibiae repräsentiert den Punkt A. Der Punkt D liegt am postero-medialen Grat des Plateaus spiegelbildlich zum Punkt C, dem vordersten Anteil des Fibulaköpfchens. Der Punkt B ist definiert durch den Schnittpunkt der antero-posterioren Mittellinie des Plateaus mit dem posterioren Tibiarand. Er spielt eine untergeordnete Rolle, da er als Verlängerung der Trennlinie der medialen und lateralen Säule lediglich die posteriore Säule in zwei Anteile teilt. Eine Teilung der posterioren Säule ist jedoch in der Klassifizierung der Frakturen mittels Luo-Systematik nicht vorgesehen.

Ist im Bereich einer Säule auf dem Tibiaplateau der knöcherne Randwall des Schienbeinkopfes durchtrennt, wird diese Säule nach Luo et al. als frakturiert gewertet. Sind mehrere Säulen betroffen werden sie analog der Fraktur zugeordnet. Eine reine Impressionsfraktur des Plateaus ohne Kontinuitätsdurchtrennung des knöchernen Randwalls wird als sog. „zero-column fracture“ gewertet.

1.5.2.2 Die „10-Segment-Klassifikation“ nach Krause et al. [8]

Krause et. al segmentierten in ihrer operativ-methodisch orientierten Veröffentlichung 2016 das Tibiaplateau in 10 Teilbereiche. Das sogenannte Mapping sollte eine gezieltere Wahl des Zugangs und der chirurgischen Strategie ermöglichen (vgl. Abbildung 2).

Die Klassifikations-Regeln nach Krause et al. im Einzelnen:

- Die Oberfläche des Tibiaplateaus wird in axialer Richtung in einen anterioren und posterioren Bereich getrennt. Die Hälften werden in zwei laterale, ein zentrales Areal und zwei mediale Areale aufgeteilt.
- In der Frontalebene erfolgt die Frakturanalyse im Bereich der obersten 3 cm des Tibiakopfes, gemessen ab dem höchsten Punkt der Artikulationsfläche.

- Sind mehr als 50% des VOI (Volume of Interest) eines Segmentes abgesenkt gilt das Segment als frakturzugehörig.
- Ist ein Segment weder frakturiert noch abgesenkt, jedoch Teil eines größeren, dislozierten Bruchstückes mit frakturierter Kortikalis unter der Artikulationsfläche, ist das Teilstück zu berücksichtigen.
- Wird ein nicht zertrümmertes Areal durch eine Frakturlinie zerteilt, darf das Segment nur gewertet werden, wenn der kleinste Bruchanteil mindestens 25% des Segment-VOI einschließt [8].

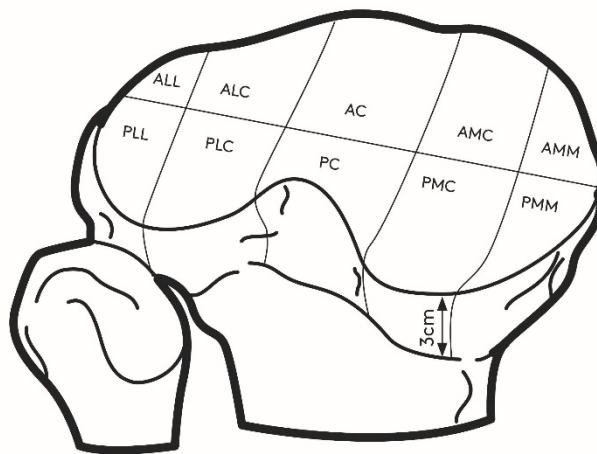


Abbildung 2: Die "10-Segment-Klassifikation" nach Krause et al.

Krause et al. machten mit einer Arbeit zu ihrer 10-Segment-Klassifikation erneut auf die Wichtigkeit des posterioren Zugangs aufmerksam. Sie demonstrierten, dass AO-C-Frakturen zu 100% und AO-B-Frakturen zu 78,3% mit posterioren Segmenten vergesellschaftet sind [8].

1.5.2.3 Die Klassifikation von Kfuri und Schatzker [27]

Die klassische und weit verbreitete Schatzker-Klassifikation wurde 2018 von Kfuri/Schatzker modifiziert und den modernen Möglichkeiten der dreidimensionalen Bildgebung angepasst. Kernpunkte der Revision beziehen

sich auf die Dokumentation von Frakturverläufen u.a. in der mit der konventionellen Projektionstechnik nicht zugänglichen Frontalebene. Den immer mehr in den chirurgischen Fokus rückenden wichtigen posterioren Anteilen wird mehr Aufmerksamkeit zuteil [52]. Die Berücksichtigung anatomischer, biomechanischer und chirurgischer Gegebenheiten soll eine gezielte Beurteilung von Instabilität, Frakturanteilen und eine entsprechend angepasste Operationsstrategie über optimierte Zugangswege möglich machen.

Im Zentrum des überarbeiteten Systems stehen weiterhin die sechs Hauptfrakturtypen aus 1979. Die proximale Tibia wird ergänzend auf Höhe der Eminentia intercondylaris in eine mediale und laterale Säule getrennt. Ein neu eingeführter, sogenannter „virtueller Äquator“, trennt einen größeren anterioren von einem kleineren posterioren Teil des Tibiaplateaus. Die Äquatorlinie verläuft vom Hinterrand des Ansatzes des Innenbandes zum lateral gelegenen anterioren Teil des Fibulakopf-Tuberkels (Ansatzpunkt des Außenbandes an der Fibula) (vgl. Abbildung 3).

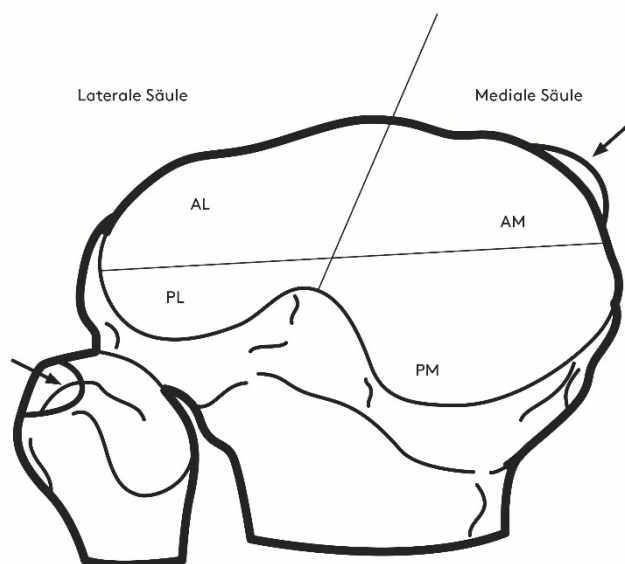


Abbildung 3: Die Klassifikation von Kfuri und Schatzker

Die vier definierten Quadranten AL=anterolateral, AM=anteromedial, PL=posterolateral und PM=posteromedial ergänzen die sechs Hauptfrakturtypen

je nachdem in welchen Quadranten der kortikale Rand des Tibiaplateaus durchtrennt ist. Die unikondylären Frakturtypen I bis IV erhalten die Zusätze A, P oder AP. Bei den bikondylären Frakturen V und VI erfolgt die Differenzierung, je nach Frakturverlauf, mit Hilfe der vier Quadranten AL, AM, PL und PM [27].

1.6 Problemstellung

Anhand des folgenden klinischen Beispiels soll die klinische und wissenschaftliche Relevanz des Themas dargestellt werden, um das Ziel der Studie und die Hypothese der vorliegenden Arbeit zu erörtern.

Es handelt sich um eine 49-jährige Patientin, die mit ihrem Fahrrad verunfallte und sich infolgedessen eine proximale, intraartikuläre Tibiafraktur links zuzog. Nach Anamnese und klinischer Untersuchung erfolgte die Röntgen- und CT-Bildgebung (vgl. Abbildung 4).



Abbildung 4: Fallbeispiel Fraktur proximale Tibia mit Gelenkbeteiligung

Eine erste Betrachtung der Bilder lässt den Schluss zu, dass das mediale Plateau, einschließlich der leicht mehrfragmentär frakturierten Area intercondylaris, keilförmig am Tibiakopf frakturiert ist. Die Hauptfrakturlinie auf dem Plateau läuft von postero-latero-zentral nach antero-zentral. Im Verlauf im Tibiakopf nach distal zieht sie schräg nach medial.

Die Analyse der Fraktur bildgebung unter Anwendung der entsprechenden Klassifikationen ergibt:

- Luo et al. „Three-Column-Concept“

In allen drei Säulen durchbrechen Frakturlinien auf dem Tibiaplateau den ossären Randwall. Es besteht nach Luo et al. eine **3-Säulen-Fraktur „posterior-medial-lateral“**.

- Krause et al. „10-Segment-Klassifikation“

Die Hauptfrakturlinie verläuft auf dem Tibiaplateau von postero-latero-zentral nach antero-zentral. Die Segmente AC und PLC sind demnach definitionsgemäß der Fraktur zugehörig. Das Segment PC ist in sich frakturiert/leicht zertrümmert und ist somit ebenfalls in die Klassifizierung einzubeziehen. Die weiter medial liegenden Segmente sind zwar weder in sich frakturiert, noch abgesenkt, werden jedoch laut Definition als Teil eines größeren, dislozierten Bruchstückes mit frakturierter Kortikalis unter der Artikulationsfläche als zugehörig gewertet. Somit lautet die Krause-Klassifikation dieser Fraktur: **AC, AMC, AMM, PLC, PC, PMC, PMM**

- Kfuri/Schatzker-Klassifikation

Die Frakturanalyse anhand der Kriterien von Kfuri/Schatzker ergibt eine „Typ IV - mediale Plateaufraktur“. Die Hauptfraktur verläuft klar von anterior nach posterior (A-P). Es liegt somit eine **Typ IV-A-P-Fraktur** vor.

- AO-Klassifikation:

Die Fraktur ist unikondylär und somit den B-Frakturen zuzuordnen. Die Intercondylar-Region ist in die „split-depression“-Fraktur einbezogen. Es liegt folglich eine B3.3-Unterklasse vor. Die mediale Lokalisation der Fraktur wird durch den Buchstaben h gekennzeichnet. Dementsprechend ist die Fraktur mit dem Typ **AO 41B3.3h** zu klassifizieren.

Wird anhand der oben beschriebenen Klassifikationen über die Fraktur und mögliche Therapie- bzw. Operationsstrategien kommuniziert können diverse Problematiken auftreten.

Die Luo-Klassifikation beschreibt die Beteiligung aller drei Säulen auf dem Tibiaplateau. Eine Information über den Hauptfrakturverlauf im Tibiakopf wird

nicht vermittelt. Ebenso wenig kann der große, mediale Frakturkeil charakterisiert werden, welcher für eine chirurgische Strategieplanung von Bedeutung ist. Prinzipiell beschreiben Luos posterior-medial-laterale Frakturen auch Trümmerfrakturen erheblichen Ausmaßes, die eventuell mehrzeitige Verfahren und multiple Zugänge nötig machen. Die genaue Beschreibung des Ausmaßes bzw. die detaillierte Beschreibung der Frakturen anhand der lediglich vier Kategorien gelingt insuffizient.

Krauses Klassifikation beschreibt die beteiligten Segmente auf den ersten kranio-kaudalen 3 cm des proximalen Tibiakopfes ab der Gelenkfläche. Hierbei wird nicht zwischen Hauptfrakturverlauf, zertrümmerten Segmenten oder Knochenteilen differenziert. Eine für die Wahl des chirurgischen Procederes nötige Ableitung der Frakturmorphologie, des Frakturverlaufs und vorhandener Bruchstücke gelingt nicht.

Kfuri/Schatzker würden im vorliegenden Fall die Fraktur als mediale Plateaufraktur beschreiben. Eine Charakterisierung der Hauptfrakturlinie gelingt anhand des Zusatzes A-P relativ präzise, auch wenn die medio-laterale Lage z.B. in Bezug auf die Area intercondylaris nicht bis ins Detail wiedergegeben werden kann. Das Hauptdefizit der Kfuri/Schatzker-Klassifikation liegt vielmehr in der fehlenden weiteren Differenzierung der medialen Plateau-Frakturen. Der Typ IV-A-P könnte definitionsgemäß ebenfalls eine reine Impressionsfraktur sein, die möglicherweise lediglich kleinere Bereiche des anteromedialen und posteromedialen Plateaus einbezieht. Die Wahl der therapeutischen Strategie wäre eine wesentlich andere als im vorliegenden Fall, bei gleicher Klassifizierung.

Die Klassifizierung nach AO gelingt anhand visueller Beschreibungen relativ zügig und einfach. Zu beachten sind hierbei die Definitionen der uni- und bikondylären Frakturen und deren Grenzen untereinander. Eine B-Fraktur ist eine unikondyläre Fraktur. Die Typen B3.3f und B3.3h beinhalten zusätzlich die Intercondylar-Region. Liegt die Fraktur, wie im o.g. Fall, am Übergang zum kontralateralen Kondylus, stellt sich die Frage, ab wann die Fraktur als bikondylärer C-Typ gewertet wird, obwohl „nur“ eine Kondyle tatsächlich abfrakturiert ist (in o.g. Fall die mediale Kondyle). Der Bereich der C-Frakturen

beinhaltet letztlich eine Auswahl an großen und größten Frakturtypen bzw. Untertypen, die wiederum chirurgisch andere Strategien des Zugangs und der Osteosynthese nötig machen können.

Der präsentierte klinische Fall verdeutlicht beispielhaft die Schwierigkeiten der Anwendbarkeit und Vergleichbarkeit bestehender dreidimensionaler Frakturklassifikationen. Sowohl in den Klassifikationen selbst als auch in der parallelen Verwendung verschiedener Klassifikationen kann es zu Missverständnissen kommen, die die Reliabilität der Klassifikationssysteme einschränken können.

1.7 Ziel und Hypothese

Primäres Ziel dieser retrospektiven, radiologischen Studie ist es, die Intrarater-Reliabilität von 3D-basierten Frakturklassifikationen am Tibiakopf anhand eines großen Studienkollektivs zu untersuchen. Zusätzlich soll die klinische Anwendbarkeit, die Präzision, die subjektive Erlernbarkeit sowie die jeweiligen Stärken und Schwächen der jeweiligen Klassifikationen beschrieben werden.

Folgende Hypothese wurde für die Untersuchung der Intrarater-Reliabilität gestellt:

Die vier Klassifikationen unterscheiden sich nicht hinsichtlich der Intrarater-Reliabilität bzw. der prozentualen Übereinstimmung und zeigen diesbezüglich auch bei unterschiedlichen Frakturausmaßen große Übereinstimmungen.

2 Material und Methoden

2.1 Studiendesign

Es handelt sich um eine retrospektive radiologische Studie (Evidenzlevel IV nach AHRQ) [53].

2.2 Ethik

Ein positives Votum der Ethik-Kommission der medizinischen Fakultät der Universität Tübingen lag zum Studienbeginn vor. Das Votum wurde für den Ethikantrag mit der Nummer 179/2019BO2 ausgestellt.

2.3 Patientenkollektiv

Erfasst wurde in dieser Studie eine konsekutive Stichprobe von Patientinnen und Patienten der Berufsgenossenschaftlichen Unfallklinik Tübingen mit Tibiakopffraktur im Zeitraum zwischen Januar 2015 und Dezember 2019. Gesammelt wurden die Daten mit Hilfe der Dokumentationsdatenbank der Klinik und spezifischer Filterung der entsprechenden ICD-10-Diagnoseschlüssel (S82.11 und S82.18) bzw. Operations- und Prozedureschlüssel (OPS 5-793X.k).

Definierte Einschlusskriterien waren Volljährigkeit, das Vorliegen einer Fraktur der proximalen Tibia mit intraartikulärer Beteiligung sowie das Vorhandensein eines konventionellen Röntgenbildes und einer Computertomographie des betroffenen Knies präoperativ. Die Beurteilbarkeit der CTs, beispielsweise hinsichtlich Artefakte, wurde anhand der initialen Befundung des Radiologen überprüft. Die Ausschlusskriterien umfassten Anzeichen einer pathologischen Fraktur, Amputationsverletzungen der unteren Extremität, periprothetische Frakturen, Frakturen bei Paraplegien, isolierte knöcherne Ausrisse von Bandstrukturen und in der Vorgeschichte bereits diagnostizierte ipsilaterale Kniefrakturen. Insgesamt konnten 535 Patienten mit o.g. ICD-10 und OPS-Codes ermittelt werden, die sich im o.g. Zeitraum mit Tibiakopffraktur/en in unserer Klinik

vorstellten. Aus dieser Gruppe erlitten 97 Patienten lediglich einen knöchernen Bandauriss, 20 Patienten stellten sich mit periprothetischer Fraktur vor und 17 waren zum Zeitpunkt der Aufnahme noch nicht volljährig. Bei 11 Patienten fehlte die bildgebende Diagnostik (Computertomographie), 5 waren nebenbefundlich paraplegisch, bei 3 Patienten wurde der Unterschenkel posttraumatisch amputiert und ein Patient verstarb noch während des Klinikaufenthaltes.

Ein Kollektiv von 381 Probanden mit insgesamt 386 Fällen erfüllte letztendlich die Kriterien und konnte zur Datenauswertung gewonnen werden.

Die Anzahl der Fälle liegt über der berechneten Stichproben-Kalkulation von $n=298$ (Signifikanz-Level $\alpha=0,05$; power $(1-\beta)=0,95$; expected $\kappa=0,7$; minimal accepted $\kappa=0,4$; proportion of outcome=0,9). Die Berechnung wurde mit einer speziell für Reliabilitäts-Studien entwickelten Software (Arifin, Vancouver) vorgenommen [54].

2.4 Begutachtung CT präoperativ und Klassifikation der Frakturen

Die Begutachtung, Interpretation und Klassifikation der jeweiligen Computertomographien wurde von einem unabhängigen Doktoranden mit abgeschlossenem Zahnmedizinstudium, Weiterbildung zum Oralchirurgen und Fachkunde in der digitalen Volumetomographie durchgeführt. Er sammelte im Vorfeld über mehr als 9 Jahre Erfahrung im Bereich der Befundung dreidimensionaler Bildgebung. Über Expertenvorkenntnisse im Gebiet der Unfall- und Wiederherstellungschirurgie im Bereich des Kniegelenkes verfügte er nicht. Nach sorgfältigem Literaturstudium der verwendeten Klassifikationen wurde er von zwei Unfallchirurgen intensiv geschult und beraten. Zusätzlich wurden vor den Klassifikationsdurchläufen 15 „Probefälle“ außerhalb des Studienkollektivs im Team erarbeitet und diskutiert.

Die definitiven Auswertungs- und Klassifikations-Prozesse wurden mit zufällig sortierten, mittels Nummern 1 bis 386 anonymisierten Patientenfällen im Abstand von 8 Wochen durchgeführt. Der zweite Durchlauf erfolgte ohne ein zwischenzeitliches Feedback hinsichtlich der Klassifikationsgüte. Der Zugang zu den Daten der ersten Sequenz war während des 8-Wochen-Intervalls und

während des zweiten Durchlaufs nicht möglich. Die Begutachtung der CTs, die radiologische Interpretation und die folgende Zuordnung in den jeweiligen Klassifikationssystemen erfolgte ohne zeitliche Limitationen.

Für die Bewertung und Klassifikation der Frakturen wurden die vier dreidimensionalen Klassifikationssysteme von Luo et al. [44], Krause et al. [8], Kfuri/Schatzker [27] sowie die AO/OTA-Klassifikation [46] verwendet. Die Kategorisierung der Frakturen erfolgte streng nach den in den jeweiligen Publikationen vorgestellten Regeln. Eine detaillierte Beschreibung der Klassifikations-Systeme ist den vorausgehenden Kapiteln 1.5.1.1 und 1.5.2 zu entnehmen.

Beim Einsatz der Luo-Klassifikation wurden die jeweiligen betroffenen Säulen des Tibiaplateaus „posterior“, „medial“ oder „lateral“, einzeln oder in Kombination, notiert. Reine Impressionsfrakturen wurden, wie vom Autor vorgeschlagen, mit dem Begriff „null“ bezeichnet. Bezüglich der Krause-Klassifikation wurden entsprechend den spezifischen Regeln die beteiligten Fraktur-Segmente (ALL, ALC, AC, AMC, AMM, PLL, PLC, PC, PMC, PMM) festgehalten. Die Schatzker/Kfuri-Klassifikation schreibt zusätzlich zur Typisierung der Frakturen nach den Ziffern I bis VI eine Quadranten-Einteilung des Tibiaplateaus vor. Mittels imaginären Äquators wurden, wie durch die Autoren beschrieben, die Typen I und II in einen anterioren Bereich A und einen posterioren Bereich P geteilt. Bei den Frakturtypen III bis VI existieren die vier Quadranten AL, AM, PL und PM. Die Teilbereiche, in denen der Frakturverlauf die Kortikalis durchbricht, wurden, wie von den Autoren vorgegeben, gewertet.

Im AO/OTA-System wurden zunächst alle Tibiakopffrakturen dem Typ 41 zugeordnet. Die Typen A, B und C beschreiben die Frakturcharakteristik grob. A-Frakturen haben keinen Bezug zur Gelenkfläche und waren somit nicht Inhalt dieser Studie. Die Frakturgruppen 1 bis 3 und die Untergruppen X.1 bis X.3 beschreiben den Frakturverlauf und die beteiligten Tibiakopfanteile präziser. Auf die Verwendung der seit 2018 vorgeschlagenen zusätzlichen Quadranten-Einteilung mittels Buchstaben (t, e, s, d, u, etc.) wurde aufgrund der ohnehin schon komplexen Systematik und der fehlenden klinischen Anwendung verzichtet. Die Ausnahme bildeten hierbei die Frakturtypen 41B1.3 h/f und

41B3.3 h/f, bei denen das h oder das f obligatorisch zur Kenntnis von lateraler Fraktur f oder medialer Fraktur h verwendet werden muss.

Eine vierstufige, ordinale Ratingskala zur subjektiven Einschätzung der Schwierigkeit einer fallspezifischen Entscheidungsfindung wurde für jede Klassifikation zusätzlich erhoben. Die vier Stufen gliedern sich in „trifft sicher zu“ – „trifft eher zu“ – „trifft eher nicht zu“ – „trifft sicher nicht zu“.

Der Doktorand hatte zur Sicherstellung der Genauigkeit während des gesamten Erhebungsprozesses Zugang zu beschreibender Literatur und veranschaulichenden Abbildungen der Klassifikationen.

