

Interuntersucher-Varianz bei der Beurteilung von Röntgenaufnahmen von Pferden nach dem Röntgenleitfaden 2007

Maren Hellige¹, Karl Rohn², Lars Buschkamp³ und Peter Stadler¹

¹ Klinik für Pferde, Stiftung Tierärztliche Hochschule Hannover

² Institut für Biometrie, Epidemiologie und Informationsverarbeitung, Stiftung Tierärztliche Hochschule Hannover

³ Tierklinik Hochmoor, Gescher-Hochmoor

Zusammenfassung: Die röntgenologische Untersuchung ist seit vielen Jahren ein wesentlicher Bestandteil der Kaufuntersuchung bei Pferden. In Deutschland hat sich die Beurteilung der Röntgenbilder nach dem Röntgenleitfaden (RöLF07) weitgehend durchgesetzt und stellt mit der Einteilung in die Röntgenklassen sowohl für den Tierarzt, den Käufer und den Verkäufer als auch für die Juristen ein Bewertungskriterium für Pferde dar. Im Rahmen der Kaufuntersuchung schuldet der Tierarzt die „richtige“ Befundung der im RöLF07 aufgeführten Röntgenaufnahmen. Allerdings entstehen nicht selten Rechtsstreitigkeiten, weil die Befundung von Röntgenbildern durch verschiedene Untersucher zu Differenzen führen kann. In der vorliegenden Studie wurde die Übereinstimmung von mehreren erfahrenen Untersuchern bei der Beurteilung von fünfzig Röntgenbildern nach dem Röntgenleitfaden07 untersucht. An der Studie nahmen vierzig erfahrene Pferdetierärzte aus dem gesamten Bundesgebiet teil. Sie wurden einzeln aufgesucht und haben die fünfzig ausgewählten Standardaufnahmen (Zehe 90°, Oxspring, Tarsus 45°, 135° und 0°, Knie 90° und 0°, Dornfortsätze der Sattellage) auf demselben Betrachtungsmonitor nach den Röntgenkennziffern (RöKZ) des RöLF 07 beurteilt. Es wurde die Übereinstimmung der Untersucher in Bezug auf die RöKZ und den Endbefund (höchste Röntgenklasse des jeweiligen Röntgenbildes) für jede Aufnahme prozentual erhoben. Außerdem wurde mittels des PABAK die Übereinstimmung aller Untersucher in Bezug auf die Endbefunde ermittelt. Für die RöKZ wurde pro Aufnahme eine Übereinstimmung von 31% bis 97,5% aller Untersucher festgestellt. In Bezug auf die Röntgenklasse des Endbefundes der 50 Röntgenaufnahmen wurde eine gute Übereinstimmung (PABAK $\kappa=0,72$) zwischen den vierzig Untersuchern festgestellt. Die Ergebnisse dieser Studie zeigen die interindividuellen Differenzen zwischen erfahrenen Pferdetierärzten, die sich insbesondere bei Befunden mit niedriger Röntgenklasse (II, II–III oder III) offenbaren.

Schlüsselwörter: Röntgen, Pferd, Kaufuntersuchung, Röntgenleitfaden, Röntgenklassen, Interindividuelle Befundung

Interobserver agreement in interpreting radiographs of horses using the German Radiology-Guidelines 2007 ("Röntgenleitfaden 2007")