2.5 Daten aus der Patientenakte und OP-Berichte

Die Erhebung der spezifischen Patientendaten wurde nach den CT-Begutachtungen bzw. der Kategorisierung der Frakturen durchgeführt, um eine Verblindung des Raters bezüglich klinischer und therapeutischer Daten sicherzustellen. Eine präzise Fall- und Patientencharakterisierung könnte als Arbeitsgrundlage für weitere Studien zur Erarbeitung einer Anwendungsempfehlung anhand dieses Kollektivs dienen. Das Ableiten eines konsekutiven Algorithmus bezüglich medizinischer Frakturversorgung, chirurgischem Zugang und Art der Frakturstabilisierung sowie ein Ausblick auf das Outcome und die Prognose könnten hierbei von besonderem Interesse sein. Anhand der digitalen Patientenakten wurden die Parameter Alter, Geschlecht, betroffener Tibiakopf, Trauma-Operations-Intervall, der Trauma-Mechanismus bzw. das Trauma-Ereignis (gruppiert in „nicht bekannt“, „VU PKW“, „VU Motorrad“, „VU Fahrrad“, „VU Fußgänger“, „Aufprall“, „Distorsion“, „Überroll“, „axiale Distorsion“ und „Luxation“) sowie der Unfallhergang (gruppiert in „nicht bekannt“, „Arbeit“, „häusliches Umfeld“, „unterwegs Freizeit“, „Skifahren“, „Pferdesport“, „Fußball“, „sonstiger Sport“) erhoben.

Aus vorhandenen Arztbriefen und OP-Berichten wurden wichtige klinische und therapeutische Informationen entnommen. Hierbei waren der Weichteilschaden

(Klassifikation nach Tscherne und Oestern), begleitende Knochenpathologien außerhalb des Tibiakopfes (u.a. Tibiaschaftfraktur, Fibulafraktur, Fraktur der Patella), begleitende Weichteilpathologien (Bänder-, Menisken- oder Sehnenbeteiligungen des Kniegelenkes), begleitende neurologische Schäden (Parese des N. peroneus und/oder des N. fibularis communis) und begleitende vaskuläre Schäden (Kompartmentsyndrome des Unterschenkels, tiefe Beinvenenthrombosen) von besonderem Interesse.

Zu den ermittelten therapeutischen Fall-Merkmalen gehörten die präoperative Versorgung (u.a. Genucast, Fixateur externe), die Art der Behandlung (konservativ oder chirurgisch), der Operateur, die Art der operativen Lagerung des Patienten, die Art und die Anzahl der chirurgischen Zugänge, die Art, Anzahl und Anbringung der Osteosynthesematerialien, die Verwendung einer diagnostischen und/oder therapeutischen Arthroskopie sowie die optionale Implantation von autogener oder allogener Spongiosa.

2.6 Definition von Subgruppen nach Frakturgröße

Die detaillierte Untersuchung der Klassifikationssysteme bezüglich Stärken und Schwächen erfordert eine differenzierte Bewertung der Klassifikationen in den jeweiligen Bereichen verschiedener Frakturgrößen. Für den Vergleich der Klassifikationen und deren erreichter statistischer Werte in Abhängigkeit von den jeweiligen Frakturgrößen (Kapitel 3.6.5 und Kapitel 3.6.6) wurden 4 Frakturgruppen definiert und der Größe nach sortiert:

- kleine Frakturen (kFs):
Luo „null“ + 1 Säule | Krause 1-3 Segmente | Schatzker I + III | AO 41B1 + 41B2
- mittlere Frakturen (mFs):
Luo 2 Säulen | Krause 4-6 Segmente | Schatzker II + IV | AO 41B3
- große Frakturen (gFs):
Luo 3 Säulen | Krause 7-9 Segmente | Schatzker V | AO 41C1.1 bis 41C3.2

- Trümmerfrakturen (TFs):
Luo 3 Säulen | Krause 10 Segmente | Schatzker VI | AO 41C3.3

2.7 Begriffsdefinitionen

In dieser Studie wurde mit medizinischen Klassifikationen gearbeitet. Für ein leichteres Verständnis verwendeter Begriffe werden diese in der Folge erläutert.

- Klassifikation:
struktureller Verbund von Kategorien mit dem Ziel der Abgrenzung und Ordnung der von ihr erfassten Objekte (z.B. AO-Klassifikation).
- Kategorie:
Gruppierung von Objekten, die ein entsprechendes Einschlusskriterium/ Attribut erfüllen (z.B. 41B-Frakturen der AO-Klassifikation).
- Klassifizierung:
Prozess der Einordnung eines Objekts (hier Fraktur) in eine Klassifikation.
- Fraktur-Typ:
Oberkategorie zur groben Zuordnung von Frakturen (übergeordnete Kriterien, z.B. 41B-Frakturen der AO-Klassifikation).
- Fraktur-Subtyp:
Additive Unterkategorie zur genaueren Differenzierung der zu einem Fraktur-Typ zugeordneten Frakturen (mehrere, untergeordnete Kriterien, z.B. 41B2-Frakturen bzw. 41B2.1-Frakturen der AO-Klassifikation).
- Klassifikations-Code:
Ergebnis einer Klassifizierung mit klassifikationsspezifisch verschlüsseltem Informationsgehalt, welches einem Objekt bzw. einer Fraktur zugewiesen wird.

- Differenzierungsmöglichkeiten:
Summe aller Kategorien einer Klassifikation.

2.8 Statistische Auswertung

2.8.1 Intrarater-Reliabilität

Alle Daten wurden mit Hilfe des Tabellenkalkulationsprogramms „Microsoft Excel“ digital dokumentiert und strukturiert

Für die Auswertung mittels deskriptiver und induktiver Statistik sowie zur Erstellung komplexerer Tabellen und Schaubildern diente die Software „IBM SPSS Statistics 28“.

Die statistische Analyse erfolgte zunächst mittels Cohens kappa (κ). Cohens kappa ist ein statistisches Maß u.a. für die Intrarater-Reliabilität. Kappa kann hierbei Werte von -1 bis +1 annehmen. Bei optimaler Übereinstimmung der Auswertungen bzw. Urteile nimmt Cohens kappa den Wert +1 an. Sind die festgestellten Übereinstimmungen mathematisch dem Ausmaß des Zufalls zuzuordnen nimmt kappa den Wert 0 an. Ist die Übereinstimmung kleiner als der Zufall, nimmt Kappa einen negativen Wert an [55].

Landis und Koch (1977) skalierten diese kappa-Werte ordinal. Die folgende Aufzählung soll die unterschiedlichen Niveaus veranschaulichen, die zur Einschätzung der kappa-Werte herangezogen werden können [56]:

Kappa (κ) nach Landis und Koch:

- <0 : schlechte Übereinstimmung („poor agreement“)
- 0,00 - 0,20: etwas Übereinstimmung („slight“)
- 0,21 – 0,40: ausreichende Übereinstimmung („fair“)
- 0,41 – 0,60: mittelmäßige Übereinstimmung („moderate“)
- 0,61 – 0,80: beachtliche Übereinstimmung („substantial“)
- 0,81 – 1,00: (fast) vollkommene Übereinstimmung („(almost) perfect“)

Unter Berücksichtigung der genannten Grundsätze und mit Vermerk auf eine vorsichtige Interpretation wird in dieser Studie neben den kappa-Werten auch eine kurze, beschreibende Interpretation der kappa-Werte anhand der Landis-Koch-Guideline vorgenommen.

Für die Analyse und den Vergleich kleinerer Datenmengen, wie in diesem Fall der Subgruppen, wurde die einfachere Auswertung nach percent-agreements (PAs) gewählt. Mit dieser Methode kann herausgefunden werden, ob sich die Fehler in den Bewertungs-Durchläufen 1 \approx „test“ und 2 \approx „re-test“ gleichmäßig auf die einzelnen Kategorien verteilen oder ob es Teil-Bereiche eines Bewertungs-Systems gibt, in denen besonders viele Fehler gemacht werden. Ein entsprechend schwaches percent-agreement kann eine Überprüfung rechtfertigen, warum es zur Häufung dieser Fehler kommt.

Sowohl der kappa-Koeffizient als auch das percent-agreement haben Stärken und Schwächen. McHugh empfiehlt die Berechnung beider Größen für eine valide Statistik [57].

2.8.2 Sicherheit, Erlernbarkeit, Präzision, Ableitung einer OP-Strategie

Die subjektive Sicherheit, die Erlernbarkeit, die Präzision sowie die Möglichkeit der Ableitung einer OP-Strategie wurden mit Hilfe einer 5-stufigen Ordinalskala (--, -, o, +, ++) bewertet. Die Bewertung der subjektiven Sicherheit stützte sich auf den individuellen Eindruck der Komplexität des Bewertungsvorgang sowie der Bewertungszeit des Raters bei einer Fraktur in den jeweiligen Klassifikationen. Zur Bewertung der Erlernbarkeit wurden Unterschiede der subjektiven Sicherheit zwischen erstem und zweitem Durchlauf herangezogen. Die Präzision der Frakturbeschreibung sowie die Option der Ableitung einer chirurgischen Strategie erfolgten subjektiv.

3 Ergebnisse

3.1 Patientenkollektiv

Die 381 Patientinnen und Patienten waren bei der fallspezifischen Erstvorstellung in unserer Klinik im Schnitt 51,9 Jahre alt (Minimum 18, Maximum 93 | SD 16,1). Zum Studienkollektiv gehörten 202 (53,0%) weibliche und 179 (47,0%) männliche Patienten. Bei 164 (43,0%) Patientinnen und Patienten war das rechte Knie betroffen, 212 (55,6%) erlitten eine Fraktur des linken Tibiakopfes und bei 5 Patientinnen und Patienten waren beide Knie mitbeteiligt.

Ein Aufprall, beispielsweise im Rahmen eines Sturzes, war in 136 (35,7%) Fällen der führende Unfallmechanismus. Dahinter folgten in absteigender Reihenfolge die Distorsion (69 Fälle | 18,1%), der VU mit dem Fahrrad (56 Fälle | 14,7%), die axiale Kompression (37 Fälle | 9,7%), der VU mit dem Motorrad (24 Fälle | 6,3%), der VU mit dem PKW (17 Fälle | 4,5%), das Überroll-Trauma (16 Fälle | 4,2%), der VU als Fußgänger (13 Fälle | 3,4%) und die Luxation (12 Fälle | 3,1%). Lediglich bei einem Fall (0,3%) konnte der Unfallmechanismus nicht ermittelt werden.

Bei einer relativen Mehrheit von 133 (34,9%) Patientinnen und Patienten resultierte die Fraktur des Tibiakopfes durch eine Verletzung in der Freizeit (u.a. motorisierte VUs, Fahrrad-Unfälle, Spazieren gehen, Wandern, Spazieren gehen mit dem Hund). Weitere häufige Unfallhergänge bezogen sich auf das häusliche Umfeld (76 Fälle | 19,9%), das Skifahren (63 Fälle | 16,5%) und den Arbeitsplatz (48 Fälle | 12,6%). Seltener erlitten Patientinnen und Patienten Tibiakopffrakturen beim Fußball (9 Fälle | 2,4%), beim Pferdesport (5 Fälle | 1,3%) oder sonstigem Sport wie Windsurfen, Kitesurfen und Eislaufen (kumulativ 11 Fälle | 2,9%). 36-mal (9,4%) konnte der Unfallhergang nicht rekonstruiert werden.

3.2 Ergebnisse zu dem „Three-Column-Concept“ von Luo et al.

3.2.1 Häufigkeiten - Deskriptive Statistik

Die Mehrheit an Frakturen in der Klassifikation nach Luo et al. lag bei beiden Durchläufen in den Säulen-Kombinationen posterior-lateral (test 145 | 37,6%; re-

test 147 | 38,1%; total 292 | 37,8%) und posterior-medial-lateral (test 150 | 38,9%; re-test 146 | 37,8%; total 296 | 38,3%). Sie repräsentierten zusammen letztendlich nahezu 8 von 10 Fällen. Die Häufigkeiten in der Klassifikation nach Luo et al. sind beispielhaft für den ersten Durchgang (test) in der Abbildung 5 dargestellt.

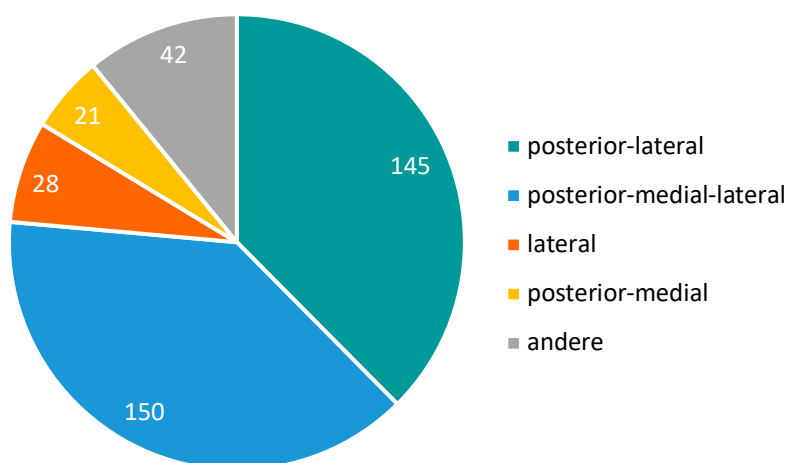


Abbildung 5: Häufigkeiten der Säulen in der Klassifikation – Luo-Klassifikation

Kumulativ war die posteriore Säule in 328 Fällen | 85,0% (test) bzw. 335 Fällen | 86,8% (re-test) beteiligt und machte mit 663 Diagnosen (total 85,9% aller Bewertungen test + re-test) den Großteil der bewerteten Frakturen aus. Äußerst selten waren hingegen Frakturen, die beide seitlichen Säulen, jedoch nicht die posteriore Säule betreffen (test 3 | 0,8%; re-test 3 | 0,8%; total 6 | 0,08%).

In der Tabelle 2 sind die relevanten Daten zu Häufigkeiten, percent-agreements und kappa-Statistiken aufgelistet.

Tabelle 2: Häufigkeiten, percent-agreements und kappa-Statistiken – Luo-Klassifikation

| | n test | n re-test | %-agree | kappa κSD p |
|---------|------------|------------|---------|------------------------|
| null | 15 3,9% | 14 3,6% | 93,3 | - |
| 1 Säule | 52 13,5% | 59 13,2% | 67,2 | 0,566 0,074 <0,001 |

| | | | | |
|---------------------|--------------------|--------------------|-------------|----------------------------------|
| posterior | 12 3,1% | 17 4,4% | 70,6 | |
| medial | 12 3,1% | 9 2,3% | 61,5 | |
| lateral | 28 7,3% | 25 6,5% | 67,7 | |
| 2 Säulen | 169 43,8% | 175 45,3% | 79,2 | 0,484 0,056 <0,001 |
| post-med | 21 5,4% | 25 6,5% | 53,3 | |
| post-lat | 145 37,6% | 147 38,1% | 84,8 | |
| med-lat | 3 0,8% | 3 0,8% | 50,0 | |
| 3 Säulen | 150 38,9% | 146 37,8% | 86,2 | -0,052 0,012 0,386 |
| post-med-lat | 150 38,9% | 146 37,8% | 86,2 | |
| Total | 386 100% | 386 100% | 89,1 | 0,848 0,022 <0,001 |

3.2.2 Percent-agreement

In 344 Fällen | 89,1% stimmten die Klassifizierungen in beiden Durchläufen überein. In 42 Fällen | 10,9% lag eine Differenz in einer Säule vor. Entweder wurde eine Säule im re-test zusätzlich gewertet (post=11-mal; med=6-mal; lat=4-mal; insg.=21-mal) oder eine Säule wurde im re-test weggelassen (post=3-mal; med=8-mal; lat=10-mal; insg.=21-mal).

Eine Zunahme der Fehlerhäufigkeit mit zunehmender Frakturgröße konnte bezüglich des percent-agreements nicht festgestellt werden (1 Säule=67,2%, 2 Säulen=79,2%, 3 Säulen=86,2%). Die kleine Gruppe um die Impressionsfrakturen „null“ (ohne Fraktur im Randwallbereich) (test=15 und re-test=14) verzeichnete einen geringen Übereinstimmungsfehler (6,7%). Bei Frakturen mit Beteiligung einer Säule (test=52 und re-test=59) wurde in 20 Fällen (32,8%) keine übereinstimmende Entscheidung getroffen. Waren 2 Säulen involviert (test=169 und re-test=175) wichen die Ergebnisse in 40 Fällen (20,8%) voneinander ab. Große Frakturen mit Beteiligung aller 3 Säulen (test=150 und re-test=146) zeigten die besten Übereinstimmungswerte mit lediglich 22 Abweichungen (13,8%).

3.2.3 Intrarater-Reliabilität gesamt

Die Intrarater-Reliabilität in der Klassifikation nach Luo et al. lag bei einem κ -Wert von $0,848 \pm 0,022$ ($p < 0,001$) und somit im Bereich „(almost) perfect“ nach Landis und Koch.

3.2.4 Intrarater-Reliabilität Subgruppen

Die Intrarater-Reliabilität der Subgruppe „1 Säule“ (Beteiligung von einer Säule) in der Klassifikation nach Luo et al. lag bei einem κ -Wert von $0,566 \pm 0,074$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch.

Die Intrarater-Reliabilität der Subgruppe „2 Säulen“ (Beteiligung von zwei Säulen) in der Klassifikation nach Luo et al. lag bei einem κ -Wert von $0,484 \pm 0,056$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch.

Bei Beteiligung aller 3 Säulen war die Intrarater-Reliabilität im Bereich „poor agreement“ nach Landis und Koch ($\kappa = -0,052 \pm 0,012$; $p = 0,386$). Demnach ist das Ergebnis eine Übereinstimmung, die noch kleiner ist als eine zufällige Übereinstimmung.

3.2.5 Sicherheits-Ranking

Die subjektive Sicherheit bezüglich der Entscheidungsfindung war in der Klassifikation nach Luo et al. überwiegend gut („trifft eher zu“: test 87 | 22,5% und re-test 103 | 26,7%) bis sehr gut („trifft sicher zu“: test 296 | 76,7% und re-test 281 | 72,8%) (vgl. Tabelle 3).

Tabelle 3: Sicherheits-Ranking – Luo-Klassifikation

| | n test | n re-test | +/- |
|------------------------|---------------|------------------|------------|
| trifft sicher zu | 296 76,6% | 281 72,8% | - 3,8% |
| trifft eher zu | 87 22,5% | 103 26,7% | + 4,2% |
| trifft eher nicht zu | 3 0,8% | 2 0,5% | - 0,3% |
| trifft sicher nicht zu | 0 0,0% | 0 0,0% | ± 0 |

3.3 Ergebnisse zur „10-Segment-Klassifikation“ von Krause et al.

3.3.1 Häufigkeiten - Deskriptive Statistik

Die Tabelle 4 zeigt die Häufigkeiten der Segmente in der Krause-Klassifikation.

Tabelle 4: Häufigkeiten der Segmente – Krause-Klassifikation

| | test | re-test |
|---------------|-----------------------|-----------------------|
| | <i>n Häufigkeit</i> | <i>n Häufigkeit</i> |
| ALL | 273 70,7% | 276 71,5% |
| ALC | 255 66,1% | 253 65,5% |
| AC | 201 52,1% | 202 52,3% |
| AMC | 159 41,2% | 163 42,2% |
| AMM | 155 40,2% | 156 40,4% |
| PLL | 307 79,5% | 309 80,1% |
| PLC | 272 70,5% | 267 69,2% |
| PC | 216 56,0% | 218 56,5% |
| PMC | 174 45,1% | 175 45,3% |
| PMM | 168 43,5% | 172 44,6% |
| | | |
| nur anterior | 28 7,3% | 34 8,8% |
| nur posterior | 30 7,8% | 30 7,8% |
| ant + post | 328 85,0% | 322 83,4% |
| | | |
| nur lateral | 128 33,2% | 122 31,6% |
| nur medial | 8 2,1% | 10 2,6% |
| med + lat | 250 64,8% | 254 65,8% |
| | | |
| nur eminentia | 0 0% | 0 0% |

Nach der „10-Segment-Klassifikation“ von Krause et al. sind die lateralen Segmente (ALL, ALC, PLL, PLC) am häufigsten involviert.

Sie waren in 93% der Fälle Teil der Fraktur (test 361 | 93,5%; re-test 360 | 93,3%; total 721 | 93,4%). In 32% traten sie isoliert auf (test 128 | 33,2%; re-test 122 | 31,6%; total 250 | 32,4%). Die Beteiligung medialer Segmente war mit 68% der Frakturen ebenfalls häufig (test 258 | 66,8%; re-test 264 | 68,4%; total 522 |

67,6%). Isoliert traten mediale Segmente jedoch mit 2,4% äußerst selten auf (test 8 | 2,1%, re-test 10 | 2,6%; total 18 | 2,3%).

Anteriore (test 356 | 92,2%; re-test 356 | 92,2%; total 712 | 92,3%) und posteriore (test 358 | 92,7%; re-test 352 | 91,2%; total 712 | 92,0%) Segmente waren mit 92% etwa gleich häufig betroffen. Frakturen, die lediglich anteriore (test 28 | 7,3%; re-test 34 | 8,8%; total 62 | 8,0%) oder posteriore Segmente betreffen, waren selten (jeweils 8%).

Betrachtet man die Verteilung der Segment-Anzahl der jeweiligen Frakturen, fällt eine leichte Tendenz zu kleineren Frakturen mit 2 (test 54 | 14,0%; re-test 50 | 13,0%; total 104 | 13,3%), 3 (test 49 | 12,7%; re-test 52 | 13,5%; total 101 | 13,1%) und 4 (test 58 | 15,0%; re-test 40 | 10,4%; total 98 | 12,7%) Segmenten sowie zusätzlich eine Häufigkeitszunahme im Bereich der Trümmerfrakturen mit 10 Segmenten (test 60 | 15,5%; re-test 65 | 16,8%; total 125 | 16,2%) ins Auge.

Die häufigsten 3 Segmentkombinationen waren in absteigender Reihenfolge die Kombination mit allen 10 Segmenten (test 60 | 15,5%; re-test 65 | 16,8%; total 125 | 16,2%), die kleinere, leicht anterior betonte laterale Fraktur mit den Segmenten ALL, ALC und PLC (test 27 | 7,0%; re-test 29 | 7,5%; total 56 | 7,3%), die laterale Fraktur ohne anteriore oder posteriore Betonung ALL, ALC, PLL, PLC (test 33 | 8,5%; re-test 23 | 6,0%; total 56 | 7,3%) und die laterale Fraktur mit anteriorer und posteriorer Beteiligung der Eminentia ALL, ALC, AC, PLL, PLC, PC (test 19 | 4,9%; re-test 22 | 5,7%; total 41 | 5,3%).

3.3.2 Percent-agreement

Die Tabelle 5 zeigt percent-agreements und kappa-Statistiken in der Klassifikation nach Krause et al..

In 249 Fällen | 64,5% stimmten die Klassifizierungen in beiden Durchläufen überein. 137 | 35,5% Fälle wurden somit im re-test abweichend beurteilt. In 106 Fällen (27,6% aller bewerteten Fälle (n=386) bzw. 77,4% der Fälle, die im re-test abweichend beurteilt wurden (n=137)) unterschieden sich die Fraktur-Bewertungen in einem Segment. In 27 Fällen (7,0% aller bewerteten Fälle (n=386) bzw. 19,7% der Fälle, die im re-test abweichend beurteilt wurden (n=137)) unterschieden sich die Fraktur-Bewertungen in zwei Segmenten, in 3

Fällen (0,8% aller bewerteten Fälle (n=386) bzw. 2,2% der Fälle, die im re-test abweichend beurteilt wurden (n=137)) lag der Rater bei drei Segmenten falsch und in einem Fall (0,3% aller bewerteten Fälle (n=386) bzw. 0,7% der Fälle, die im re-test abweichend beurteilt wurden (n=137)) waren 4 Segmente verschieden.

Tabelle 5: Häufigkeiten, percent-agreements und kappa-Statistiken – Krause-Klassifikation

| | test | re-test | %-agree | kappa κSD p |
|---------------------------|-------------------|-------------------|----------------|------------------------------|
| <i>Kleine Frakturen</i> | 113 29,3% | 119 30,8% | 63,1 | 0,585 0,048 <0,001 |
| 1 Segment | 10 2,6% | 17 4,4% | 55,6 | 0,438 0,121 <0,001 |
| 2 Segmente | 54 14,0% | 50 13,0% | 51,4 | 0,424 0,064 <0,001 |
| 3 Segmente | 49 12,7% | 52 13,5% | 47,8 | 0,354 0,067 <0,001 |
| <i>Mittlere Frakturen</i> | 118 30,6% | 111 28,8% | 47,2 | 0,411 0,046 <0,001 |
| 4 Segmente | 58 15,0% | 40 10,4% | 42,6 | 0,300 0,063 <0,001 |
| 5 Segmente | 29 7,5% | 32 8,3% | 23,4 | 0,148 0,060 <0,001 |
| 6 Segmente | 31 8,0% | 39 10,1% | 40,8 | 0,246 0,071 <0,001 |
| <i>Große Frakturen</i> | 95 24,6% | 91 23,6% | 51,0 | 0,482 0,051 <0,001 |
| 7 Segmente | 34 8,8% | 32 8,3% | 38,3 | 0,324 0,068 <0,001 |
| 8 Segmente | 35 9,0% | 39 10,1% | 37,3 | 0,296 0,064 <0,001 |
| 9 Segmente | 26 6,7 % | 20 5,2% | 50,0 | 0,437 0,095 <0,001 |
| <i>Trümmerfrakturen</i> | | | | |
| 10 Segmente | 60 15,5% | 65 16,8% | 92,3 | 0,000 0,000 - |
| Total | 386 100% | 386 100% | 64,5 | 0,621 0,025 0,000 |

In den kleineren Frakturtypen mit 1 bis 3 Segmenten konnte von insgesamt 113 Fällen eine relativ geringe Anzahl von 77 | 68,1% im zweiten Durchlauf gleich klassifiziert werden. Im mittelgroßen Frakturbereich mit 4 bis 6 Segmenten gab es bei 118 Fällen lediglich in 60 (50,8%) Frakturklassifizierungen eine Übereinkunft. Fast die Hälfte der Fälle unterschied sich folglich in mindestens einem Segment. Im Bereich der großen Frakturen (7 bis 9 Segmente) mit 95 Fällen wurden im re-test 52 (54,7%) Fälle gleich beurteilt. Schließlich konnte bei 60 Trümmerfrakturen mit Beteiligung von 10 Segmenten eine komplette

Übereinkunft erzielt werden. Ein Anstieg der Fehlerquote mit wachsender Frakturgröße und Komplexität konnte somit bezüglich des percent-agreements für die Krause-Klassifikation nicht festgestellt werden.

3.3.3 Intrarater-Reliabilität gesamt

Die Intrarater-Reliabilität total in der Klassifikation nach Krause et al. lag bei einem κ -Wert von $0,621 \pm 0,025$ ($p < 0,0001$) und somit im Bereich „substantial“ nach Landis und Koch.

3.3.4 Intrarater-Reliabilität Subgruppen

Die Intrarater-Reliabilität der Subgruppe „kleine Frakturen“ (Beteiligung von 1 bis 3 Segmenten) in der Klassifikation nach Krause et al. lag bei einem κ -Wert von $0,585 \pm 0,048$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch. Die Intrarater-Reliabilität der Subgruppe „mittlere Frakturen“ mit 4 bis 6 beteiligten Segmenten in der Klassifikation nach Krause et al. lag bei einem κ -Wert von $0,411 \pm 0,046$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch. Bei Beteiligung von 7 bis 9 Segmenten (Subgruppe „große Frakturen“) war die Intrarater-Reliabilität im Bereich „moderate“ nach Landis und Koch ($\kappa = 0,482 \pm 0,051$; $p < 0,001$).

Für „Trümmerfrakturen“ mit Beteiligung aller 10 Segmente konnten keine Intrarater-Reliabilitäten errechnet werden.

3.3.5 Sicherheits-Ranking

Die subjektive Sicherheit bezüglich der Entscheidungsfindung war in der Klassifikation nach Krause et al. überwiegend gut („trifft eher zu“: test 186 | 48,2% und re-test 122 | 31,6%) bis sehr gut („trifft sicher zu“: test 192 | 49,7% und re-test 261 | 67,6%). Im re-test konnte erheblich an subjektiver Sicherheit gewonnen werden (Kat. „trifft eher zu“ mit wesentlichen Verlusten von 16,6% zugunsten der Kat. „trifft sicher zu“ mit +17,8%) (vgl. Tabelle 6).

Tabelle 6: Sicherheits-Ranking – Krause-Klassifikation

| | n test | n re-test | +/- |
|------------------------|---------------|------------------|---------------|
| trifft sicher zu | 192 49,9% | 261 67,6% | +17,8% |
| trifft eher zu | 186 48,2% | 122 31,6% | - 16,6% |
| trifft eher nicht zu | 8 2,1% | 3 0,8% | - 1,3% |
| trifft sicher nicht zu | 0 0,0% | 0 0,0% | ± 0 |

3.4 Ergebnisse zur Klassifikation von Kfuri und Schatzker

3.4.1 Häufigkeiten - Deskriptive Statistik

In der Tabelle 7 sind die alle wichtigen Daten zu Häufigkeiten, percent-agreements und kappa-Statistiken aufgelistet.

Tabelle 7: Häufigkeiten, percent-agreements und kappa-Statistiken - Kfuri/Schatzker-Klassifikation

| | test | re-test | %-agree | kappa κSD p |
|----------------|-----------------------|-----------------------|----------------|------------------------|
| | <i>n Häufigkeit</i> | <i>n Häufigkeit</i> | | |
| <i>Typ I</i> | 22 5,7% | 18 4,7% | 73,9 | 0,570 0,133 <0,001 |
| A | 5 1,3% | 4 1,0% | | |
| P | 1 0,3% | 0 0,0% | | |
| AP | 16 4,1% | 14 3,6% | | |
| <i>Typ II</i> | 106 27,5% | 111 28,8% | 82,8 | 0,513 0,090 <0,001 |
| A | 13 3,4% | 10 2,6% | | |
| P | 5 1,3% | 4 1,0% | | |
| AP | 88 22,8% | 97 25,1% | | |
| <i>Typ III</i> | 61 15,8% | 57 14,8% | 70,8 | 0,566 0,076 <0,001 |
| A | 14 3,6% | 15 3,9% | | |
| P | 17 4,4% | 17 4,4% | | |
| AP | 30 7,8% | 25 6,5% | | |
| <i>Typ IV</i> | 26 6,7% | 26 6,7% | 92,6 | 0,856 0,095 <0,001 |
| A | 8 2,1% | 8 2,1% | | |
| P | 0 0,0% | 0 0,0% | | |
| AP | 18 4,7% | 18 4,7% | | |

| | | | | |
|-----------------|------------------|------------------|-------------|------------------------------|
| <i>Typ V</i> | 114 29,5% | 115 29,8% | 71,1 | 0,666 0,046 <0,001 |
| AL-AM | 1 0,3% | 2 0,5% | | |
| AL-PL | 2 0,5% | 2 0,5% | | |
| AL-PM | 1 0,3% | 0 0,0% | | |
| AM-PL | 19 4,9% | 29 7,5% | | |
| AM-PM | 1 0,3% | 2 0,5% | | |
| PL-PM | 9 2,3% | 9 2,3% | | |
| AL-AM-PL | 18 4,7% | 15 3,9% | | |
| AL-AM-PM | 3 0,8% | 4 1,0% | | |
| AL-PL-PM | 18 4,7% | 25 6,5% | | |
| AM-PL-PM | 19 4,9 % | 12 3,1% | | |
| AL-AM-PL-PM | 23 6,0% | 15 3,9% | | |
| | | | | |
| <i>Typ VI</i> | 57 14,8% | 59 15,3% | 71,7 | 0,653 0,070 <0,001 |
| AL-AM | 0 0,0% | 0 0,0% | | |
| AL-PL | 7 1,8% | 9 2,3% | | |
| AL-PM | 1 0,3% | 1 0,3% | | |
| AM-PL | 5 1,3% | 8 2,1% | | |
| AM-PM | 1 0,3% | 1 0,3% | | |
| PL-PM | 0 0,0% | 0 0,0% | | |
| AL-AM-PL | 16 4,1% | 17 4,4% | | |
| AL-AM-PM | 1 0,3% | 1 0,3% | | |
| AL-PL-PM | 6 1,6% | 6 1,6% | | |
| AM-PL-PM | 1 0,3% | 3 0,8% | | |
| AL-AM-PL-PM | 19 4,9% | 13 3,4% | | |
| | | | | |
| Total | 100% | 100% | 81,1 | 0,793 0,022 0,000 |

Eine relative Mehrheit der Frakturtypen in der Klassifikation nach Kfuri/Schatzker lag bei den Frakturtypen II und V. Beim Frakturtyp II (test 106 | 27,5%; re-test 111 | 28,8%; total 217 | 28,1%) dominierte wiederum der Subtyp II-A-P (test 88 | 22,8%; re-test 97 | 25,1%; total 179 | 23,2%). Der Frakturtyp Schatzker II-AP stellte somit nahezu jede vierte Fraktur im vorliegenden Kollektiv dar.

Im Bereich der Frakturen Schatzker I bis IV mit definitionsgemäßer Unterteilung in anteriore und posteriore Teilbereiche (A, P, A-P vgl. Schatzker V und VI mit Quadranteneinteilung AM, AL, PM, PL) überwog in allen Typen I bis IV die

Frakturklassifizierung A-P (Typ I in 75,0% aller Fälle; in Typ II in 85,3% aller Fälle; in Typ III in 46,6% aller Fälle und in Typ IV in 69,2% aller Fälle).

Der Frakturtyp V (test 114 | 29,5%; re-test 115 | 29,8%; total 229 | 29,7%) war gefolgt von Frakturtyp II (test 106 | 27,5%; re-test 111 | 28,8%; total 217 | 28,1%) am häufigsten vertreten. Ein dominierender Subtyp war im Frakturtyp V nicht auszumachen. Relativ häufig war u.a. eine diagonale Fraktur von antero-medial nach postero-lateral „Typ V-AM-PL“ (test 19 | 4,9%; re-test 29 | 7,5%; total 48 | 6,2%).

Bei beiden großen Frakturtypen V und VI zeigten sich, im Vergleich zu anderen Subtypen, vermehrt Frakturen mit Beteiligung aller vier Teilbereiche AL-AM-PL-PM (V-AL-AM-PL-PM: test 23 | 6,0%; re-test 15 | 3,9%; total 217 | 4,9% und VI AL-AM-PL-PM: test 19 | 4,9%; re-test 13 | 3,4%; total 32 | 4,1%).

3.4.2 Percent-agreement

In der Klassifikation nach Kfuri/Schatzker konnte in 313 Fällen | 81,1% eine Übereinstimmung der Klassifizierungen erzielt werden. 73 | 18,9% Fälle wurden folglich im re-test abweichend beurteilt.

Das höchste percent-agreement erzielte der im Vergleich zu den anderen Typen seltene Frakturtyp IV mit 92,6% | 25 Übereinstimmungen | 2 Fehler. Der wesentlich häufigere Frakturtyp II (82,8% | 96 Übereinstimmungen | 20 Fehler) erzielte ebenfalls ein gutes Ergebnis im percent-agreement. Die weiteren Frakturtypen I (73,9% | 17 Übereinstimmungen | 6 Fehler), III (70,8% | 46 Übereinstimmungen | 19 Fehler), V (71,1% | 86 Übereinstimmungen | 35 Fehler) und VI (71,7% | 43 Übereinstimmungen | 17 Fehler) lagen alle mit ca. 70% percent-agreement auf vergleichbarem Niveau.

3.4.3 Intrarater-Reliabilität gesamt

Die Intrarater-Reliabilität in der Klassifikation nach Kfuri/Schatzker lag bei einem κ -Wert von $0,793 \pm 0,022$ ($p < 0,0001$) und somit am oberen Ende des Bereichs „substantial“ nach Landis und Koch.

3.4.4 Intrarater-Reliabilität Subgruppen

Die Intrarater-Reliabilität des Frakturtyps I (laterale Spaltfrakturen) in der Klassifikation von Kfuri/Schatzker lag bei einem κ -Wert von $0,570 \pm 0,133$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch.

Bezüglich des Kfuri/Schatzker-Frakturtyps II „Laterale Spalt-Impressionsfrakturen“ lag die Intrarater-Reliabilität bei einem κ -Wert von $0,513 \pm 0,090$ ($p < 0,001$) und somit im Bereich „moderate“ nach Landis und Koch.

Bei den Impressionsfrakturen des lateralen Tibiaplateaus (Frakturtyp III) lag die Intrarater-Reliabilität im Bereich „moderate“ nach Landis und Koch ($\kappa = 0,566 \pm 0,076$; $p < 0,001$).

Bei den etwas selteneren Typ IV-Frakturen (mediale Plateau-Frakturen) lag die Intrarater-Reliabilität im Bereich „almost perfect“ ($\kappa = 0,856 \pm 0,095$; $p < 0,001$).

„Bikondyläre Frakturen ohne Unterbrechung der Epi- und Metaphyse“ \approx Frakturtyp V zeigten eine Intrarater-Reliabilität von $\kappa = 0,666 \pm 0,046$; $p < 0,001$ (nach Landis und Koch im Bereich „substantial“).

Der bikondyläre Frakturtyp VI („Trümmerfrakturen bzw. Frakturen mit vollständiger Unterbrechung der Epi- und Metaphyse“) wies einen κ -Wert von $0,653 \pm 0,070$; $p < 0,001$ (nach Landis und Koch im Bereich „substantial“) auf.

3.4.5 Häufige Fehler bzw. abweichende Bewertungen

In 73 von 386 Fällen (18,9%) konnte in der Schatzker-Klassifikation keine Übereinstimmung getroffen werden.

Ein kompletter Frakturtypen-Wechsel kam in 26 Fällen vor. Zum Subtypen-Wechsel kam es bei 47 Fällen.

Bei den Fällen mit Frakturtypenwechsel kam es am häufigsten (7 Fälle) zum Wechsel zwischen Typ II (laterale Spalt-Impressionsfrakturen) und Typ III (Impressionsfrakturen des lateralen Tibiaplateaus). In diesen Fällen wechselte das Klassifizierungs-Ergebnis 6-mal von II-A-P zu III-A-P.

In den höheren Kategorien änderte sich das Ergebnis hauptsächlich von Typ V auf Typ VI (4 Fälle).

Ein Subtypen-Wechsel kann in der Schatzker-Klassifikation mit dem Fehlen bzw. Addieren von Quadranten gleichgesetzt werden. Vermehrt und in insgesamt 5 Fällen wurde vom Typ II-A in den Typ II-A-P gewechselt (vgl. Abbildung 6).

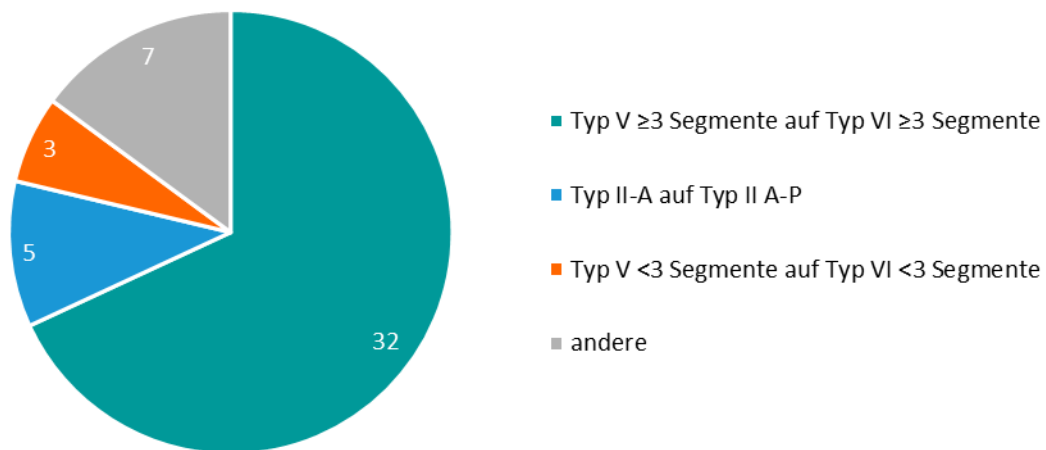


Abbildung 6: Häufigkeiten der Subtypen-Wechsel bei abweichender Bewertung - Kfuri/Schatzker

Die Häufigkeit von Subtypen-Fehlern in den Kategorien V und VI stieg mit der Anzahl der beteiligten Quadranten leicht an. Waren im ersten Durchlauf 2 beteiligte Quadranten diagnostiziert worden, kam es in lediglich 3 Fällen zu Subtypen-Wechseln. Waren 3 oder mehr Quadranten im ersten Durchlauf ermittelt worden, wurde 32-fach (68,1%) der Subtyp gewechselt (davon 16-mal mit Beteiligung aller 4 Segmente).

3.4.6 Sicherheits-Ranking

Die subjektive Sicherheit bezüglich der Entscheidungsfindung war in der Klassifikation nach Kfuri/Schatzker überwiegend gut („trifft eher zu“: test 148 | 38,3% und re-test 115 | 29,8%) bis sehr gut („trifft sicher zu“: test 222 | 57,5% und re-test 264 | 68,4%). Die subjektive Sicherheit nahm im re-test leicht zu („trifft sicher zu“ +10,7% im re-test und moderate Verluste in den Kategorien mit weniger Sicherheit) (vgl. Tabelle 8).

Tabelle 8: Sicherheits-Ranking - Kfuri/Schatzker-Klassifikation

| | n test | n re-test | +/- |
|------------------------|---------------|------------------|------------|
| trifft sicher zu | 222 57,5% | 264 68,4% | +10,7% |
| trifft eher zu | 148 38,3% | 115 29,8% | - 8,5% |
| trifft eher nicht zu | 16 4,1% | 7 1,8% | - 2,3% |
| trifft sicher nicht zu | 0 0,0% | 0 0,0% | ± 0 |

3.5 Ergebnisse zur Klassifikation der AO/OTA

3.5.1 Häufigkeiten - Deskriptive Statistik

Tabelle 9: Häufigkeiten, percent-agreements und kappa-Statistiken – AO-Klassifikation

| | test | re-test | %-agree | kappa κSD p |
|----------------------|-----------------------|-----------------------|----------------|----------------------------------|
| | <i>n Häufigkeit</i> | <i>n Häufigkeit</i> | | |
| <i>41B-Frakturen</i> | 225 58,3% | 218 56,5% | 82,0 | 0,762 0,033 <0,001 |
| 41B1.1 | 21 5,4% | 18 4,7% | 77,3 | |
| 41B1.2 | 9 2,3% | 8 2,1% | 70,0 | |
| 41B1.3h | 5 1,3% | 5 1,3% | 66,7 | |
| 41B1.3f | 1 0,3% | 0 0,0% | 100,0 | |
| 41B2.1 | 61 15,8% | 58 15,0% | 83,1 | |
| 41B2.2 | 5 1,3% | 5 1,3% | 100,0 | |
| 41B3.1 | 90 23,3% | 90 23,3% | 78,2 | |
| 41B3.2 | 7 1,8% | 8 2,1% | 87,5 | |
| 41B3.3h | 3 0,8% | 2 0,5% | 25,0 | |
| 41B3.3f | 23 6,0% | 24 6,2% | 38,2 | |
| <i>41C-Frakturen</i> | 161 41,7% | 168 43,5% | 69,0 | 0,570 0,048 <0,001 |
| 41C1.1 | 10 2,6% | 14 3,6% | 60,0 | |
| 41C1.2 | 11 2,8% | 14 3,6% | 56,3 | |
| 41C2.2h | 2 0,5% | 1 0,3% | 50,0 | |
| 41C2.2f | 0 0,0% | 0 0,0% | - | |
| 41C2.3 | 6 1,6% | 9 2,3% | 36,4 | |
| 41C3.1 | 75 19,4% | 70 18,1% | 61,1 | |
| 41C3.2 | 11 2,8% | 6 1,6% | 41,7 | |
| 41C3.3 | 46 11,9% | 54 14,0% | 53,8 | |
| Total | 100% | 100% | 79,0 | 0,756 0,024 <0,001 |

In der Tabelle 9 sind alle wichtigen Daten zu Häufigkeiten, percent-agreements und kappa-Statistiken zur AO-Klassifikation aufgelistet.

Die 41B-Frakturen (n=222 | 57,4%) wurden in dieser Studie im Durchschnitt aus test und re-test etwas häufiger diagnostiziert als die 41C-Frakturen (n=165 | 42,6%). Die häufigsten Frakturtypen waren in absteigender Reihenfolge der Typ 41B3.1 „monokondyläre, laterale Spalt-Impressions-Fraktur“ mit einer Häufigkeit von 90 Fällen (23,3%), der Typ 41C3.1 „bikondyläre Fraktur mit fragmentiertem lateralem Plateau“ mit durchschnittlich 73 Fällen (18,8%), der Typ 41B2.1 „monokondyläre, laterale Impressionsfraktur“ mit 60 Fällen (15,4%) und die Trümmerfraktur vom Typ 41C3.3 „bikondyläre Fraktur mit fragmentiertem lateralem und medialem Plateau“ mit 50 Fällen (13,0%). Alle anderen Frakturtypen wurden deutlich seltener diagnostiziert.

Insgesamt dominierten folglich laterale Frakturen unterschiedlichen Typs und Ausmaßes. Bei 222 41B-Frakturen (monokondylär) war in 193 Fällen eine laterale Beteiligung zu finden (87,1%). 41C-Frakturen beinhalten definitionsgemäß einen lateralen Anteil, da sie bikondylär sind.

3.5.2 Percent-agreements

Das percent-agreement lag in der Klassifikation nach AO bei insgesamt 79,0% (305 Übereinstimmungen | 81 Fehler). Die 41B-Frakturen zeigten ein höheres percent-agreement (82% | 182 Übereinstimmungen | 40 Fehler) als die 41C-Frakturen (69% | 114 Übereinstimmungen | 51 Fehler).