For many years, radiography has formed an integral part of the equine pre-purchase examination. In Germany, the radiological guidelines "Röntgenleitfaden 2007 (RöLF07)" are widely used by equine veterinarians to assist with the interpretation of radiographic findings of riding horses. The radiological guidelines include over 200 radiographic findings, which are registered, and each finding has an individual number and a corresponding class. The radiological findings are divided into four radiological classes (I–IV) and three intermediate classes (I–II, II–III and III–IV), which were developed as an assessment criterion of the horse for the owner, the vendor and the purchaser. Normal anatomical variations are classified as class I, whereas severe radiographic findings were classified as class III–IV or IV. The highest radiological class of a single finding is the final class of the horse for the radiological examination. Due to the forensic situation, the veterinarian has to classify every radiographic finding correctly. As there are variable interpretations of radiographic findings of different observers' litigations are quite common. The aim of the study was to characterise the inter-observer agreement of several experienced observers in orthopaedic findings observed on radiographs using the German radiological guidelines "Röntgenleitfaden 07". The 50 sets of radiographs were interpreted once by 41 experienced equine veterinarians. One of the authors visited every observer to make sure that the technical equipment and circumstances of the examinations were similar. The observers were asked to examine the radiographic sets as they would have been examined at a pre-purchase examination. A set of 50 orthopaedic radiographs (fore feet: lateromedial and dorsoproximal-palmarodistal oblique view, hind feet: lateromedial view, tarsi: dorsomedial-plantarolateral oblique (DMPLO), dorsolateral-plantaromedial oblique (DLPMO) and dorsoplantar views, stifles: lateromedial and caudo-cranial view, thoracolumbar spine: latero-lateral view of the spinous processes) were examined using the "Röntgenleitfaden 07" to characterise and classify the findings. Two of the authors selected the orthopaedic radiographs from the archive of the Clinic for Horses of the University of Veterinary Medicine in Hannover. The number of observers to identify each finding, per radiograph, was calculated and reported. A representative selection of radiographs of every location (feet, tarsi, stifle and spinous processes) is documented in the paper. Furthermore, the Kappa statistic for inter-rater agreement was calculated using the bias weighted PABAK. Larger variation on the number of radiographic findings were noted for each anatomical location. For example, up to 35 different findings were observed in the foot whereas between 5 and 18 findings per radiograph were detected at other locations. The frequency of reporting the same radiographic number was between 29% and 97% for all radiographs. Related to the final classification of every single radiograph the percentage of agreement was between 36% and 100%. There was a good interrater agreement (PABAK $\kappa=0.72$) between all observers referring to the fifty radiographs for the overall final classification. The agreement was excellent for larger and easy to categorise radiographic findings for example an osseous cyst like lesion in the medial femoral condyle or obvious findings like osteochondral fragments, whereas more subtle lesions resulted in poorer agreement. In the radiological guidelines "Röntgenleitfaden 07", several orthopaedic findings are described and classified, however poor to fair agreement between experienced observers for ambiguous findings occur.

Keywords: horse, radiology, purchase-examination, interobserver, agreement

Zitation: Hellige M., Rohn K., Buschkamp L., Stadler P. (2018) Interuntersucher-Varianz bei der Beurteilung von Röntgenaufnahmen von Pferden nach dem Röntgenleitfaden 2007. *Pferdeheilkunde* 34, 212–222; DOI 10.21836/PEM20180301

Korrespondenz: Dr. Maren Hellige, Klinik für Pferde, Stiftung Tierärztliche Hochschule Hannover, Bünteweg 9, 30559 Hannover; maren.hellige@tiho-hannover.de

Einleitung

In der Pferdemedizin ist die röntgenologische Untersuchung eine tragende Säule der bildgebenden Diagnostik bei der Untersuchung von Pferden mit orthopädischen Erkrankungen der Gliedmaßen. Das Röntgen klinisch unauffälliger Pferde hingegen bildet seit vielen Jahren einen Bestandteil der tierärztlichen Kaufuntersuchung und dient als Hilfestellung bei der Beurteilung des orthopädischen Status Präsens zum Zeitpunkt der Kaufuntersuchung (Bladon und Main 2003, Bemmann et al. 2008). Um die Anfertigung, Interpretation und Beurteilung der Standardaufnahmen im Rahmen einer Kaufuntersuchung zu vereinheitlichen, wurde der Röntgenleitfaden mit einer Einteilung in Klassen und Zwischenklassen entwickelt und inzwischen mehrfach überarbeitet (1993, Utrecht und Zürich, überarbeitet 2002 und 2007). Auch mit der letzten Überarbeitung hielt die Röntgenkommission an dem Prinzip fest, die Röntgenbefunde in Klassen einzuteilen (Gerhards et al. 2008). Die Einteilung in Klassen und Zwischenklassen war ausschließlich auf die Ergebnisse der Röntgenuntersuchung zu beziehen. Mit den Röntgenkennziffern sind alle Befunde einfach zu dokumentieren und zu vergleichen, falls die Röntgenaufnahmen mehreren Tierärzten zur Beurteilung vorgelegt werden (Bemmann et al. 2008). Dennoch kommt es durch verschiedene Untersucher häufig zu einer unterschiedlichen Interpretation einzelner Befunde, welche sich in einer deutlichen Diskrepanz bei der Klassifizierung von einzelnen Röntgenbildern niederschlagen kann. In der Humanradiologie haben die meisten Fehler in der radiologischen Befundung geringe bzw. nicht signifikante Auswirkungen auf das Wohl des Patienten (Brady et al. 2012). In der Pferdemedizin hingegen wird insbesondere im Rahmen der Kaufuntersuchung eine „richtige“ Beurteilung der Röntgenbilder geschuldet und vom Auftraggeber erwartet, auch wenn nicht selten die klinische Relevanz fraglich bleibt. Dies kann zu gerichtlichen Auseinandersetzungen führen, bei denen allein das Ergebnis der röntgenologischen Untersuchung, d.h. die Interpretation der bildgebenden Diagnostik der Ankaufuntersuchung beanstandet wird (Stadler 2005). Es wird in der Regel vermutet, dass der Käufer selbst bei Mitteilung eines Befundes der Röntgenklasse II–III vom Kauf Abstand genommen hätte, wenn er nachvollziehbar vorträgt, dass es ihm darauf ankam, ein Pferd zu erwerben, das keine Befunde aufweist, die im ungünstigsten Fall in die RöKl II einzustufen sind (Plewa 2014).

Aus der Humanmedizin ist bekannt, dass bei der Interpretation von Röntgenaufnahmen mit orthopädischen Fragestellungen eine schlechte bis mäßige Übereinstimmung zwischen mehreren Untersuchern vorliegt und die subjektive Einschätzung aufgrund der Erfahrung des Untersuchers zu unterschiedlichen Beurteilungen führt (Engel et al. 2002, Lindenhovius et al. 2012). Von den praktizierenden Pferdetierärzten, die Kaufuntersuchungen durchführen, wird hingegen eine detaillierte, korrekte und insbesondere einheitliche Befundung der Standardröntgenbilder erwartet.

Bei der Interpretation von Röntgenaufnahmen der Gliedmaßen von Pferden liegen einige Studien über die interindividuelle Übereinstimmung von bis zu vier Untersuchern vor (Groth et al. 2009, Beccati et al. 2013, Jackson et al. 2013). Bisher ist allerdings in keiner Untersuchung der Röntgenleitfaden als Grundlage für eine Beurteilung von unterschiedlichen

Röntgenbildern durch mehrere Untersucher verwendet worden.

Im Rahmen eines „Workshop Röntgenleitfaden 07“ wurden im Sinne einer Feldstudie von 50 Tierärzten (Teilnehmer und Referenten) jeweils 10 Röntgenbilder von drei Pferden nach dem RÖLF07 ausgewertet. Dabei wurde die Übereinstimmung der unterschiedlichen Untersucher sowohl in Bezug auf die Röntgenklassen als auch auf die Röntgenkennziffern dokumentiert. Bei den Röntgenklassen pro Pferd wurde eine mittelmäßige Übereinstimmung zwischen den verschiedenen Tierärzten festgestellt. In Bezug auf die Röntgenkennziffern gab es unter den 50 Untersuchern keine zwei Tierärzte, die die gleichen Röntgenkennziffern vergeben haben (Zöttl et al. 2009). Auch diese Ergebnisse waren Anstoß für die hier vorliegende Studie.

Material und Methode

Es wurden bundesweit ca. 50 Pferdekliniken und -praxen sowie die Universitätskliniken schriftlich eingeladen, sich an der vorliegenden Studie zu beteiligen. Nach einer positiven Rückmeldung wurde mit den 41 teilnehmenden Kolleginnen bzw. Kollegen ein Termin vor Ort vereinbart. Die TeilnehmerInnen waren zumeist Fachtierärzte für Pferde, für Chirurgie oder Dipl. ECVS/ACVS und Inhaber oder Teilhaber von Pferdekliniken bzw. -praxen. Es wurden 50 digitale Röntgenbilder (DICOM 3.0) von durchschnittlicher bis guter diagnostischer Qualität von 50 verschiedenen Pferden mit folgender Verteilung der anatomischen Lokalisation für die vorliegende Studie ausgewählt:

- 10 × Zehe latero-medial (90°) (Vorder- und Hintergliedmaße)
- 10 Röntgenbilder nach Oxspring (Vordergliedmaße)
- 15 Aufnahmen von Tarsalgelenken (jeweils 5 × 0°-Aufnahmen, 5 × 45–70° Projektion und 5 × 90°–135°-Projektion)
- 10 Aufnahmen von Kniegelenken (5 × 90°-Projektion und 5 × 0°-Projektionsrichtung)
- 5 latero-laterale Aufnahmen der Dornfortsätze der Sattellarge

Es lag jeweils nur eine Aufnahme-richtung vor, d.h. es war nicht möglich, eine weitere Projektionsrichtung desselben Gelenkes zum Vergleich heranzuziehen.