Die Abbildung 7 zeigt u.a. die percent-agreements jedes Subtyps der 41B-Frakturen. Insgesamt zeigte sich ein hohes percent-agreement in 8 von 10 Subtypen (Ø82,5% | range 66,7%-100%). Die zwei verbleibenden Subtypen 41B3.3h „monokondyläre, mediale Impressions-Spalt-Fraktur mit Einbeziehung der Eminentia intercondylaris“ (25,0% agreement | 1 Übereinstimmung | 3 Fehler) bei sehr geringer Anzahl an Fällen n=4) und 41B3.3f „monokondyläre, laterale Impressions-Spalt-Fraktur mit Einbeziehung der Eminentia intercondylaris“ (nur 38,2% agreement | 13 Übereinstimmungen | 21 Fehler | bei immerhin n=34 Fällen) hatten im Vergleich zu den anderen Subtypen deutlich niedrigere Werte. Beim Frakturtyp 41B3.3h wurde in 7 von 34 Fällen (20,6% | 5-mal B->C | 2-mal

C->B) der Frakturtyp gewechselt. Beim medialen Gegenstück 41B3.3f sogar in 3 von 4 Fällen (75% | 2-mal B->C | 1-mal C->B).

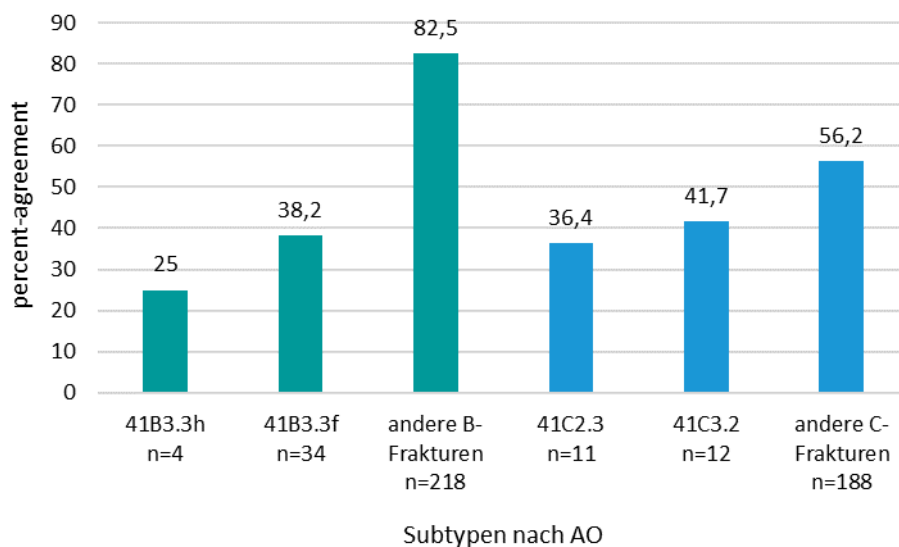


Abbildung 7: Percent-agreements bei 41B- und 41C-Fraktersubtypen – AO-Klassifikation

Im Vergleich zu den 41B-Frakturen zeigten die 41C-Frakturen ein geringeres percent-agreement. Beim Subtyp 41C2.3 „bikondyläre Fraktur mit multifragmentärer metaphysärer Fraktur“ mit 63,6% Fehlerquote total (inkl. Typenwechsel) (36,4% percent-agreement | 4 Übereinstimmungen | 7 Fehler | n=11) lag der Fehler in 63,3% bei der Wahl eines anderen Subtyps im re-test. Im Falle des Subtyps 41C3.2 „bikondyläre Fraktur mit fragmentiertem lateralem Plateau“ mit 58,3% Fehlerquote total (41,7% percent-agreement | 5 Übereinstimmungen | 7 Fehler | n=12) lag der Anteil eines anders gewählten Subtyps bei 50% (50% der Fehler auf die Wahl eines unterschiedlichen Frakturtyps zurückzuführen). In der Kategorie der kompletten Trümmerfrakturen 41C3.3 lag der Anteil der falschen Subtyp-Wahl am Gesamtfehler-Ergebnis (53,8% percent-agreement | 35 Übereinstimmungen | 30 Fehler | n=65) bei immerhin 44,6% (55,4% anderer Frakturtyp).

3.5.3 Intrarater-Reliabilität gesamt

Die Intrarater-Reliabilität in der Klassifikation nach AO lag bei einem κ -Wert von $0,756 \pm 0,024$ ($p < 0,0001$) und somit im Bereich „substantial“ nach Landis und Koch.

3.5.4 Intrarater-Reliabilität Subgruppen

Die Intrarater-Reliabilität der 41B-Frakturen (monokondylär) lag bei einem κ -Wert von $0,762 \pm 0,033$ ($p < 0,001$) und somit im Bereich „substantial“ nach Landis und Koch.

Die Intrarater-Reliabilität der 41C-Frakturen (bikondylär) lag bei einem κ -Wert von $0,570 \pm 0,048$ ($p < 0,001$) und somit im oberen Bereich der Kategorie „moderate“ nach Landis und Koch.

3.5.5 Häufige Fehler bzw. abweichende Bewertungen

In der AO-Klassifikation konnte in 81 von 386 (21,0%) Fällen keine Übereinstimmung getroffen werden.

Ein Typen-Wechsel von Typ B (monokondylär) zu Typ C (bikondylär) kam in 13 Fällen (16,0% aller Fehler) vor (10-mal B->C und 3-mal C->B). Die Anzahl der Subtypen-Wechsel betrug 68 (84,0% aller Fehler).

Am häufigsten und in insgesamt 25 Fällen (36,8% der Subtypen-Wechsel) wurde vom Typ 41C3.1 (bikondyläre Fraktur mit fragmentiertem lateralem Plateau) zu 41C3.3 (bikondyläre Fraktur mit fragmentiertem lateralem und medialem Plateau) gewechselt. Vermehrt wurde zusätzlich der Typ 41B3.1 (monokondyläre, laterale Spalt-Impressions-Fraktur) vom Typ 41 B3.3f (laterale Impressions-Spalt-Fraktur mit Einbeziehung der Eminentia intercondylaris) ersetzt (13 Fälle | 15,1% aller Subtypen-Wechsel).

3.5.6 Sicherheits-Ranking

Die subjektive Sicherheit bezüglich der Entscheidungsfindung war in der Klassifikation nach AO mehrheitlich sehr gut („trifft sicher zu“: test 230 | 59,6%

und re-test 256 | 66,3%). Eine Entscheidung mit guter subjektiver Sicherheit konnte in etwas mehr als einem Drittel der Fälle („trifft eher zu“: test 138 | 35,8% und re-test 120 | 31,1%) getroffen werden. Die subjektive Sicherheit konnte im zweiten Umlauf leicht erhöht werden („trifft sicher zu“ +6,7% im re-test und moderate Verluste in den Kategorien mit weniger Sicherheit) (vgl. Tabelle 10).

Tabelle 10: Sicherheits-Ranking – AO-Klassifikation

| | n test | n re-test | +/- |
|------------------------|-------------|-------------|--------|
| trifft sicher zu | 230 59,6% | 256 66,3% | +6,7% |
| trifft eher zu | 138 35,8% | 120 31,1% | - 4,7% |
| trifft eher nicht zu | 18 4,7% | 10 2,6% | - 2,1% |
| trifft sicher nicht zu | 0 0,0% | 0 0,0% | ± 0 |

3.6 Zusammenfassung der Ergebnisse

3.6.1 Intrarater-Reliabilitäten und percent-agreements

Die Tabelle 11 zeigt einen Überblick über die Intrarater-Reliabilitäten und percent-agreements der vier Klassifikationen. Luo et. al zeigte mit $\kappa=0,848$ die höchste Intrarater-Reliabilität und kann als einziges System als „(almost) perfect“ nach Landis und Koch bewertet werden. Luos Klassifikation zeigte ebenfalls das höchste percent-agreement (89,1).

Tabelle 11: Intrarater-Reliabilitäten und percent-agreements der Klassifikationen im Vergleich

| | Luo et al. | Krause et al. | Kfuri/Schatzker | AO |
|----------------------------------|--------------------|--------------------|--------------------|--------------------|
| Kappa κ | 0,848±0,022 | 0,621±0,025 | 0,790±0,022 | 0,756±0,024 |
| κ SD | 0,022 | 0,025 | 0,022 | 0,024 |
| p-Wert | <0,001 | 0,000 | 0,000 | 0,000 |

| percent-agreement | 89,1 | 64,5 | 81,1 | 79,9 |
|-------------------|------|------|------|------|
| | | | | |

Die Klassifikation von Krause et al. hatte die geringste Reliabilität ($\kappa=0,621$) und kann nach Landis und Koch als „substantial“ bewertet werden. In Bezug auf das percent-agreement zeigte Krauses Klassifikation mit 64,5 ebenfalls den niedrigsten Wert im Vergleich. Die Klassifikationen nach Kfuri/Schatzker ($\kappa=0,790$) und AO ($\kappa=0,756$) erreichten einen ähnlich Wert der Intrarater-Reliabilität an der oberen Grenze des Bereiches „substantial“ nach Landis und Koch. Im Vergleich der percent-agreements zeigten beide Klassifikationen vergleichbare Werte (Kfuri/Schatzker 81,1 und AO 79,9).

3.6.2 Sicherheits-Rankings

Im subjektiven Sicherheits-Ranking (Bewertungszeit des Raters bei einer Fraktur je Klassifikation) wies die Klassifikation nach Luo et al. die besten Werte auf. Sie zeigte in beiden Durchläufen einen hohen Wert in der Kategorie „trifft sicher zu“ (test 76,6% | re-test 2 72,8%). Die Luo-Klassifikation war die einzige Klassifikation, die im re-test (leicht) an subjektiver Sicherheit verlor (-3,8% „trifft sicher zu“).

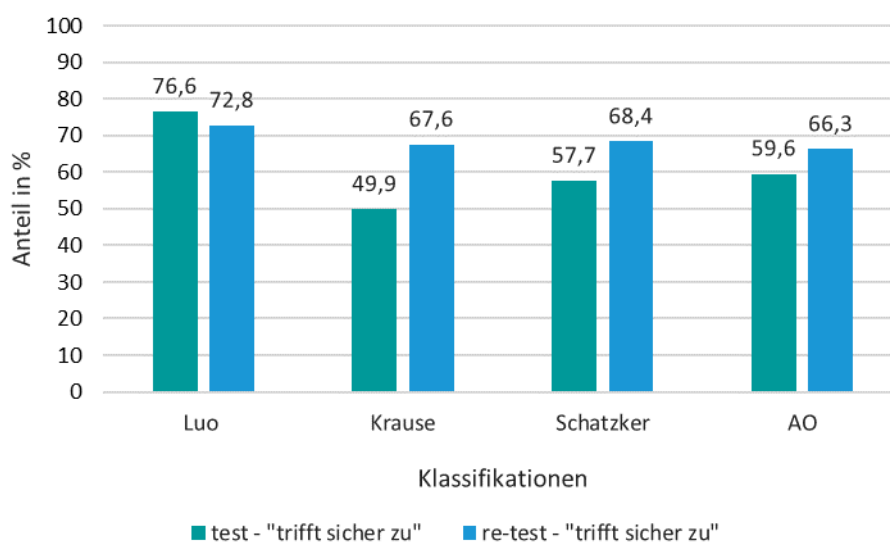


Abbildung 8: Vergleich des Sicherheits-Rankings der Klassifikationen in Kat.1 "trifft sicher zu"

Die Klassifikationen von Kfuri/Schatzker (test 57,7% | re-test 68,4% | +10,7%) und AO (test 59,6% | re-test 66,3% | +6,7%) erreichten vergleichbare Werte in

der Kategorie „trifft sicher zu“. Beide konnten im re-test an Sicherheit gewinnen. Die Klassifikation von Krause et al. zeigte die größte Zunahme an Bewertungssicherheit zwischen test und re-test (+17,8%). Sie erreichte im zweiten Durchlauf (67,6% „trifft sicher zu“) ein vergleichbares Ergebnis wie die Klassifikationen von Kfuri/Schatzker und AO (vgl. Abbildung 8).

3.6.3 Statistische Werte der Klassifikationen und Vergleich nach Frakturgröße

Die Tabelle 12 verdeutlicht die gemessenen percent-agreement- und kappa-Werte der vier Klassifikationen in den definierten Frakturgrößen-Gruppen (siehe Abschnitt 2.6).

Die Klassifikation nach Luo et al. zeigte das höchste percent-agreement im Bereich der größeren Frakturen (gFs 86,2 | TFs 86,2 im Vergleich zu kFs 73,3 | mFs 79,2).

Die Klassifikation von Krause et al. zeigte die höchsten Werte im percent-agreement bei den Trümmerfrakturen (TFs 92,3 im Vergleich zu kFs 63,1 | mFs 47,2 | gFs 51,0) Die Klassifikation von Kfuri/Schatzker wies in 3 Frakturgrößen-Gruppen mit gut 70% vergleichbare Werte auf (kFs 74,1 | gFs 71,1 | TFs 71,7). Der höchste Wert zeigte sich in der Gruppe der mittleren Frakturgrößen (84,6). Die AO-Klassifikation zeigte im Bereich der kleinen Frakturen den höchsten Wert. Mit zunehmender Frakturgröße verringerten sich die Werte (kFs 83,7 | mFs 74,6 | gFs 61,5 | TFs 53,8).

Im folgenden Abschnitt werden die Ergebnisse differenziert nach Frakturgrößen-Gruppen beschrieben.

Bei den „kleinen Frakturen“ lagen die höchsten Werte in der AO-Klassifikation (83,7 percent-agreement | $\kappa=0,739$ mit $p<0,001$).

In der Gruppe „mittlere Frakturgrößen“ erzielte die Kfuri/Schatzker-Klassifikation den höchsten Wert (84,6 percent-agreement | $\kappa=0,724$ mit $p<0,001$).

Im Bereich der „großen Frakturen“ zeigte die Kfuri/Schatzker-Klassifikation den höchsten kappa-Wert ($\kappa=0,724$ mit $p<0,001$), wohingegen die Luo-Klassifikation das höchste percent-agreement erzielte (86,2%). Ein kappa-Wert konnte in der

Luo-Klassifikation aufgrund der eindimensionalen Kategorie (nur post-med-lat) und der hohen Übereinstimmungsrate mathematisch nicht errechnet werden. Im Bereich der „Trümmerfrakturen“ erzielten die Systeme nach Luo et al. (86,2%) und Krause et al. (92,3%) hohe percent-agreements. Die AO-Klassifikation wies den niedrigsten Wert auf (53,8%). Einen auswertbaren kappa-Werte zeigte die Klassifikation nach Kfuri/Schatzker (Typ VI | $\kappa=0,653$ bei $p=0,001$). Die anderen Klassifikationen haben definitionsgemäß für die maximale Ausprägung von Tibiakopffrakturen nur eine Kategorie (Luo = 3 Säulen | Krause = 10 Segmente | AO = 41C3.3) und lassen deshalb keine Berechnung eines kappa-Wertes zu.

Tabelle 12: Vergleich der Klassifikationen nach Frakturgröße

| | Anzahl n | %-agree | kappa κ SD p |
|---------------------------------|--------------------|-------------|-------------------------|
| <i>kleine Frakturen (kFs)</i> | | | |
| Luo et al. | 75 19,4% | 73,3 | 0,674 0,061 <0,001 |
| Krause et al. | 122 31,6% | 63,1 | 0,585 0,048 <0,001 |
| Kfuri/Schatzker | 85 22,0% | 74,1 | 0,660 0,059 <0,001 |
| AO | 104 26,9% | 83,7 | 0,739 0,055 <0,001 |
| <i>mittlere Frakturen (mFs)</i> | | | |
| Luo et al. | 192 49,7% | 79,2 | 0,484 0,056 <0,001 |
| Krause et al. | 127 32,9% | 47,2 | 0,411 0,046 <0,001 |
| Kfuri/Schatzker | 143 37,0% | 84,6 | 0,724 0,053 <0,001 |
| AO | 134 34,7% | 74,6 | 0,506 0,063 <0,001 |
| <i>große Frakturen (gFs)</i> | | | |
| Luo et al. | 159 41,2% | 86,2 | -0,052 0,012 0,386 |
| Krause et al. | 102 26,4% | 51,0 | 0,482 0,051 <0,001 |
| Kfuri/Schatzker | 121 31,3% | 71,1 | 0,666 0,046 <0,001 |
| AO | 135 35,0% | 61,5 | 0,432 0,058 <0,001 |
| <i>Trümmerfrakturen (TFs)</i> | | | |
| Luo et al. | 159 41,2% | 86,2 | -0,052 0,012 0,386 |
| Krause et al. | 65 16,8% | 92,3 | 0,000 0,000 - |
| Kfuri/Schatzker | 60 15,5% | 71,7 | 0,653 0,070 <0,001 |
| AO | 65 16,8% | 53,8 | -0,229 0,050 0,025 |

3.6.4 Anzahl der unterschiedlich bewerteten Klassifikationen je Fallbewertung

Im folgenden Abschnitt werden die Bewertungsfehler je Fall hinsichtlich der 4 Klassifikationen betrachtet. Dabei kann ein Fall in 0 (komplette Übereinstimmung der Bewertung in allen Klassifikationen) bis in maximal 4 Klassifikationen (Übereinstimmung in keiner Klassifikation) abweichend bewertet werden.

Die Tabelle 13 zeigt die Anzahl der Fälle mit entsprechenden Bewertungsdifferenzen in den Klassifikationen kumulativ.

Tabelle 13: Anzahl unterschiedlich bewerteter Klassifikationen je Fallbewertung

| | Differenzen in n Klassifikationen | | | | |
|----------------|-----------------------------------|-------------|------------|-----------|----------|
| | n=0 | n=1 | n=2 | n=3 | n=4 |
| Total | 158 40,9% | 147 38,1% | 58 15,0% | 22 5,7% | 1 0,3% |
| test | | | | | |
| Kleine Fs | 74 49,3% | 47 31,3% | 18 12,0% | 11 7,3% | 0 0,0% |
| Mittlere Fs | 91 40,3% | 85 37,6% | 33 14,6% | 16 7,1% | 1 0,4% |
| Große Fs | 54 30,7% | 81 46,0% | 29 16,5% | 12 6,8% | 0 0,0% |
| Trümmer Fs | 51 32,2% | 72 45,6% | 28 17,7% | 7 4,4% | 0 0,0% |
| re-test | | | | | |
| Kleine Fs | 74 51,0% | 48 33,1% | 17 11,7% | 6 4,1% | 0 0,0% |
| Mittlere Fs | 91 41,2% | 87 39,4% | 31 14,0% | 12 5,4% | 0 0,0% |
| Große Fs | 54 30,2% | 82 45,8% | 29 16,2% | 13 7,3% | 1 0,6% |
| Trümmer Fs | 51 33,1% | 65 42,2% | 28 18,2% | 9 5,8% | 1 0,6% |

In 158 Fällen | 40,9% gab es eine komplett übereinstimmende Bewertung in test und re-test. In 147 Fällen | 38,1% lag eine Differenz in einer Klassifikation vor. In 58 Fällen | 15,0% wurde in 2 Kategorien abweichend bewertet. Seltener waren verschiedene Bewertungen in 3 Klassifikationen (22 | 5,7%) und 4 Klassifikationen (1 | 0,3%).

In der Folge werden die Ergebnisse der kumulativen Differenzen in Bezug auf die Frakturgrößen-Gruppen aufgezeigt. Die Übereinstimmungen (n=0) sinken mit zunehmender Frakturgröße bis zur Subgruppe „große Frakturen“ (kFs mit n=0 bei 49,3%/51,0% | mFs mit n=0 bei 40,3%/41,2% | gFs mit n=0 bei 30,7%/30,2% | TFs mit n=0 bei 32,2%/33,1%). Die „große Frakturen“ und „Trümmerfrakturen“ zeigten vergleichbare Werte im Bereich n=0 (Übereinstimmung in allen 4 Klassifikationen).

In den Kategorien mit nicht übereinstimmenden Bewertungen in einer oder zwei Klassifikationen waren die Werte bei den „kleinen Frakturen“ am niedrigsten (kFs mit n=1 bei 31,3%/33,1% und kFs mit n=2 bei 12,0%/11,7%). Höhere Werte im Bereich der nicht übereinstimmenden Bewertung in einer oder zwei Klassifikationen zeigten die „großen Frakturen“ (gFs mit n=1 bei 46,0%/45,8% und gFs mit n=2 bei 16,5%/16,2) und die „Trümmerfrakturen“ (TFs mit n=1 bei 45,6%/42,2% und TFs mit n=2 bei 17,7%/18,2%). Die Anzahl der Fälle mit dreifachem Bewertungsfehler n=3 lag für alle Frakturgrößen-Subgruppen vergleichbar im einstelligen Prozentbereich (4,1% - 7,3%).

Die Durchläufe „test“ und „re-test“ müssen aufgrund der Definition des Ausmaßes über die Klassifikation separat ausgewiesen werden. Die Ergebnisse von test und re-test unterschieden sich letztlich marginal (vgl. Abbildung 9).

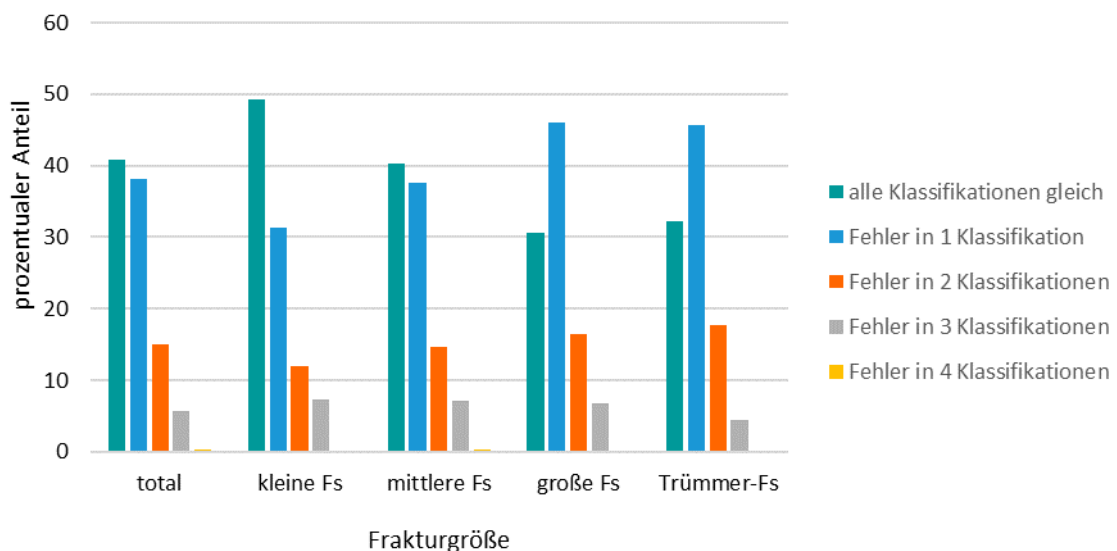


Abbildung 9: Bewertungsfehler je Fall nach Frakturgröße

3.6.5 Zusammenfassung der Ergebnisse aller vier Klassifikationen

Die Tabelle 14 veranschaulicht und vergleicht in einer übersichtlichen Zusammenfassung die Ergebnisse der einzelnen Klassifikationen.

Tabelle 14: Zusammenfassung der Qualitäten aller vier Klassifikationen

| | Luo | Krause | Schatzker | AO |
|---------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| | <i>kappa</i> <i>p.a.</i> | <i>kappa</i> <i>p.a.</i> | <i>kappa</i> <i>p.a.</i> | <i>kappa</i> <i>p.a.</i> |
| Total | 0,848 89,1 | 0,621 64,5 | 0,793 81,1 | 0,756 79,9 |
| <i>Landis und Koch</i> | alm. perfect | substantial | substantial | substantial |
| Kleine Frakturen | 0,674 73,3 | 0,585 63,1 | 0,660 74,1 | 0,739 83,7 |
| Mittlere Frakturen | 0,484 79,2 | 0,411 47,2 | 0,742 84,6 | 0,506 74,6 |
| Große Frakturen | n.s. 86,2 | 0,482 51,0 | 0,666 71,1 | 0,432 61,5 |
| Trümmerfrakturen | n.s. 86,2 | n.a. 92,3 | 0,653 71,7 | n.s. 53,8 |
| Subjektive Sicherheit | + | - | o | o |
| Erlernbarkeit | + | - | o | o |
| Präzision/Frakturverlauf | -- | o | ++ | + |
| Ableitung OP-Strategie | - | o | ++ | + |

Die bisher beschriebenen Ergebnisse aus den oberen Kapiteln werden im Folgenden noch ergänzt um die subjektive Bewertung des Informationsgehaltes der Klassifikationen (Präzision/Frakturverlauf) und um die subjektive Bewertung der Möglichkeit der Ableitung einer Operationsstrategie.

Die Luo-Klassifikation ist per Definition in 4 Kategorien eingeteilt. Eine detaillierte Differenzierung der Frakturen und eine exakte Beschreibung des Frakturverlaufs gelingt häufig nur unzureichend. Die Ableitung einer OP-Strategie kann somit

ebenfalls erschwert sein. Krauses 10-Segment-Klassifikation zeigt eine gute Informationsdichte im Bereich der ersten 3cm unterhalb der Gelenkfläche. Der weiter distal liegende Bereich wird in Krauses Klassifikation nicht berücksichtigt. Das Fehlen dieser Daten kann die Ableitung einer chirurgischen Strategie beeinträchtigen. Die Klassifikation von Kfuri/Schatzker ist laut den Autoren eine gezielt auf die chirurgische Anwendung modifizierte Version der klassischen Schatzker-Klassifikation. Die Beschreibung des Frakturverlaufs mit Hilfe der Quadranten-Einteilung gelingt gut. Die spezifischen Modifikationen hinsichtlich der chirurgischen Planung lassen eine Strategie hinsichtlich des Zugangs und der Osteosynthesetechniken zu. Die AO-Klassifikation zeigt eine gute Verständlichkeit durch eine Visualisierung der definierten Frakturen und Frakturgruppen. Der Informationsgehalt ist solide, jedoch im Bereich der C-Frakturen überdifferenziert. Eine Therapieableitung kann gelingen.

4 Diskussion

4.1 Allgemein

Die modernen, speziell für das CT entwickelten oder überarbeiteten, 3D-basierten Klassifikationssysteme zur Frakturanalyse und Dokumentation am Tibiakopf von Krause et al. (2016) [8] und Kfuri/Schatzker (2018) [27] sind erst in den letzten Jahren entwickelt worden. Audigé et al. [29] postulierte wissenschaftliche und statistische Standards für Reliabilitäts-Studien für Fraktur-Klassifikationssysteme und beurteilte bisherige Arbeiten zu Klassifikationen nach 10 Qualitäts-Kriterien. Er beschreibt das methodische Vorgehen der begutachteten Reliabilitäts-Studien als pragmatisch, unvollständig und inadäquat. Unter anderem in den Bereichen „Repräsentatives Studienkollektiv“, „Begründung der Kollektiv-Größe“, „Verblindung“ und „korrekte statistische Methoden“ konnten weniger als die Hälfte der Studien überzeugen. Bewertet man die bisher in der Literatur zu findenden Reliabilitätsanalysen anhand Audigés 10 Qualitäts-Kriterien ergaben sich deutliche Limitationen. Die Autoren der von Audigé begutachteten Reliabilitätsstudien versuchen möglichst viele

Aussagen bezüglich Reliabilitäten und Qualitäten der Klassifikationssysteme anhand relativ kleiner Studienkollektive zu treffen. Differenzierte, auch subgruppen-spezifische Aussagen über Reliabilitäten und Qualitäten konnten nicht getroffen werden. In der Arbeit von Audigé et al. [29] variierten die eingeschlossenen Klassifikationen stark. Vor allem die speziell für die CT entwickelten, neuen Systeme, sind in den Analysen überhaupt nicht zu finden [10, 21, 26, 40, 51, 58-73]. Die Datenlage ist unvollständig. Es fehlt eine Basis an Daten, mit Hilfe derer Aussagen u.a. über die Stärken, Schwächen und Reliabilitäten der Klassifikationssysteme möglich sind. Darüber hinaus fehlen diese Daten und Bewertungen auch als Startpunkt für weitere Schritte (z.B. Multicenter-agreement-Studien mit professionellen Anwendern und anschließender prospektiver, klinischer Anwendungsstudien) hin zu einer Anwendungsempfehlung oder im besten Fall zur Definition eines Goldstandards. In der vorliegenden Studie sollten methodische Mängel minimiert werden. Es wurde sich streng an den geforderten wissenschaftlichen Qualitäts-Kriterien nach Audigé et al. [13] orientiert. Mit Hilfe eines sehr großen und im Umfang begründeten, repräsentativen Studienkollektivs, sollten Aussagen fundiert getroffen werden. Eine klare Verblindung hinsichtlich klinischer Informationen und eine enge Zusammenarbeit mit Experten auf dem Gebiet der medizinischen Statistik waren essenzielle Eckpunkte.

4.2 Patientenkollektiv, Methoden und Limitationen

In dieser Studie wurden erstmals Reliabilitäten der am häufigsten untersuchten und verwendeten, modernen, CT-basierten Tibiakopf-Frakturklassifikation mit Hilfe eines großen Patientenkollektivs (n=386) evaluiert. Das Kollektiv beinhaltete im Vergleich zu ähnlichen Arbeiten vergleichbare und ausgeglichene Werte in den Bereichen Geschlecht, Alter und betroffene Extremität (links/rechts). Aufgrund der großen Fallzahl konnten auch Subgruppen (z.B. Frakturgröße) der Klassifikationen hinsichtlich ihrer Reliabilitäten detailliert verglichen werden. Die Kfuri/Schatzker-Klassifikation sowie die 10-Segment-

Klassifikation wurden durch die vorliegende Arbeit erstmalig hinsichtlich der Reliabilität überprüft.

Es bestehen nur wenige vergleichbare Studien. Im letzten Jahrzehnt setzte sich die Diagnostik mittels Computertomographie am Tibiakopf durch. Mehrheitlich wurden Studien durchgeführt, die die Reliabilitäten der Klassifikationssysteme hinsichtlich des Einsatzes konventioneller Röntgentechnik oder der Anwendung der Computertomographie vergleichen. Sie treffen mehrheitlich Aussagen über die Vorteile 3D-basierter Diagnostik [10, 40, 41, 59, 62, 65-67, 74]. Alle weiteren Arbeiten in der Literatur, die sich mit Reliabilitäten von Klassifikations-Systemen (CT-Diagnostik) am Tibiakopf beschäftigen, zählen inkl. der o.g. Studien im Schnitt lediglich 45 Fälle (15-114 Fälle) [51, 58, 60, 61, 63, 64, 72, 75, 76]. Die Intrarater-Reliabilität wurde in 5 Studien nicht beschrieben [41, 62-64, 76]. Inhaltlich beschäftigen sich 16 von 17 Studien mit der älteren, nicht für die 3D-Bildgebung überarbeiteten Schatzker-Klassifikation. 9 Autoren analysierten u.a. die Reliabilität der AO-Klassifikation und 6 inkludierten u.a. das „Three-Column-Concept“ von Luo et al. in ihre Studie. Martinez-Rondanelli et al. [60], Anwar et al. [75] (jeweils eigene 4-Säulen-Systeme) und Hoekstra et al. [61] (modifizierte Luo-Klassifikation) verglichen die gängigsten Klassifikationen mit selbst formulierten Klassifikationen. Die Analyse von Intrarater-Reliabilitäten in Subgruppen (u.a. verschiedene Trauma-Mechanismen) wird lediglich in der Arbeit von Zhang et al. [51] beschrieben. Insgesamt leiden die bisherigen Veröffentlichungen unter kleinen Studienkollektiven und den damit verbundenen Schwierigkeiten (Signifikanzlevel, Analyse von Subgruppen, Power). In einem methodischen Qualitäts-Rating von Millar et al. [21], aufgestellt mittels COSMIN-Checkliste für Reliabilitätsstudien [77], wurden die Publikationen von Chan et al. [74], Hu et al. [67], Mellema et al. [62] und te Stroet et al. [66] mit „poor“ bewertet. Für Reliabilitäts-Analysen gibt es eine Vielzahl möglicher statistischer Methoden. In dieser Studie entschieden wir uns für die Verwendung des Korrelationskoeffizienten „Cohens kappa“ und der Berechnung des percent-agreements. Es sind die gebräuchlichsten Techniken für Reliabilitäts-Analysen [78]. McHugh et al. [57] empfahlen die additive Verwendung beider Methoden. Die 17 o.g. Vergleichsarbeiten verwendeten ebenfalls alle Cohens kappa. Audigé

et al. [13] empfehlen eine Anwendung von Cohens kappa für Reliabilitäts-Analysen, fordern jedoch eine umsichtige Interpretation und warnen vor zu schnellen und falschen Interpretationen. Im besten Fall sollten die Klassifikationen zusätzlich separat bezüglich ihrer Teil-Kategorien verglichen werden. Die konkreten Zahlen sollten keinesfalls zur Beantwortung aller wesentlicher Fragen während der Evaluation eines Klassifikationssystems verwendet werden. Sie mahnen zusätzlich zur Vorsicht bei Vergleichen zwischen Studien, da kappa-Werte stark von der Verteilung der Frakturtypen in der Studienpopulation abhängen würden. Die Interpretations-Rangskalen u.a. von Landis und Koch [56] sind willkürlich festgelegt. Es gibt keine Übereinkunft, ab wann ein kappa-Wert als akzeptierbar eingeschätzt werden kann. Die Verwendung der Interpretations-Skalen sollte mit Bedacht, den theoretischen Hintergrund im Blick behaltend und nicht isoliert von anderen statistischen Werten geschehen. Eine abschließende Einschätzung der Reliabilität von Klassifikationen und die Beantwortung aller wichtigen Fragen zur Qualität der Klassifikationen können sie letztendlich nicht abbilden [29].

Die Studien von Maripuri et al. [68] und Albuquerque et al. [79] verwendeten ebenfalls das percent-agreement additiv. Es war bis zur Einführung von Cohens kappa die Standard-Technik zur Reliabilitäts-Analyse. Sie ist vor allem für binäre Fragestellungen einfach durchzuführen, detektiert problematische Variablen und lässt im Vergleich zu den kappa-Werten direkt Interpretationen zu. Schwächen des percent-agreements beziehen sich auf den falschen Einschluss von „geschätzten“ oder „geratene“ Kategorisierungen bei unerfahrenen oder untrainierten Bewertern. Cohen zog diese Schwäche als Anlass für die Entwicklung seines kappas (chance-corrected) heran. Letztendlich werden im percent-agreement errechnete Übereinstimmungen eher überschätzt [57].

In den folgenden Abschnitten sollen die Limitationen dieser Studie dargestellt werden. Wie in 1.2 aufgeführt orientierte sich die vorliegende Arbeit strikt an den von Audigé et al. [13], speziell für Fraktur-Klassifikationen aufgestellten, zehn Qualitätskriterien für Reliabilitäts-Analysen. Sie sind in Tabelle 1 aufgeführt. Die Kriterien 6 (\approx ausreichend hohe Anzahl an Ratern) und 7 (\approx Rater klassifizieren unabhängig voneinander) sind im Bereich der Messung von Interrater-

Reliabilitäten von zentraler Bedeutung (mehrere unabhängige Bewerter). Sie passen methodisch nicht zum Studiendesign, das für diese Doktorarbeit verwendet wurde (Messung der Intrarater-Reliabilität \approx 1 Bewerter). Zukünftige Arbeiten unserer Arbeitsgruppe werden jedoch entsprechende Interrater-Reliabilität durch zwei weitere Observer ergänzen. 6 der verbleibenden 8 Kriterien (Punkte 1 bis 4, sowie Punkte 8 und 10) konnten in der vorliegenden Studie ohne Einschränkungen umgesetzt werden. Die letztlich insuffizient erfüllten oder aus bestimmten Gründen nicht erfüllbaren Gütekriterien 5 (repräsentative Bewerter) und 9 (Durchführung einer Goldstandard-Bewertung \approx Schätzung der „wahren“ Befunde durch Experten, intraoperative Befunde, etc.) werden im Folgenden als Kritikpunkte beschrieben.

Ein bedeutsamer Kritikpunkt dieser Studie liegt im Bereich der grundsätzlichen Aussagekraft von Intra- und Interrater-Reliabilität. Es ist wichtig zu verstehen, dass die Intrarater-Reliabilität die Reproduzierbarkeit unabhängig von der Richtigkeit (Validität) misst. Eine falsche Einschätzung kann infolgedessen immer noch eine gute Intrarater-Reliabilität ergeben, auch wenn sie eigentlich widerspiegelt, dass der Rater ständig falsch lag. Es existieren zwei Möglichkeiten die Richtigkeit der Einschätzung, zumindest näherungsweise, zu überprüfen (Gütekriterium 9 nach Audigé et al. [13]). Zum einen kann die Klassifizierung mit intraoperativen Befunden verglichen werden. Ein ethisch unzulässiger Weg, da während einer Operation aus Gründen der Invasivität meistens nicht das ganze Frakturausmaß dargestellt wird. Im postoperativen Stadium (retrospektive Studie) kann diese Möglichkeit nicht umgesetzt werden. Zum anderen besteht die Möglichkeit, die Befunde mehrerer unabhängiger Rater oder eines Expertengremiums zunächst miteinander zu vergleichen und anschließend mittels Übereinstimmungs-Quote (Konsensus) einen „Goldstandard“ zu definieren. So kann der wahren Begebenheit zumindest nähergekommen werden (sog. gold-standard-assessment) [13, 29, 72]. Für den Versuch, sich an die Richtigkeit anzunähern, ist der von Kottnerus and Muris [80] geprägte Begriff „Referenz-Standard“ die wahrscheinlich bessere Wahl.

Als weiterer Kritikpunkt der vorliegenden Studie kann genannt werden, dass kein Experte als Rater eingesetzt wurde und somit die späteren Nutzer der

Klassifikation nicht ausreichend repräsentiert wurden (Gütekriterium 5 nach Audigé et al. [13]). Fünf Autoren, die sich mit Reliabilitäts-Studien am Tibiakopf beschäftigten, konnten in ihren Publikationen keinen Einfluss der Erfahrung des Raters auf die Intrarater-Reliabilität nachweisen [26, 40, 59, 72, 81]. Methodisch sollte bei unerfahreneren Ratern bezüglich der internen Validität ein sogenannter „instrument change“ zwischen erstem und zweitem Durchlauf angesprochen werden. Hierbei bewertet der Rater im re-test mit mehr Erfahrung durch das Training im ersten Durchgang. Dies zählt methodisch zu einem Fehler bei der zweiten Messung, kann aber bei Intrarater-Reliabilitäts-Analysen kausal nicht gänzlich ausgeschlossen werden.

4.3 Reliabilität, Stärken und Schwächen der Klassifikationen

4.3.1 Luo et al.

Das im Jahre 2010 publizierte „Three-Column-Concept“ von Luo et al. [44] erreichte ein signifikantes Ergebnis der Intrarater-Reliabilität von $\kappa=0,848$. Es lag somit als einzige Klassifikation im Bereich „(almost) perfect“ nach Landis und Koch. In den jeweiligen Gruppen der Fraktur-Größen war die Intrarater-Reliabilität auf konstant gutem Niveau. Die subjektive Bewertungssicherheit als Maß der Komplexität in der Anwendung bzw. der Erlernbarkeit erreichte in der mit 4 Kategorien und einfachen Bewertungsregeln publizierten Klassifikation von Luo et al. ebenfalls Höchstwerte in beiden Durchläufen. Eine OP-Strategie scheint anhand der fehlenden Präzision in Bezug auf die Anatomie des genauen Frakturverlaufs und der beteiligten Fragmente im Vergleich zu den anderen Systemen schwierig ableitbar.

Die Aussagen zur klinischen und operativen Entscheidungsfindung stehen im Widerspruch zur damaligen Erstpublikation von Luo et al.. Sie hatten damals die Klassifikation als entschiedene Verbesserung gegenüber der klassischen Schatzker und AO-Klassifikation bezeichnet [44]. Zhu et al. [51, 73] bestätigten 2012 Luos Aussagen. Sie publizierten 2016 und validierten 2019 trotzdem ein „updated Three-Column-Concept“, das zusätzlich zur Klassifizierung den Trauma-Mechanismus in die operative Planung miteinbezieht. Hoekstra et al.

[61] modifizierten und validierten ebenfalls die klassische Luo-Variante. Sie vergrößerten die laterale Säule nach distal und versetzten somit das sehr häufig vorkommende postero-laterale Segment in die laterale Säule (sog. „revised 3-column“). Ihr Ziel war es, die Indikation eines lateralen Zugangs zu erweitern. Martinez-Rondanelli et al. [60] und Anwar et al. [75] beschrieben in ihren Arbeiten die Einteilung des Tibiaplateaus in 4 Säulen. Die Informationsbreite und Präzision der 3-Säulen-Methode scheint heute für die Dokumentation, Kommunikation und Ableitung einer OP-Strategie nicht mehr ausreichend [21]. Bis heute sind konkurrierende, speziell für die 3D-Bildgebung und das chirurgische Vorgehen entwickelte Systeme publiziert worden. Im Hinblick auf Präzision und Informationssumme sind sie womöglich angesichts ihrer größeren Anzahl an Kategorien und den damit verbundenen erweiterten Differenzierungs- und Beschreibungsmöglichkeiten dem „Three-Column-Concept“ aus dem Jahre 2010 überlegen. Luos Aussagen waren damals, soweit in der Literatur beschrieben, korrekt, müssen jedoch Stand heute, aufgrund neuer Entwicklungen, hinterfragt werden.

Die Intrarater-Reliabilitäten der Luo-Klassifikation waren in vergleichbaren Arbeiten ebenfalls hoch (Patange et al. | $\kappa=0,874$ [65]; Taşkesem et al. | $\kappa=0,67-0,84$ [59]; Hoekstra et al. „revised“ | $\kappa=0,782$ [61]; Zhang et al. | $\kappa=0,739$ [51]). Luos Klassifikation war in allen Publikationen der klassischen, röntgenbasierten Schatzker-Klassifikation überlegen. Millar et al. [21] konnten dies ebenfalls in seinem Review bestätigen. Ihre Arbeit kam zu ähnlichen Schlussfolgerungen wie die vorliegende Studie. Ein einfaches Klassifikationssystem mit wenigen Kategorien würde zwar eine gute Reliabilität aufweisen, habe allerdings den Nachteil einer geringeren Informationsübermittlung hinsichtlich Frakturmorphologie oder Begleitverletzungen. Dhillon et al. [70] bezeichneten ihr modifiziertes 4-Säulen-Luo-Konzept hingegen in ihrem Kritikscheiben an Krause als einfaches, biomechanisch funktionierendes und kostensparendes System. Im gleichen Paper veröffentlichten sie eine klare, chirurgische Strategie-Empfehlung. Krause et al. [82] widersprachen wenig später deutlich und zeigten in ihrer Antwort beispielhafte Schwächen des Dhillon-Systems auf, die sie bei eigener Anwendung des „Three-Column-Concepts“ erlebt hatten. Sie kritisierten

das von Dhillon et al. [70] postulierte chirurgische Fokussieren auf das Wiederherstellen der im Zentrum der Hauptfraktur stehenden Säulen. Kleinere Fragmente im Umfeld der Hauptfrakturen oder in der Peripherie, ohne direkten Bezug zur Hauptfraktur, würden von Dhillon et al. [70] nicht konsequent beachtet werden. Dieses Vorgehen würde laut Krause et al. [82] ein immenses Risiko für postoperative Kniearthrosen mit eventuellem Prothesenersatz mit sich bringen. Zhang et al. [51] untersuchten als einzige Autoren die Intrarater-Reliabilitäten in Abhängigkeit von der Anzahl an beteiligten Säulen. Die Reliabilitäten ließen marginal von einer beteiligten Säule ($\kappa=0,801$), über zwei beteiligte Säulen ($\kappa=0,770$) bis hin zur Beteiligung aller drei Säulen ($\kappa=0,708$) nach. Die Intrarater-Reliabilität blieb jedoch auch bei größeren Frakturtypen hoch. Unsere Auswertung der Frakturen nach Säulen-Beteiligung brachte aufgrund der hohen Übereinstimmung der Klassifizierungs-Ergebnisse aus erstem und zweitem Durchlauf keine auswertbaren kappa-Werte.

Percent-agreements zum „Three-Column-Concept“ wurden in keiner bisher veröffentlichten Arbeit untersucht. Somit können die Werte der vorliegenden Arbeit nicht mit anderen Arbeiten verglichen werden.

Die Klassifikation bewertet die Frakturmorphologie auf dem Tibiaplateau. Die Frakturmorphologie im Bereich unter der Gelenkfläche, welche für Frakturausläufer und demnach auch für die Versorgungsstrategie relevant ist, wird im „Three-Column-Concept“ nicht berücksichtigt. Dies ist im Vergleich zu den anderen untersuchten Klassifikationen (Kfuri/Schatzker und AO) eine Limitation der Klassifikation. Häufig verlaufen Frakturlinien nicht senkrecht nach unten durch den Tibiakopf, sondern schlagen nach distal eine schräge bis horizontale Richtung ein. Im Verlauf können sie sich multipel teilen und Fragmente bilden. Die dreidimensionale Frakturarchitektur wird durch einen solitären Blick auf das Tibiaplateau im Rahmen der Luo-Klassifikation nicht erfasst. Es fehlt demzufolge die Beschreibung von Frakturmustern (Spalt-, Impressions-, Kombinationsfrakturen), von Extensionen und Austrittspunkten in die Meta- und Diaphyse sowie eine Dokumentation der Anzahl und der Lage von möglichen Fragmenten im distalen Bereich. Die Beschreibung von mechanischen Informationen wie der Frakturstabilität sowie die Dokumentation

von wichtigen Weichgewebspathologien sind ebenfalls nicht möglich [21]. Martinez-Rondanelli et al. [60] diskutieren in ihrer Arbeit übereinstimmende Kritikpunkte. Für eine komplette Bewertung der Fraktur empfehlen sie abschließend die zusätzliche Verwendung einer klassischen Klassifikation. Zusammengefasst zeigte die Luo-Klassifikation die höchsten Werte bzgl. Reliabilität, percent-agreement und subjektive Bewertungssicherheit. Luos Klassifikation kann des Weiteren im klinisch wichtigen Bereich „Zeitfaktor“ überzeugen. Die Kategorisierung der Frakturen gelingt aufgrund der einfachen Regeln und der Beschränkung auf einen Blick auf das Tibiaplateau im Vergleich zügig. In Anbetracht der mangelnden Präzision und der daraus entstehenden unbefriedigenden klinisch- und chirurgischen Planungsmöglichkeiten muss die hohe Reliabilität im Gesamtkontext jedoch relativiert werden.

4.3.2 Krause et al.

Die ermittelte Intrarater-Reliabilität der „10-Segment-Klassifikation“ von Krause et al. [8] lag in der vorliegenden Studie bei $\kappa=0,621$. Im Vergleich mit den anderen Klassifikationen war dieser Wert am niedrigsten. Er liegt dennoch im Bereich „substantial“ nach Landis und Koch (zweithöchste von 6 Stufen). In den Subgruppen schnitt die Bewertung kleiner Frakturen mit 1 bis 3 Segmenten am besten ab ($\kappa=0,585$). Große ($\kappa=0,482$) und mittlere Frakturen ($\kappa=0,411$) erzielten geringere Reliabilitäten. Mit einem percent-agreement von 64,5 erzielte die Klassifikation von Krause et al. ebenfalls den niedrigsten Wert im Vergleich. Vergleicht man die percent-agreements im Hinblick auf die definierten Frakturgrößen-Gruppen, zeigte die „10-Segment-Klassifikation“ lediglich im Teilbereich der Trümmerfrakturen gute Werte (92,3%). Kleine, mittlere und große Frakturen erreichten im Vergleich zu den anderen Klassifikationen dieser Arbeit niedrigere Reliabilitäten und percent-agreements. Vor allem bei mittelgroßen (4 bis 6 Segmente) und großen (7 bis 9 Segmente) Frakturen wurde jedes zweite CT im zweiten Umlauf abweichend bewertet. Das gute Ergebnis im Bereich der „Trümmerfrakturen“ sollte mit Vorsicht betrachtet werden. Die Befundung eines durchweg in multiple, kleine Fragmente zertrümmerten Tibiakopfes und die Zuordnung zur „Maximalstufe“ (alle bzw. 10 Segmente beteiligt) gelingt einfacher

als die Differenzierung zwischen 6-, 7-, 8- oder gar 9-Fragment-Frakturen. Demzufolge ist eine Interpretation des percent-agreements der Klassifikation im Bereich der Trümmerfrakturen schwierig. Schließlich zeigte die Klassifikation ebenfalls in den Bereichen subjektive Bewertungssicherheit und Komplexität niedrigere Werte im Vergleich mit den anderen Klassifikationen. Immerhin konnte im zweiten Bewertungsdurchlauf in der subjektiven Bewertungssicherheit im Vergleich zu den anderen Systemen der höchste Zugewinn erzielt werden (+17,8% „trifft sicher zu“). Dies könnte als vorsichtiges Indiz für eine Abhängigkeit der Qualität der Kategorisierung von der Erfahrung und Übungszeit des Untersuchers mit der Klassifikation gewertet werden. Eine präzise Frakturverlaufsbeschreibung ist eingeschränkt möglich. Mittels Aufzählung der an der Fraktur partizipierenden Segmente im Bereich der proximalen 3 cm unter der Gelenkfläche können differenzierte Aussagen über Frakturmuster und Frakturlinien nur begrenzt getroffen werden. Insbesondere die weiteren Verläufe der Frakturlinien und deren Austrittspunkte sowie die Anatomie der Fragmentierung im weiter distal liegenden Bereich des Tibiakopfes sind nicht abgebildet. Die Austrittspunkte der Frakturlinien sowie die Anzahl, das Aussehen und die Größe der Fragmente sind chirurgisch bedeutsame Informationen, um Zugangswege und die Plattenpositionierung planen zu können. Lediglich anhand der Segmentbeschreibung wird des Weiteren nicht klar, ob es sich um einen Spaltbruch, eine Impressionsfraktur oder eine Kombination von beidem handelt. Für die Dokumentation und Kommunikation fehlen die spezifischen Informationen zu Hauptfrakturen, Nebenfrakturen, dislozierten Segmenten und Trümmeranteilen. Schließlich fehlen die Informationen bezüglich der Stabilität der Fraktur (Beteiligung Meta-/Diaphyse) sowie die Beschreibung von Weichgewebspathologien. Dieser Kritik stellten sich Krause et al. in einer Stellungnahme im Jahre 2017 [82]. Sie befürworteten die Verwendung ihrer Klassifikation additiv zu und nicht anstatt klassischer Frakturklassifikationen wie z.B. dem „Three-Column-Concept“ von Luo et al. [44]. Chirurgische Planungsmöglichkeiten anhand des 10-Segment-Systems werden von Krause et al. [82] und Meulenkamp et al. erläutert [83]. Sie beschrieben das Frakturmuster als führend in Bezug auf den chirurgischen Zugang und die operative

Versorgung. Das klinische Outcome würde im Wesentlichen von der Genauigkeit der Wiederherstellung der Gelenkfläche beeinflusst. Andere Autoren wie Dhillon et al. [70] teilten diese Ansichten nicht. Sie postulierten eine schlechte Eignung des Krause-Systems für den klinischen Alltag und für die operative Planung.

Gründe für die schwachen Werte hinsichtlich der Reliabilitäten und der percent-agreements liegen womöglich in der abzählenden Kombinatorik (Mathematik) und in den komplexen Regeln für die Einschlusskriterien der Segmente. In Krauses „10-Segment-Klassifikation“ existieren insgesamt 1023 Möglichkeiten der Kategorisierung. Mit steigender Anzahl der Kategorien sinkt die Wahrscheinlichkeit für übereinstimmende Bewertungen.

Der folgende Abschnitt soll einen kurzen Einblick in die vielfältigen Kriterien der Kategorisierung in Krauses Klassifikation vermitteln. Die in der Publikation beschriebenen Bewertungskriterien beschreiben differenziert die zu bewertende vertikale Höhe von 3 cm ab Gelenkfläche. Alternativ können auch die obersten 10 CT-Schichten unter dem Tibiaplateau bewertet werden. Krause et al. [8] beschreiben in ihren Einschlusskriterien für ein Segment ein sogenanntes Volume of Interest (VOI). Für die Wertung eines Segmentes muss dieses VOI zu 50% abgesenkt sein. Zusätzlich werden nicht abgesenkte Segmente, die jedoch Teil eines größeren, dislozierten Fragments sind, in die Bewertung eingeschlossen. Ein nicht zertrümmertes Segment wird als zugehörig gewertet, wenn mindestens ein Viertel des VOI vom restlichen Segment abgetrennt ist. Trümmerfrakturen werden nach Krause et al. [8] diagnostiziert, sobald 3 oder mehr Fragmente bei Aufsicht auf die Gelenkfläche vorhanden sind.

Die Anwendung aller Regeln ist häufig trotz präziser, dreidimensionaler Bildgebung mittels hochauflösender CTs erschwert. Der Zeitaufwand ist erhöht, vor allem bei qualitativen Mängeln in der Bildgebung. Insbesondere im Bereich zwischen 4 und 9 Segmenten mit Trümmeranteilen und Extension in die Metaphyse ist eine Differenzierung erschwert. Zu diskutieren ist, inwiefern sich bei den 106 Fällen (77,4% der Fehler) mit einem unterschiedlich bewerteten Segment eine klinisch-chirurgische Konsequenz ergibt. Unterschiedliche Bewertungen in mehr als einem Segment kamen lediglich in 8% aller Fälle vor. Die Anwendbarkeit im klinischen Alltag wird aufgrund der oben genannten

Gründe diskutiert. Positiv ist die zumindest theoretische, linear aufsteigende Sortierung der Kategorien in kleinen Schritten in Bezug auf das Frakturausmaß, welche z.B. in der Klassifikation von Kfuri/Schatzker nicht zu finden ist. Hierbei sollte, wie bereits oben beschrieben, beachtet werden, dass nicht in allen Fällen das wahre Ausmaß der Fraktur mit der Menge an beteiligten Segmenten auf dem Tibiaplateau korreliert. So kann sich beispielhaft hinter einer 4- bzw. 5-Segment-Fraktur nach Krause et al. eine mediale oder laterale Impressionsfraktur nach AO (Typ 41B2.1 und 41B2.2) verbergen (12 Fälle), ebenso aber auch eine wesentlich umfangreichere bikondyläre C-Fraktur (12 Fälle). Die Beschreibung der Größe und der Lokalisation der betroffenen Gelenkfläche gehört ebenfalls zu den Stärken der 10-Segment-Klassifikation. Arbeiten zu Reliabilitäten oder percent-agreements der Klassifikation existieren nicht. In der Publikation von Krause et al. werden exemplarisch Segment-Häufigkeiten und Fraktur-Verteilungen in Bezug auf die AO-Klassifikation beschrieben. Es konnte eine Anhäufung von posterioren Segmenten in B- und C-Frakturen der AO-Klassifikation dargestellt werden [8]. Dhillon et al. [70] beschrieben in einem „Letter to the editor 2017“ 7 Kritikpunkte an Krauses Publikation und Klassifikation. Sie kommentierten u.a., dass bei der Analyse der oberen 3 cm des Tibiakopfes chirurgisch wichtige Informationen zur distalen Spitze eines Fragments fehlen würden. Die Lage dieses Apex sei entscheidend für den Zugang und die Wahl des richtigen Osteosynthesematerials. Sie bezeichneten die Aufteilung des Plateaus in 10 Segmente als „schwerfällig“ und untauglich für klinische Szenarien.

Die rege Diskussion der Klassifikation zeigt, dass die Kombination aus 1023 Differenzierungsmöglichkeiten und komplexer Einschlusskriterien lediglich einen fraglichen klinisch-therapeutischen Vorteil mit sich bringt. Unbestritten hingegen ist Krauses elementare Darstellung der Wichtigkeit posteriorer Segmente bei Frakturen der proximalen Tibia mit Gelenkbeteiligung.

4.3.3 Kfuri und Schatzker

Die überarbeitete klassische Schatzker-Klassifikation von Kfuri/Schatzker aus dem Jahre 2018 zeigte in der vorliegenden Studie eine höhere Intrarater-Reliabilität (percent-agreement von 81,1 $\kappa=0,790$) als die AO-Klassifikation

($\kappa=0,756$). Die Intraraterreliabilität war für alle Frakturgrößen vergleichbar. Im Bereich der mittelgroßen Frakturen (84,6 | $\kappa=0,724$) zeigte die Klassifikation den höchsten Wert im Vergleich der Klassifikationen. Insgesamt wurden für die Kfuri/Schatzker-Klassifikation gute und v.a. die stabilsten Reliabilitätswerte und percent-agreements hinsichtlich der definierten „Frakturgrößen-Subgruppen“ gemessen. Bzgl. Komplexität und Bewertungssicherheit ist die Klassifikation vergleichbar mit Krause et al. [8] und Luo et al. [44]. Die verständliche und übersichtlich visualisierte, klassische Variante aus dem Jahre 1979 sollte von den Autoren gezielt und aus chirurgischer Sicht sinnvoll auf die genaue Beschreibung des Frakturverlaufs und der beteiligten Gelenkflächen-Segmente modifiziert werden. Anatomische Hilfsmittel wie die Eminentia intercondylaris und die Definition eines Plateau-Äquators erleichtern das systematische und schrittweise Vorgehen zur Definition des Typs und der beteiligten Segmente. Das gezielt auf die operative Therapie ausgerichtete Konzept sollte das Verständnis für den Chirurgen erleichtern. Die Mischung aus Frakturmustern (Typen, übernommen aus der klassischen Variante) und einem „4-Säulen-Konzept“ (Subtypen, Frakturverlauf auf dem Plateau) kann eine symbiotische Beschreibung der Gelenkflächenbeteiligung und des Verlaufs der Muster nach distal hin zu Meta- und Diaphyse inklusive Beschreibung der Lage des chirurgisch wichtigen Fragment-Apex (Zugangsweg und Plattenpositionierung) ermöglichen. Die in der Literatur von unterschiedlichen Autoren beschriebenen Defizite der klassischen Schatzker-Klassifikation in Bezug auf die schwierige Beschreibung posteriorer Frakturanteile sind durch die Neuauflage weitestgehend eliminiert worden [21, 49, 84, 85].

Die Kfuri/Schatzker-Klassifikation staffelt die Frakturtypen nicht stringent nach der Frakturgröße. Ursächlich ist die primäre Typisierung der Fraktur nach einem bestimmten Frakturmuster. Sind die bikondylären Frakturen V und VI nahezu immer größer und klinisch komplexer als die monokondylären, so sind Typ III-Frakturen (monokondyläre, laterale Impressionsfrakturen) häufig kleiner als die Typen I (monokondyläre, laterale Spaltfrakturen) und II (monokondyläre, laterale Spalt-Impressionsfrakturen). Drei der vier monokondylären Typen beschreiben laterale Frakturen. Im Typ IV (monokondyläre mediale Plateaufrakturen) sind alle

Typen I bis III für die mediale Gelenkfläche subsummiert. Eine weitere, präzisere Differenzierung medialer Frakturen ist nicht möglich. In unserer Studie zeigte sich dies eindrucksvoll in deutlich höheren percent-agreement-Werten (92,6) und kappa-Werten ($\kappa=0,856$) der Typ IV-Frakturen im Vergleich zu den anderen Fraktur-Typen. Diese höheren Messwerte gehen mit Einbußen in Bezug auf die Präzision einher. Ähnliche Zusammenhänge werden in dieser Arbeit bei der Luo-Klassifikation beschrieben. Die Kfuri/Schatzker-Klassifikation zeigt folglich in der Beschreibung medialer Anteile Schwächen. Eine weitere Differenzierung dieser medialen Frakturareale ist für das chirurgische Vorgehen zum derzeitigen Wissensstand nicht von Relevanz.

Bei Abweichung der ersten und zweiten Messung zeigte sich am häufigsten ein Wechsel zwischen Frakturtyp II und III (25%). Eine Differenzierung zwischen lateraler Impressionsfraktur (Typ III) und lateraler Impressions-Spaltfraktur (Typ II) ist oft schwierig, da viele Impressionsfrakturen kleine Spaltausläufer nach distal haben. Diese verlaufen jedoch häufig multipel, zeigen keinen klaren Apex und sind anterior-posterior nicht durchgängig.

Die fallspezifische Definition des „virtuellen Äquators“ gelingt in vielen Fällen nur mühevoll. Lateral liegt der Äquator tangential dem Fibulaköpfchen vorne auf. Das Köpfchen erstreckt sich in kranio-kaudaler Richtung häufig nicht bis zur Gelenkfläche und ist in der coronaren Schichtung beim Einstellen der Gelenkfläche nicht zu sehen. Dies erschwert die Orientierung. In dieser Arbeit zeigte sich dieses Problem in einem für die monokondylären Typen überproportional häufigen Subtypen-Wechsel von Typ II-A zu Typ II-A-P. Die Trennlinie zwischen medialen und lateralen Segmenten in Form einer anatomischen Struktur direkt auf der tibialen Gelenkfläche (Eminentia intercondylaris) scheint leichter zu finden zu sein. 9 von 10 Subtypen-Wechsel betrafen Frakturen mit Beteiligung von 3 oder mehr Segmenten. Die Entscheidung welches Segment die Einschlusskriterien erfüllt, welche Frakturverläufe zur Hauptfraktur gehören (Durchbruch des kortikalen Plateau-Rings) und welche zu Nebenfrakturen gezählt werden, gelingt in den Frakturgruppen mit definitionsgemäß vier zu differenzierenden Quadranten (Typ V und VI) erwartungsgemäß schwieriger.

Arbeiten zu Reliabilitäten und percent-agreements der Klassifikation von Kfuri/Schatzker sind bisher nicht publiziert. Somit ist ein Vergleich mit anderen Studien nicht möglich. Das klassische System von Schatzker, welches auf Röntgenbildern basiert, ist hingegen die Klassifikation, die am häufigsten in Reliabilitätsstudien eingeschlossen wurde [21]. 16 der 17 in der Literatur zu findenden Studien beinhalten die Analyse der klassischen Schatzker-Klassifikation. Die Intrarater-Reliabilität der älteren Schatzker-Klassifikation war im konventionellen Röntgenbild vergleichbar mit dem Wert der AO-Klassifikation ($\kappa=0,57-0,91$ | $\emptyset\kappa=0,69$) [26, 40, 41, 66, 68, 69, 74, 86]. Die Werte waren bei Anwendung der alten Klassifikation in der CT-Diagnostik höher ($\kappa=0,68-0,89$ | $\kappa=0,73$) [10, 65, 67]. Sowohl in der konventionellen als auch in der CT-Diagnostik zeigte die Schatzker-Klassifikation in der Hälfte aller Reliabilitätsstudien die höchste Reliabilität im Vergleich zu anderen Klassifikationen.

Zhang et al. [51] verglichen 2019 erstmals Intrarater-Reliabilitäten von Schatzker-Fraktur-Typen und schlossen den Unfallmechanismus in die Auswertung mit ein. „High-energy fractures“ (Schatzker Typen IV bis VI) zeigten eine Intrarater-Reliabilität von $\kappa=0,721$. Bei „Low-energy fractures“ (Schatzker Typen I bis III) wurde eine etwas höhere Intrarater-Reliabilität von $\kappa=0,832$ gemessen. Diese Ergebnisse decken sich nicht mit den Ergebnissen dieser Studie. Der Wert für die „High-energy fractures“ (Schatzker Typen IV bis VI) lag u.a. aufgrund eines nahezu perfekten Messergebnisses des Frakturtyps IV bei $\kappa=0,725$. Die Intrarater-Reliabilität der „Low-energy fractures“ (Schatzker Typen I bis III) in dieser Arbeit lag bei $\kappa=0,550$. In der Arbeit von Zhang et al. [51] wurde die klassische Schatzker-Variante aus dem Jahre 1979 untersucht.

Kfuri/Schatzker [27] beschreiben in ihrer Publikation eine zusätzliche, genauere Frakturkeil-Charakterisierung mittels Kleinbuchstaben a, p, a_x und p_x . Die Kleinbuchstaben a und p charakterisieren die Lage der Eckpunkte eines Hauptfragments anterior oder posterior des Äquators auf dem Tibiaplateau. Die Variablen a_x und p_x definieren die Lage der Fragmentspitze im metaphysären Bereich. Diese Beschreibung ist nicht für alle Frakturtypen möglich (z.B. Typ III). Sie stößt des Weiteren in den größeren Frakturkategorien mit vielen und kleineren Fragmenten an ihre Grenzen. Die zusätzliche Differenzierung führt zu

einer erheblichen Erweiterung der Anzahl an Fraktur-Kategorien, die eine vergleichende Reliabilitäts-Studie nicht mehr möglich machen würde. Die Erweiterung kann, vor allem bei Frakturen mit gut differenzierbaren, großen Fragmentkeilen, hilfreich sein. Eine generelle Anwendung bei allen Frakturtypen kann hingegen nicht empfohlen werden.

4.3.4 AO-Klassifikation

Die AO-Klassifikation zeigte mit einer Intrarater-Reliabilität von $\kappa=0,756$ und einem percent-agreement von 79,9 vergleichbare Werte wie die Kfuri/Schatzker-Klassifikation. Die gemessenen Ergebnisse lagen folglich ebenfalls im Bereich „substantial“ nach Landis und Koch. Vor allem im Bereich der kleinen Frakturen (83,7 | $\kappa=0,739$) erzielte die Klassifikation die höchsten Werte im Vergleich zu den anderen Klassifikationen. In den größeren Fraktur-Subgruppen zeigte die AO-Klassifikation geringere Werte sowohl hinsichtlich der Intrarater-Reliabilität als auch im percent-agreement.

Spezifische Regeln oder eine detaillierte Vorgehensweise beim Klassifizieren einer Fraktur wie in den Klassifikationen von Krause et al. [8] und Kfuri/Schatzker et al. [27] sind in der klassischen Publikation nicht beschrieben. Vielmehr wird die zu klassifizierende Fraktur einem kurz umschriebenen, bildlich dargestellten Typ bzw. Subtyp zugewiesen. 2017 wurde die AO-Klassifikation modernisiert und um eine Quadranten-Einteilung analog zur Kfuri/Schatzker-Klassifikation erweitert. Kleinbuchstaben (bis zu 8 in der C3-Kategorie) sollen die Frakturlokalisierung und den Charakter näher beschreiben. Diese Erweiterung hat aufgrund der Komplexität und des erhöhten Zeitaufwandes noch keinen Zugang in den klinischen Alltag gefunden. Eine aussagekräftige Reliabilitätsstudie wäre hinsichtlich der extrem großen Anzahl verschiedener Fraktur-Kategorien nur mit sehr hohen Fallzahlen möglich.

Die visuelle Darstellung der 20 Frakturmuster ist subjektiv übersichtlich und verständlich. Die Differenzierung zwischen 41C2- und 41C3-Frakturen scheint aufgrund multipler Frakturlinien mit hoher Ähnlichkeit initial schwierig. Die Ergebnisse der Messung der subjektiven Bewertungssicherheit waren vergleichbar mit der Klassifikation von Kfuri/Schatzker. Beide weisen eine

ähnliche Klassifikations-Struktur (Typen, Subtypen, bildliche Darstellung) und eine vergleichbare Anzahl an Zuteilungsmöglichkeiten auf. Die primäre Unterteilung in mono- und bikondylär (41B bzw. 41C) mit sukzessiver Steigerung der Frakturschwere in den Subgruppen (Ausnahme Subgruppe B2 „laterale und mediale Impressionsfrakturen“) ist für den Nutzer nachvollziehbar. Im Vergleich zur Schatzker-Klassifikation sind in der AO-Klassifikation für die laterale und die mediale Kondyle alle Frakturmuster zugänglich. Eine relativ genaue Frakturverlaufsbeschreibung kann gelingen. Zur exakten Beschreibung der Lokalisation und des Charakters muss zusätzlich die Quadranteneinteilung herangezogen werden. Der Klassifikations-Code kann hierbei unübersichtlich lang werden. Für eine beispielhaft aufgeführte Fraktur 41C3.3estuvwx kann die Interpretation im klinischen Austausch die Zuhilfenahme der Publikation nötig machen. Die Ableitung einer OP-Strategie kann gelingen. Im Vergleich zu den Klassifikationen von Luo et al. und Krause et al. enthält sie mehr chirurgisch wertvolle Informationen. Die genaue Beschreibung der Lokalisation und des Charakters kann über die oben beschriebenen Kleinbuchstaben erfolgen, die zusätzlich zu den 20 Typen und Subtypen verwendet werden. Die Kfuri/Schatzker-Klassifikation scheint in der Informationsübermittlung schlanker und übersichtlicher. Bei vergleichbarer Präzision beschränkt sich die Kfuri/Schatzker-Klassifikation auf 34 Kategorien.

In der AO-Klassifikation wurden im Vergleich zur Klassifikation von Kfuri/Schatzker lediglich halb so viele Typen-Wechsel gemessen. Die AO-Klassifikation scheint demzufolge weniger anfällig für Typen-Wechsel zu sein als die Klassifikation von Kfuri/Schatzker. Vermutlich ist das Zusammenfassen zu lediglich 3 Fraktur-Typen (A-, B- und C-Frakturen) in der AO-Klassifikation dafür verantwortlich (vgl. Schatzker Frakturtypen I-VI).

Gründe für die niedrigen kappa-Werte und percent-agreements in den großen Frakturtyp-Subgruppen könnten in den subjektiv schwer differenzierbaren Subtypen der 41C-Fraktursubtypen (bikondylär) liegen (vgl. Abbildung 10).



Abbildung 10: Verschiedene Subtypen (3 von insg. 8) der C-Frakturen in der AO-Klassifikation

41C-Frakturen erreichten ein percent-agreement von lediglich 69,0 (vgl. 41B-Frakturen mit 82,0). 41C2 und 41C3-Frakturen ähneln sich in ihrer Morphologie. Die Subtypen sind noch komplexer zu bestimmen. Hierbei muss die Anzahl und die Lage der Segmente sowie die Differenzierung zwischen mehrfragmentär und zertrümmert zielgenau für beide Kondylen erfolgen. Beispielhaft sei hier der Subtyp 41C2.3 (bikondyläre Fraktur mit multifragmentärer metaphysärer Fraktur) genannt, der in 63,6% keine Übereinstimmung im re-test zeigte. Mehr als jeder dritte Subtypen-Wechsel bezog sich auf den Übergang von 41C3.1 (bikondyläre Fraktur mit fragmentiertem lateralem Plateau) zu 41C3.3 (bikondyläre Fraktur mit fragmentiertem lateralem und medialem Plateau) (vgl. Abbildung 11).

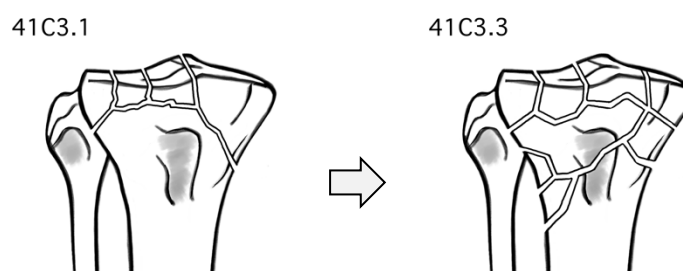


Abbildung 11: Häufigster Subtypenwechsel in der AO-Klassifikation

Die Notwendigkeit der komplexen Kategorisierung der C-Frakturen sollte hinsichtlich der Übermittlung klinischer und OP-strategisch wichtiger Informationen hinterfragt werden.

Ein weiterer Schwachpunkt der AO-Klassifikation scheinen die monokondylären Fraktur-Subtypen mit schrägem Frakturverlauf unter Einbeziehung der Eminentia intercondylaris zu sein. Insbesondere die Kategorien 41B3.3h (monokondyläre mediale Impressions-Spalt-Fraktur mit Einbeziehung der Eminentia intercondylaris) und ihr laterales Korrelat 41B3.3f zeigten für 41B-Frakturen ($\kappa=0,82,0$) verhältnismäßig niedrige Messwerte im percent-agreement (41B3.3h mit 25,0; 41B3.3f mit 38,2). Hierbei scheint die Grenze zur monokondylären Fraktur ohne Eminentia-Beteiligung schwierig festzulegen zu sein. Eine explizite anatomische Grenze wo die Eminentia beginnt bzw. ab welcher Größe der Beteiligung die Eminentia in die Klassifizierung miteinbezogen wird ist in der Publikation weder beschrieben noch illustriert. Gleiches gilt in Bezug auf die Grenze zu bikondylären 41C-Frakturen (vgl. Abbildung 12).

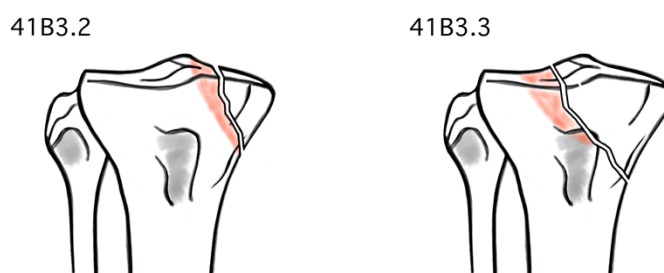


Abbildung 12: AO 41B3.2 und 41B3.3f ohne bzw. mit Einbeziehung der E. intercondylaris

Die Klassifikation von AO ist vergleichbar häufig Inhalt von Reliabilitätsstudien [21]. Intrarater-Reliabilitäts-Analysen mittels CT-Diagnostik wurden in sieben Arbeiten beschrieben [10, 40, 58-60]. Variierende Ergebnisse ($\kappa=0,36-0,89$) werden postuliert. Die Fallzahlen der veröffentlichten Studien betragen 15 bis 115 Fälle. Durchschnittlich zeigte die AO-Klassifikation eine Intrarater-Reliabilität von $\kappa=0,65$. Der Messwert der vorliegenden Arbeit ($\kappa=0,756$) ist vergleichbar mit den Ergebnissen der Studie von Wennergren et al. [72] ($\kappa=0,74-0,79$). Diese zeigte nach Millar et al. [21] die beste Methodik unter den veröffentlichten Reliabilitätsstudien zum Tibiakopf und beinhaltete die größte Fallzahl (n=117).

Wennergren et al. ermittelte als einziger Autor percent-agreement-Werte und differenzierte in der Reliabilitäts-Analyse nach den Subgruppen „AO-Type“ (65,0-68,0 | $\kappa=0,74-0,79$) und „AO Group“ (79,0-83,0 | $\kappa=0,62-0,64$). Ein sinnhafter Vergleich mit unserer Arbeit ist nicht möglich, da diese die AO-Typen 41 A-C, 42 A-C und 43 A-C und somit alle Frakturentitäten der Tibia (auch diejenigen ohne Gelenkbeteiligung) für die Analyse einschlossen. Darüber hinaus wurde lediglich in 54 Fällen ein CT befundet.

4.3.5 Schlussfolgerungen

Im folgenden Abschnitt sollen die Messergebnisse der jeweiligen Klassifikationen selbst, deren Interpretation und Diskussion sowie die Einordnung in bereits bestehende Literatur im Gesamtkontext der Intention der vorliegenden Arbeit beschrieben werden.

Die Klassifikationen von Luo et al. [44] und Krause et al. [8] sind aufgrund der o.g. Schwächen hinsichtlich der Entwicklung einer Anwendungsempfehlung eingeschränkt empfehlenswert. Es sollten Modifikationen zur detaillierteren Beschreibung der Art, der Lokalisation sowie der Ausdehnung der Frakturen diskutiert und entwickelt werden. Zhu et al. [73, 86], Dhillon et al. [70], Hoekstra et al. [61], Anwar et al. [75] und Martinez-Rondanelli et al. [60] erkannten bei eigener Anwendung früh die Informationsdefizite der Luo-Klassifikation und modifizierten das System auf unterschiedlichste Weise (zusätzliche Säule, Einschluss des Traumamechanismus). Diese Modifikationen konnten sich im klinischen Alltag nicht etablieren. Krause et al. und Dhillon et al. führten eine öffentliche Diskussion [82]. In der Arbeit von Millar et al. [21] sind vergleichbare Kritikpunkte hinsichtlich Luos Klassifikation beschrieben. Die 2010 von Luo et al. [44] publizierte Klassifikation war die erste Klassifikation, die für die dreidimensionale Bildgebung entwickelt wurde und in besonderem Maße die posteriore Säule respektiert. Luo et al. betonten die chirurgisch, praktische Orientierung ihrer Klassifikation. Letztlich scheint der Informationsgehalt des „Three-Column-Concept“ im Vergleich zu den speziell für die CT-Bildgebung weiterentwickelten Klassifikationen, u.a. von Kfuri/Schatzker, für eine fachgerechte Kommunikation und für die Ableitung einer therapeutischen

Strategie zu gering. Veröffentlichte Modifikationen des klassischen „Three-Column-Concepts“ kritischer Autoren konnten sich nicht etablieren.

Die Klassifikation von Krause et al. [8] wirkt aufgrund der komplexen Regeln und den vielen Möglichkeiten der Kategorisierung akademisch. Die „10-Segment-Klassifikation“ zeigte in der vorliegenden Arbeit die niedrigsten Reliabilitätswerte und percent-agreements. Arbeiten zur Reliabilitätsanalyse fehlen bisher.

Die Klassifikationen von Kfuri/Schatzker und AO könnten bei der Entwicklung eines Goldstandards für die Klassifikation von Frakturen der proximalen Tibia mit Gelenkbeteiligung eine Rolle spielen.

Die Basis der Klassifikation von Kfuri/Schatzker bildet eine spezifisch auf die aktuelle 3D-Diagnostik weiterentwickelte Kategorisierung mittels Quadranteneinteilung des Tibiaplateaus. Diese Differenzierungsmöglichkeiten lassen eine relativ präzise Beschreibung des Frakturverlaufs zu. Im Zusammenspiel mit den übernommenen klassischen Frakturtypen und Subtypen (Beschreibung des Frakturcharakters) scheint eine übersichtliche, verständliche und dennoch detaillierte Beschreibung der Pathologie zu gelingen, welche eine für chirurgische Techniken und Zugangswege ausreichende Informationsmenge beinhaltet. Die in der vorliegenden Arbeit gemessenen Reliabilitäten der Klassifikation sind trotz deutlich mehr Differenzierungsmöglichkeiten (34) im Vergleich zur Luo-Klassifikationen (4) über alle Frakturgrößen im Bereich „substantial“ nach Landis und Koch (zweithöchste Kategorie). Die wenigen Schwächen des Systems scheinen in einer oft schwierigen Definition des Äquators aufgrund des tiefer stehenden Fibulaköpfchens, einer fehlenden Differenzierungsmöglichkeit von medialen Frakturen, einer fehlenden linearen Staffelung nach Frakturgröße und einer komplexen und klinisch fragwürdig anwendbaren Strategie zur genauen Beschreibung der Frakturlinien und -keile mittels weiterer Kleinbuchstaben zu liegen. Der beschriebene Klassifizierungs-Workflow der Kfuri/Schatzker-Klassifikation wirkt durchdacht und lässt sich rasch erlernen. Der Zeitaufwand zur Kategorisierung einer Fraktur ist im Vergleich zu den anderen Klassifikationen der vorliegenden Arbeit mittelmäßig.

Die AO-Klassifikation ist keine spezifisch für Frakturen der proximalen Tibia entwickelte Klassifikation. Die übergeordnete AO-Klassifikationsstrategie scheint

auch am Tibiakopf eine reliable Methode zu sein. Drei definierte Frakturtypen werden hierbei von mehreren Frakturtypuntergruppen ergänzt. Die Frakturgruppen definieren die Seite der Fraktur, die Seite des zertrümmerten Bereiches und/oder häufige und spezifische Frakturmorphologien. Die Beschreibung des Frakturverlaufs scheint hingegen nicht so präzise wie in der Kfuri/Schatzker-Klassifikation zu sein. Mit Hilfe der weiter oben beschriebenen, neuen und klinisch komplexen Quadranteneinteilung kann sie dennoch gelingen. Der Zeitaufwand kann hierbei entsprechend höher sein. Die AO-Klassifikation zeigte in der vorliegenden Arbeit beim Vergleich der nach Größe sortierten Subgruppen niedrigere Werte als die Kfuri/Schatzker-Klassifikation. Vor allem im Bereich der bikondylären C-Frakturen zeigte sich nur eine moderate Reliabilität. Womöglich erschwert die Anzahl ähnlicher, multifragmentärer Subgruppen eine sichere Einordnung. Der Klassifizierungs-Workflow gelingt im Allgemeinen mit Hilfe des hauptsächlich bildlichen Regelwerks gut. Das System scheint, womöglich ebenfalls aufgrund der übersichtlichen Illustrationen, etwas leichter erlernbar zu sein als die Kfuri/Schatzker-Klassifikation. Der Zeitaufwand der Klassifizierung ist vergleichbar mit der Klassifikation von Kfuri/Schatzker.

4.4 Ausblick

Die Konzeption und Entwicklung einer Anwendungs-Empfehlung bzw. eines Goldstandards kann, wie oben beschrieben, nicht durch eine singuläre Arbeit dargestellt werden. Vielmehr ist dieser Prozess mehrstufig. Bisher publizierte Reliabilitätsstudien zu Klassifikationen von Frakturen der proximalen Tibia mit Gelenkbeteiligung konnten laut Audigé et al. [29] und Millar et al. [21] nicht überzeugen. Die genannten Autoren sowie Maripuri et al. [68] forderten die Entwicklung eines Validierungs-Konzeptes für eine zukünftige, evidenzbasierte Klassifikations-Entwicklung.

Audigé et al. [13] publizierten ein Jahr später im Jahre 2005 eine 3-Phasen-Validierung speziell für die Erarbeitung eines Goldstandards im Bereich von Frakturklassifikationen. In Schritt 1 sollten bei der Entwicklung gänzlich neuer Systeme die Klassifikationskategorien definiert und mittels Reliabilitätsstudien

auf Zuverlässigkeit und Genauigkeit überprüft werden. Bei der Validierung bereits bestehender Systeme stehen die Evaluation hinsichtlich Reliabilitäten und die Suche nach bestehenden Schwächen im Zentrum der wissenschaftlichen Tätigkeit. In Schritt 2 könnte eine multizentrische Übereinstimmungsstudie mit Hilfe einer repräsentativen Gruppe zukünftiger Benutzer der Klassifikation durchgeführt werden. Werden bereits bestehende Systeme validiert, sollte der Fokus auf einer Auswahl von ein bis zwei Systemen liegen, die im ersten Schritt überzeugen konnten und deren Schwächen eventuell bereits bearbeitet wurden. Letztendlich kann in Schritt 3 der Klassifizierungsvorschlag im Rahmen einer prospektiven klinischen Studie angewendet werden. Mit Hilfe dieser Arbeit kann der praktische Nutzen bewertet und die wichtige „construct validity“ fixiert werden. Die „construct validity“ bezeichnet die Verknüpfung einer Frakturkategorie des Klassifizierungssystems mit einem spezifischen Therapie-Plan und einer entsprechenden Prognose für den Patienten (Beweglichkeit, quality of life, Rezidiv, Risiko einer Arthrose bzw. eines Gelenkersatzes).

Für den beschriebenen, mehrstufigen Entwicklungsprozess muss in Schritt 1 eine Basis gelegt werden. Im Zentrum der Arbeit mit bestehenden und verbreitet im klinischen Alltag angewandten Klassifikationen steht die Feststellung und Bewertung aktuell gebräuchlicher und in der Literatur häufig zu findender Klassifikationssysteme (\approx status quo). Eine weitere Maßnahme zur Grundsteinlegung besteht in der Sichtung und Auswahl eines geeigneten Studienkollektivs. Die über den Inhalt der Studie hinausgehenden und somit zusätzlich gesammelten Informationen können für den weiteren Prozess (Schritte 1-3) verwendet werden. In der vorliegenden Arbeit wurden zusätzlich wichtige Daten zur Biometrie des Patienten, zum Unfallhergang, zu Begleitverletzungen, OP-Techniken, Zugangswegen und Osteosynthesematerial gesammelt und dokumentiert. Diese Daten könnten beispielsweise bei der Entwicklung eines CT-OP-Strategie-Algorithmus hilfreich sein. Wang et al. [73] eruierte einen solchen Algorithmus exemplarisch in Abhängigkeit vom Traumamechanismus. Dhillon et al. [70] erstellten einen Workflow anhand verschiedener betroffener Säulen. Rossmann et al. [87] versuchten eine Beziehung zwischen Fraktur-Klassifizierung und operativem Zugang herzustellen. Zhang et al. [51] versuchten

anhand der Fraktur-Morphologie Rückschlüsse auf den Traumamechanismus zu ziehen. Sie stellten den Unfall-Hergang und die einwirkenden Kräfte ins Zentrum der therapeutischen Herangehensweise.

Der erste Schritt im Entwicklungsprozess (u.a. Reliabilitäten, Qualitäten, Stärken und Schwächen bestehender Klassifikationen) ist durch diese Studie nicht abgeschlossen. Die vorliegende Arbeit sollte durch eine Studie zur Interrater-Reliabilität der vier evaluierten Klassifikations-Systeme ergänzt werden. Hierzu kann, wie erwähnt, das generierte Studienkollektiv mit entsprechendem Datensatz Verwendung finden. Im Anschluss sollte zielführend überlegt werden, ob und wenn ja, wie die dargestellten Schwächen der Klassifikationen behoben werden könnten. Beispielhaft soll hier das Hinzufügen von Weichgewebspathologien erwähnt werden, wie es von Carlson [88] und Tscherne [49] vorgeschlagen wurde. Wahlquist et al. [89] revidierten in ihrer Publikation die Schwäche des nicht weiter differenzierbaren Schatzker-Typ-IV, indem sie die mediale Säule weiter unterteilten. Ähnliche Empfehlungen hinsichtlich der Kfuri/Schatzker-Klassifikation und speziell für den Typ IV der Klassifikation können anhand der in der vorliegenden Arbeit gemessenen Ergebnisse ebenfalls gegeben werden. Ein weiterer Gedanke zur Kfuri/Schatzker-Klassifikation wäre die konsequent linear aufsteigende Staffelung der Kategorien hinsichtlich der Frakturgröße. In der AO-Klassifikation könnte indes erwogen werden, die Subgruppen der 41C-Frakturen sinnvoll und klinisch prägnant zu verschlanken. Selbstverständlich muss ein neues oder verändertes System erneut verbindlich auf Reliabilität und Validität überprüft werden, bevor in Schritt 2 eine multizentrische Übereinstimmungsstudie in Angriff genommen werden kann.

Eine attraktive Überlegung ist die Entwicklung einer Software, die mit Hilfe von künstlicher Intelligenz (KI) Computertomographien auswertet und eine digitale Klassifizierung möglich machen kann. Der Beginn der Programmierung einer Software kann hierbei bestenfalls zeitgleich mit dem Prozess der Erarbeitung einer Anwendungsempfehlung stattfinden. Eventuell sollte sogar die Abstimmungsmöglichkeit mit einer zukünftigen Software und die technische Umsetzbarkeit der KI-Anwendung in den Entwicklungsprozess eines modernen

Klassifikationssystem integriert werden. Suero et al. [90] publizierten 2010 bisher als einzige Autoren ein Programm auf Basis der AO-Klassifikation. Seine deskriptiv arbeitende Software rekonstruiert Frakturen dreidimensional und segmentiert die Frakturanatomie farblich, kann jedoch keine Klassifizierung durchführen.

5 Zusammenfassung

Einleitung: Die proximale Tibiafraktur mit Gelenkbeteiligung stellt eine chirurgische Herausforderung dar. Der Schlüssel zu einer erfolgreichen Therapieplanung ist eine akkurate Diagnostik mit einer präzisen dreidimensionalen Beschreibung der Fraktur. Hierzu wurden dreidimensionale Klassifikationssysteme für den Tibiakopf beschrieben, welche jedoch bisher unzureichend hinsichtlich Vergleichbarkeit und Reliabilität analysiert wurden. Ziel dieser retrospektiven radiologischen Studie war es die vier gängigsten, dreidimensionalen, CT-basierten Klassifikationssysteme u.a. hinsichtlich Intrarater-Reliabilität und Percent-Agreements zu untersuchen.

Methoden: Retrospektiv wurden 386 Computertomographien von 381 Patienten, die von 2015-2019 in der BG Unfallklinik Tübingen auf Grund einer Tibiakopffraktur therapiert wurden, analysiert. Die Bildgebung wurde durch einen Observer analysiert und die Fraktur zu zwei Zeitpunkten klassifiziert. Hierzu wurde das 3-Säulen-Konzept nach Luo, die 10-Segment-Klassifikation nach Krause/Frosch, die Kfuri/Schatzker-Klassifikation sowie die AO-Klassifikation verwendet. Das Zeitintervall zwischen erster und zweiter Bewertung betrug 8 Wochen. Die Häufigkeit der Frakturtypen für jedes Klassifikationssystem wurde deskriptiv dargestellt. Des Weiteren erfolgte die Berechnung der Intrarater-Reliabilität mittels Cohens-Kappa (κ) und des Percent-Agreements (p.a.). Der Grad der Übereinstimmung wurde anhand der Empfehlungen von Landis und Koch bewertet. Jede Klassifikation wurde mittels subjektiver, 5-stufiger Ordinalskala bezüglich des Informationsgehaltes (Differenzierung zw. Frakturen), der Lernkurve sowie der Möglichkeit der chirurgischen Therapieplanung bewertet.

Ergebnisse: Hinsichtlich der AO-Klassifikation wurden 36 B1, 66 B2, 123 B3, 21 C1, 8 C2 und 132 C3-Frakturen befundet. In der Luo-Klassifikation zeigte sich eine Beteiligung der posterioren Säule in 331 (86%) Fällen, der lateralen Säule in 323 Fällen (84%) und der medialen Säule in 184 Fällen (48%).

Die Intrarater-Reliabilität der Klassifikation nach Luo betrug $\kappa=0,848$, p.a.=89,1, nach Krause/Frosch $\kappa=0,621$ p.a.=64,5, nach Kfuri/Schatzker $\kappa=0,793$, p.a.=81,1 sowie nach AO $\kappa=0,756$, p.a.=79,9. Die Luo-Klassifikation zeigte eine leichte Erlernbarkeit. Die Differenzierung zwischen Frakturen ist hingegen erschwert.

Die höchsten Reliabilitäts-Werte hinsichtlich Frakturgröße zeigten folgende Klassifikationen: die AO-Klassifikation bei kleinen Frakturen, die Kfuri/Schatzker-Klassifikation bei mittleren Frakturen, die Luo-Klassifikation bei großen Frakturen und die Krause-Klassifikation bei Trümmerfrakturen.

In der Kfuri/Schatzker-Klassifikation ist die Differenzierung zwischen lateraler Impressions- (Typ III) und lateraler Impressions-Spaltfraktur (Typ II) erschwert. Eine präzise Konstruktion des „virtuellen Äquators“ ist erschwert, wenn das kraniale Ende des Fibulakopfes unterhalb der Gelenkfläche liegt und folglich in der axialen CT-Schichtung (Konstruktionsebene) nicht sichtbar ist. Die AO-Klassifikation zeigte niedrigere Reliabilitäten in der Bewertung von C-Frakturen. Ein Grund hierfür scheint die Vielzahl ähnlicher Subtypen in der bikondylären Frakturkategorie C.

Schlussfolgerung/Ausblick: Die Klassifikation von Luo zeigte die höchste Reliabilität. Der Informationsgehalt ist aufgrund vergleichsweise weniger und größerer Fraktursubgruppen gering. Die Therapieableitung ist erschwert. Die Klassifikationen von Kfuri/Schatzker und AO bieten geeignete Differenzierungsmöglichkeiten bei hohen Reliabilitäten und adäquater Komplexität für den klinischen Alltag. Weitere Studien zur Interrater-Reliabilität sind nötig. Des Weiteren ist die Anwendung im klinischen Setting bzgl. Ableitung einer chirurgischen Strategie sowie die Verknüpfung mit klinischen Outcomes in weiteren Arbeiten zukünftig zu untersuchen.

6 Literaturverzeichnis

1. Elsoe R, Larsen P, Nielsen NP, Swenne J, Rasmussen S, Ostgaard SE (2015) Population-based epidemiology of tibial plateau fractures. *Orthopedics* 38(9): p. 780-786.
2. Lüring C, Baumann P (2011) Erkrankungen und Verletzungen von Kniegelenk und Unterschenkel. In: *Orthopädie und Unfallchirurgie*, Grifka J, Kuster M (Hrsg), Springer, Heidelberg, 1109 Seiten.
3. Aurich M, Koenig V, Hofmann G (2018) Comminuted intraarticular fractures of the tibial plateau lead to posttraumatic osteoarthritis of the knee: Current treatment review. *Asian J Surg* 41(2): p. 99-105.
4. Honkonen SE (1995) Degenerative arthritis after tibial plateau fractures. *J Orthop Trauma* 9(4): p. 273-277.
5. Singleton N, Sahakian V, Muir D (2017) Outcome After Tibial Plateau Fracture: How Important Is Restoration of Articular Congruity? *J Orthop Trauma* 31(3): p. 158-163.
6. Elsoe R JM, Larsen P (2019) Tibial plateau fractures are associated with a long-lasting increased risk of total knee arthroplasty a matched cohort study of 7,950 tibial plateau fractures. *Osteoarthritis Cartilage* 27(5): p. 805-809.
7. Bobrich E, Haupt C, Grass R, Zwipp H (2009) Tibiakopffraktur und Luxationsfraktur. *Trauma und Berufskrankheit* 11(2): p. 154-159.
8. Krause M, Preiss A, Müller G, Madert J, Fehske K, Neumann MV, Domnick C, Raschke M, Südkamp N, Frosch KH (2016) Intra-articular tibial plateau fracture characteristics according to the "Ten segment classification". *Injury* 47(11): p. 2551-2557.
9. Assink N, Kraeima J, Slump CH, Ten Duis K, de Vries J, Meesters AML, van Ooijen P, Witjes MJH, FFA IJ (2019) Quantitative 3D measurements of tibial plateau fractures. *Sci Rep* 9(1). Art.No.: 14395
10. Doornberg JN, Rademakers MV, van den Bekerom MP, Kerkhoffs GM, Ahn J, Steller EP, Kloen P (2011) Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. *Injury* 42(12): p. 1416-1425.
11. Molenaars RJ, Mellema JJ, Doornberg JN, Kloen P (2015) Tibial Plateau Fracture Characteristics: Computed Tomography Mapping of Lateral, Medial, and Bicondylar Fractures. *J Bone Joint Surg Am* 97(18): p. 1512-1520.
12. Molenaars RJ, Solomon LB, Doornberg JN (2019) Articular coronal fracture angle of posteromedial tibial plateau fragments: A computed tomography fracture mapping study. *Injury* 50(2): p. 489-496.
13. Audigé L, Bhandari M, Hanson B, Kellam J (2005) A concept for the validation of fracture classifications. *J Orthop Trauma* 19(6): p. 401-406.
14. Krause M, Krüger S, Müller G, Püschel K, Frosch KH (2019) How can the articular surface of the tibial plateau be best exposed? A comparison of specific surgical approaches. *Arch Orthop Trauma Surg* 139(10): p. 1369-1377.

15. Krause M, Müller G, Frosch KH (2018) Surgical approaches to tibial plateau fractures. *Unfallchirurg* 121(7): p. 569-582.
16. Kfuri M, Schatzker J, Castiglia MT, Giordano V, Fogagnolo F, Stannard JP (2017) Extended Anterolateral Approach for Complex Lateral Tibial Plateau Fractures. *J Knee Surg* 30(3): p. 204-211.
17. Bertrand ML, Pascual-López FJ, Guerado E (2017) Severe tibial plateau fractures (Schatzker V-VI): open reduction and internal fixation versus hybrid external fixation. *Injury* 48 Suppl 6: p. 81-85.
18. Elabjer E, Benčić I, Ćuti T, Cerovečki T, Ćurić S, Vidović D (2017) Tibial plateau fracture management: arthroscopically-assisted versus ORIF procedure - clinical and radiological comparison. *Injury* 48 Suppl 5: p. 61-64.
19. Sun H, Zhu Y, He QF, Shu LY, Zhang W, Chai YM (2017) Reinforcement strategy for lateral rafting plate fixation in posterolateral column fractures of the tibial plateau: The magic screw technique. *Injury* 48(12): p. 2814-2826.
20. Menghi A, Mazzitelli G, Marzetti E, Barberio F, D'Angelo E, Maccauro G (2017) Complex tibial plateau fractures: a retrospective study and proposal of treatment algorithm. *Injury* 48 Suppl 3: p. 1-6.
21. Millar SC, Arnold JB, Thewlis D, Fraysse F, Solomon LB (2018) A systematic literature review of tibial plateau fractures: What classifications are used and how reliable and useful are they? *Injury* 49(3): p. 473-490.
22. Schatzker J (1974) Compression in the surgical treatment of fractures of the tibia. *Clin Orthop Relat Res*(105): p. 220-239.
23. Marsh JL, Slongo TF, Agel J, Broderick JS, Creevey W, DeCoster TA, Prokuski L, Sirkin MS, Ziran B, Henley B, Audigé L (2007) Fracture and dislocation classification compendium - 2007: Orthopaedic Trauma Association classification, database and outcomes committee. *J Orthop Trauma* 21(10): p. 1-133.
24. Schatzker J, McBroom R, Bruce D (1979) The tibial plateau fracture. The Toronto experience 1968-1975. *Clin Orthop Relat Res*(138): p. 94-104.
25. Moore TM (1981) Fracture-dislocation of the knee. *Clin Orthop Relat Res*(156): p. 128-140.
26. Charalambous CP, Tryfonidis M, Alvi F, Moran M, Fang C, Samarji R, Hirst P (2007) Inter- and intra-observer variation of the Schatzker and AO/OTA classifications of tibial plateau fractures and a proposal of a new classification system. *Ann R Coll Surg Engl* 89(4): p. 400-404.
27. Kfuri M, Schatzker J (2018) Revisiting the Schatzker classification of tibial plateau fractures. *Injury* 49(12): p. 2252-2263.
28. Tuladhar R, Hu SJ, Chang SM (2019) The articular and non-articular proportion in segment classification of tibial plateau fractures. *Injury* 50(3): p. 818-819.
29. Audigé L, Bhandari M, Kellam J (2004) How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand* 75(2): p. 184-194.
30. Martin JS, Marsh JL (1997) Current classification of fractures. Rationale and utility. *Radiol Clin North Am* 35(3): p. 491-506.

31. Burstein AH (1993) Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am* 75(12): p. 1743-1744.
32. Garbuz DS, Masri BA, Esdaile J, Duncan CP (2002) Classification systems in orthopaedics. *J Am Acad Orthop Surg* 10(4): p. 290-297.
33. Martin J, Marsh JL, Nepola JV, Dirschl DR, Hurwitz S, DeCoster TA (2000) Radiographic fracture assessments: which ones can we reliably make? *J Orthop Trauma* 14(6): p. 379-385.
34. Kohn D, Knie. Expertise Orthopädie und Unfallchirurgie. Vol. 1. 2015, Stuttgart: Georg Thieme Verlag. 500 Seiten.
35. Jagodzinski M, Das Knie - Form, Funktion und ligamentäre Wiederherstellungschirurgie. Vol. 2. 2016, Berlin: Springer Verlag. 210 Seiten.
36. Drenckhahn D, Anatomie. Makroskopische Anatomie, Histologie, Embryologie, Zellbiologie. Band 1: Zelle, Gewebe, Entwicklung, Skelett- und Muskelsystem, Atemsystem, Verdauungssystem, Harn- und Genitalsystem. Vol. 17. 2008, München: Urban & Fischer Verlag/Elsevier GmbH. 968 Seiten.
37. Bhattacharyya T, McCarty LP, Harris MB, Morrison SM, Wixted JJ, Vrahas MS, Smith RM (2005) The posterior shearing tibial plateau fracture: treatment and results via a posterior approach. *J Orthop Trauma* 19(5): p. 305-310.
38. Barei DP, O'Mara TJ, Taitzman LA, Dunbar RP, Nork SE (2008) Frequency and fracture morphology of the posteromedial fragment in bicondylar tibial plateau fracture patterns. *J Orthop Trauma* 22(3): p. 176-182.
39. Higgins TF, Kemper D, Klatt J (2009) Incidence and morphology of the posteromedial fragment in bicondylar tibial plateau fractures. *J Orthop Trauma* 23(1): p. 45-51.
40. Brunner A, Horisberger M, Ulmar B, Hoffmann A, Babst R (2010) Classification systems for tibial plateau fractures; does computed tomography scanning improve their reliability? *Injury* 41(2): p. 173-178.
41. Gicquel T, Najihi N, Vendevre T, Teyssedou S, Gayet LE, Hutten D (2013) Tibial plateau fractures: reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification. *Orthop Traumatol Surg Res* 99(7): p. 805-816.
42. Yacoubian SV, Nevins RT, Sallis JG, Potter HG, Lorich DG (2002) Impact of MRI on treatment plan and fracture classification of tibial plateau fractures. *J Orthop Trauma* 16(9): p. 632-637.
43. Gardner MJ, Yacoubian S, Geller D, Suk M, Mintz D, Potter H, Helfet DL, Lorich DG (2005) The incidence of soft tissue injury in operative tibial plateau fractures: a magnetic resonance imaging analysis of 103 patients. *J Orthop Trauma* 19(2): p. 79-84.
44. Luo CF, Sun H, Zhang B, Zeng BF (2010) Three-column fixation for complex tibial plateau fractures. *J Orthop Trauma* 24(11): p. 683-692.
45. Müller ME, Nazarian S (1981) Classification of fractures of the femur and its use in the A.O. *Rev Chir Orthop Reparatrice Appar Mot* 67(3): p. 297-309.

46. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF (2018) Fracture and Dislocation Classification Compendium-2018. *J Orthop Trauma* 32(1): p. 1-170.
47. Gerard-Marchant P (1939) Fractures des plateaux tibiaux. *Rev Chir Orthop Reparatrice Appar Mot* 26: p. 499-546.
48. Duparc J (1960) Fractures articulaires de l'extrémité supérieure du tibia. *Rev Chir Orthop* 46: p. 399-486.
49. Tscherne H, Lobenhoffer P (1993) Tibial plateau fractures. Management and expected results. *Clin Orthop Relat Res* 292: p. 87-100.
50. Müller M, Allgöwer M (1977) AO Technik. In: *Manual der Osteosynthese*, Springer, Heidelberg, 409 Seiten.
51. Zhang BB, Sun H, Zhan Y, He QF, Zhu Y, Wang YK, Luo CF (2019) Reliability and repeatability of tibial plateau fracture assessment with an injury mechanism-based concept. *Bone Joint Res* 8(8): p. 357-366.
52. Yang G, Zhai Q, Zhu Y, Sun H, Putnis S, Luo C (2013) The incidence of posterior tibial plateau fracture: an investigation of 525 fractures by using a CT-based classification system. *Arch Orthop Trauma Surg* 133(7): p. 929-934.
53. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E (2011) Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: *Clinical Practice Guidelines We Can Trust*, National Academies Press (US) Copyright 2011 by the National Academy of Sciences. All rights reserved, Washington (DC), 266 Seiten.
54. Arifin WN. A Web-based Sample Size Calculator for Reliability Studies. 2022; Available from: <http://wnarifin.github.io>.
55. Lydersen S (2018) Cohen's kappa – a measure of agreement between observers. *Tidsskr Nor Laegeforen* 138(5).
56. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1): p. 159-174.
57. McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22(3): p. 276-282.
58. Alencar JB, Souza CJD, Coelho PRR, Freitas PG, Lopes MBG, Pinto FJM (2020) Inter- and intraobserver reliability of Schatzker, AO and Luo Classifications for tibial plateau fractures. *Acta Ortop Bras* 28(5): p. 216-220.
59. Taşkesen A, Demirkale İ, Okkaoğlu MC, Özdemir M, Bilgili MG, Altay M (2017) Intraobserver and interobserver reliability assessment of tibial plateau fracture classification systems. *Eklemler Hastalıkları Cerrahisi* 28(3): p. 177-181.
60. Martínez-Rondanelli A, Escobar-González SS, Henao-Alzate A, Martínez-Cano JP (2017) Reliability of a four-column classification for tibial plateau fractures. *Int Orthop* 41(9): p. 1881-1886.
61. Hoekstra H, Kempnaers K, Nijs S (2017) A revised 3-column classification approach for the surgical planning of extended lateral tibial plateau fractures. *Eur J Trauma Emerg Surg* 43(5): p. 637-643.
62. Mellema JJ, Doornberg JN, Molenaars RJ, Ring D, Kloen P (2016) Interobserver reliability of the Schatzker and Luo classification systems for tibial plateau fractures. *Injury* 47(4): p. 944-949.

63. de Lima Lopes C, da Rocha Cândido Filho CA, de Lima ESTA, Gonçalves MC, de Oliveira RL, de Lima PR (2014) Importance of radiological studies by means of computed tomography for managing fractures of the tibial plateau. *Rev Bras Ortop* 49(6): p. 593-601.
64. Zhu Y, Hu CF, Yang G, Cheng D, Luo CF (2013) Inter-observer reliability assessment of the Schatzker, AO/OTA and three-column classification of tibial plateau fractures. *J Trauma Manag Outcomes* 7(1): p. 7.
65. Patange Subba Rao SP, Lewis J, Haddad Z, Paringe V, Mohanty K (2014) Three-column classification and Schatzker classification: a three- and two-dimensional computed tomography characterisation and analysis of tibial plateau fractures. *Eur J Orthop Surg Traumatol* 24(7): p. 1263-1270.
66. te Stroet MA, Holla M, Biert J, van Kampen A (2011) The value of a CT scan compared to plain radiographs for the classification and treatment plan in tibial plateau fractures. *Emerg Radiol* 18(4): p. 279-283.
67. Hu YL, Ye FG, Ji AY, Qiao GX, Liu HF (2009) Three-dimensional computed tomography imaging increases the reliability of classification systems for tibial plateau fractures. *Injury* 40(12): p. 1282-1285.
68. Maripuri SN, Rao P, Manoj-Thomas A, Mohanty K (2008) The classification systems for tibial plateau fractures: how reliable are they? *Injury* 39(10): p. 1216-1221.
69. Walton NP, Harish S, Roberts C, Blundell C (2003) AO or Schatzker? How reliable is classification of tibial plateau fractures? *Arch Orthop Trauma Surg* 123(8): p. 396-398.
70. Dhillon MS, Patel S, K P (2017) Simple four column classification can dictate treatment for intra articular tibial plateau fractures much better than ten segment classification. *Injury* 48(6): p. 1276-1278.
71. Mansur H, Corrêa V, Abdo B, Ramos L, Castiglia M (2022) Avaliação da reprodutibilidade da classificação de Schatzker revisada por Kfuri para as fraturas do planalto tibial. *Revista Brasileira de Ortopedia* 57(5): p. 502-510.
72. Wennergren D, Ekholm C, Sundfeldt M, Karlsson J, Bhandari M, Möller M (2016) High reliability in classification of tibia fractures in the Swedish Fracture Register. *Injury* 47(2): p. 478-482.
73. Wang Y, Luo C, Zhu Y, Zhai Q, Zhan Y, Qiu W, Xu Y (2016) Updated Three-Column Concept in surgical treatment for tibial plateau fractures - A prospective cohort study of 287 patients. *Injury* 47(7): p. 1488-1496.
74. Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK, Heppenstall RB (1997) Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. *J Orthop Trauma* 11(7): p. 484-489.
75. Anwar A, Zhang Y, Zhao Z, Gao Y, Sha L, Lv D, Zhang Z, Lv G, Zhang Y, Nazir MU, Qasim W, Wang Y (2019) Two column classification of tibial plateau fractures; description, clinical application and reliability. *Injury* 50(6): p. 1247-1255.
76. Khan MA, Pal S, Ahmed SK, Chinoy MA (2021) Inter-observer variation of the Schatzker and Khan classification of Tibial plateau fractures: Cohort study. *J Pak Med Assoc* 71(Suppl 5)(8): p. 51-54.

77. de Vet HCW, Terwee CB, Mokkink LB, Knol DL (2011) Practical Guides to Biostatistics and Epidemiology. In: Measurement in Medicine: A Practical Guide, Cambridge University Press, Cambridge. 350 Seiten.
78. Mandrekar JN (2011) Measures of interrater agreement. *J Thorac Oncol* 6(1): p. 6-7.
79. Albuquerque RP, Giordano V, Pallottino A, Sassine T, Canedo R, Pina J, do Amaral NP (2009) Analysis of the reproducibility of tibial plateau fractures' classification. *Rev Bras Ortop* 44(3): p. 225-229.
80. Knottnerus JA, Muris JW (2003) Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* 56(11): p. 1118-1128.
81. Martin JS, Marsh JL, Bonar SK, DeCoster TA, Found EM, Brandser EA (1997) Assessment of the AO/ASIF fracture classification for the distal tibia. *J Orthop Trauma* 11(7): p. 477-483.
82. Krause M, Frosch KH (2017) Response to the letter-to-the-editor by Dhillon et al. "Simple four column classification can dictate treatment for intra articular tibial plateau fractures much better than ten segment classification", *Injury* 2017. *Injury* 48(10): p. 2369-2370.
83. Meulenkamp B, Martin R, Desy NM, Duffy P, Korley R, Puloski S, Buckley R (2017) Incidence, risk factors, and location of articular malreductions of the tibial plateau. *J Orthop Trauma* 31(3): p. 146-150.
84. Lobenhoffer P, Gerich T, Bertram T, Lattermann C, Pohlemann T, Tscheme H (1997) Particular posteromedial and posterolateral approaches for the treatment of tibial head fractures. *Unfallchirurg* 100(12): p. 957-967.
85. Yao X, Xu Y, Yuan J, Lv B, Fu X, Wang L, Yang S, Meng S (2018) Classification of tibia plateau fracture according to the "four-column and nine-segment". *Injury* 49(12): p. 2275-2283.
86. Zhu Y, Yang G, Luo CF, Smith WR, Hu CF, Gao H, Zhong B, Zeng BF (2012) Computed tomography-based Three-Column Classification in tibial plateau fractures: introduction of its utility and assessment of its reproducibility. *J Trauma Acute Care Surg* 73(3): p. 731-737.
87. Rossmann M, Fensky F, Ozga AK, Rueger JM, Märdian S, Russow G, Brunnemer U, Schmidmaier G, Hofmann A, Herlyn P, Mittlmeier T, Amer A, Gössling T, Grossterlinden LG (2020) Tibial plateau fracture: does fracture classification influence the choice of surgical approach? A retrospective multicenter analysis. *Eur J Trauma Emerg Surg* 48(5): p. 3635-3641.
88. Carlson DA (2005) Posterior bicondylar tibial plateau fractures. *J Orthop Trauma* 19(2): p. 73-78.
89. Wahlquist M, Iagulli N, Ebraheim N, Levine J (2007) Medial tibial plateau fractures: a new classification system. *J Trauma* 63(6): p. 1418-1421.
90. Suero EM, Hübner T, Stübiger T, Krettek C, Citak M (2010) Use of a virtual 3D software for planning of tibial plateau fracture reconstruction. *Injury* 41(6): p. 589-591.

7 Erklärung zum Eigenanteil

Die Arbeit wurde in der Klinik für Unfall- und Wiederherstellungschirurgie der BG Unfallklinik Tübingen unter Betreuung von Herrn Prof. Dr. med. C. Ihle durchgeführt. Die Konzeption der Studie erfolgte in Zusammenarbeit mit Herrn Prof. Dr. med. C. Ihle und Herrn PD Dr. med. M.-D. Ahrend. Die Detektion der Patienten aus der Patientenliste nach Einschlusskriterien, die Erhebung der Daten aus den digitalen Patientenakten, die Begutachtung der CTs, die Klassifizierung der Frakturen sowie die digitale Dokumentation erfolgten durch mich. Die statistische Auswertung erfolgte, nach Beratung durch das Institut für klinische Epidemiologie und angewandte Biometrie des Universitätsklinikums Tübingen (Frau You-Shang Feng), durch mich. Ich versichere, dass Manuskript selbstständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

Tübingen, den 08.05.2023

Dr. med. dent. Felix Eggensperger

8 Danksagung

An dieser Stelle möchte ich mich für die vielfältig erfahrene Hilfe bedanken. Ohne die Unterstützung zahlreicher Personen und Institutionen hätte diese Dissertation nicht realisiert werden können.

Ein besonderer Dank gilt zunächst meinem Doktorvater Herrn Prof. Dr. med. Christoph Ihle für die spontane und unkomplizierte Annahme als Doktorand, für die stets wertvollen, konstruktiven und fachlich präzisen Ratschläge eines erfahrenen Orthopäden und Unfallchirurgen sowie für die freundliche, aufgeschlossene und humorvolle Art.

Herrn Prof. Dr. med. Steffen Schröter möchte ich für sein ausgesprochenes Interesse am Thema dieser Dissertation, für den wissenschaftlichen Impuls und die inhaltliche Grundsteinlegung der Studie danken.

Ein spezieller Dank gilt Frau Prof. Dr. med. Tina Histing, der ärztlichen Direktorin der BG Unfallklinik Tübingen, für die Möglichkeit der erneuten Promotion in ihrem Hause. Ich habe mich während meiner Pflegepraktika und Famulaturen, meinen beiden Promotionen sowie unzähligen Kursen stets wohl gefühlt und konnte viel lernen.

Frau You-Shang Feng vom „Institut für klinische Epidemiologie und angewandte Biometrie“ des Universitätsklinikums Tübingen möchte ich danken für die hervorragende Unterstützung im Bereich der medizinischen Statistik.

Ein Dankeschön möchte ich auch meinem langjährigen Freund Herrn Sebastian Schrof aussprechen, für die großartige Hilfe beim Entwerfen der Abbildungen und Illustrationen.

Vielen Dank meiner lieben Mutter und meiner Freundin Sara für die wundervolle Unterstützung während des Projektes.

PD Dr. med. Marc-Daniel Ahrend hat mich als Initiator, Förderer, Kritiker, Ratgeber, Motivator und Freund den Weg vom ersten Treffen bis zur Disputation begleitet. Ein Betreuer und Mentor wie man ihn sich nur wünschen kann.

Vielen Dank lieber Marc für deine Zeit, deinen Einsatz, deine Kompetenz, deine Hilfsbereitschaft sowie dein fachliches und menschliches Interesse.