Den Untersuchern wurden die Röntgenbilder auf einem MacBook Pro 15[®] (Apple Inc.) mit einem externen hochauflösenden Monitor (EIZO[®] FlexScan MX 190 S, Hakusan, Japan) präsentiert. Es wurde der DICOM Viewer „OsiriX3.6“ (Pixmeo, Genf, Schweiz) zur Betrachtung und Bildbearbeitung verwendet, so dass eine übliche Nachbearbeitung durch individuelle Veränderung von Helligkeit und Kontrast möglich war. Jeder Teilnehmer erhielt eine Einweisung in die Bedienung der Software und als Hilfsmittel auf einem separaten Notebook die Röntgen – CD RöKl[™] 4.0 (fzmb GmbH, Bad Langensalza), die neben beispielhaften Röntgenbildern für die einzelnen RöKZ auch den RÖLF07 in digitaler Form beinhaltet.

Jeder Untersucher erhielt eine individuelle Untersuchernummer, so dass die statistische Auswertung anonymisiert vorgenommen wurde. Da jeder Untersucher die Beurteilung der

Röntgenbilder allein vorgenommen hat, wurden Absprachen oder Konsensmeinungen verhindert.

In einem Befundbogen sollte der jeweilige Untersucher für die einzelnen Röntgenaufnahmen den Befund, die Röntgenkennziffer des erhobenen Befundes sowie die zugehörige Röntgenklasse eintragen. Wie auch bei der Befundung im Rahmen einer Kaufuntersuchung war es möglich, mehrere unterschiedliche RöKZ für eine Röntgenaufnahme zu dokumentieren. Nach dem RÖLF07 entsprach die höchste Klassifizierung eines Einzelbefundes der jeweiligen Aufnahme dem röntgenologischen Endbefund. Zudem sollte sowohl die Qualität (Projektionsrichtung, Artefakte, Belichtung) als auch der Schwierigkeitsgrad für jede einzelne Aufnahme dokumentiert werden. Für die Befundung der insgesamt 50 Röntgenbilder wurde den Untersuchern kein Zeitlimit gesetzt.

Statistik

Als Maß der Übereinstimmungen zwischen jeweils zwei Untersuchern in Bezug auf den Endbefund (höchste Röntgenklasse je Aufnahme) wurde der aus dem Fleiss-Cohen's kappa (Fleiss et al. 2003) entwickelte PABAK (prävalenz- und biasadjustierter Kappa, Byrt et al. 1993) gewählt. Der PABAK basiert auf dem Kappa-Test nach Cohen (1960). Nach Byrt et al. (1993) ermittelt der Kappa-Koeffizient die Übereinstimmung zweier Datenreihen unter Berücksichtigung der zufälligen Übereinstimmung. Der gewichtete Kappa-Koeffizient bewertet zudem größere Differenzen schwerwiegender als ähnliche Beurteilungen von verschiedenen Untersuchern. Weil der Kappa-Wert bei asymmetrischer Befundbeurteilung (z.B. überwiegend hohe Endbefunde zu einem Bild) trotz guter Übereinstimmung zwischen Beurteilern in diesem Fall zu niedrige Kappa-Werte errechnet (Byrt et al. 1993, Flight und Julious 2015), wurde zur Beurteilung der Übereinstimmung der PABAK-Wert berechnet. Der PABAK-Wert kann wie der Kappa-Wert zwischen -1 und 1 liegen, wobei der Wert 0,00 für rein zufällige Übereinstimmungen steht und ein Wert von 1,00 eine vollständige Übereinstimmung repräsentiert. Negative Werte (-1,00) weisen auf eine mehr als zufällig erwartete Abweichung hin. Für die Beurteilung der Kappa-Werte (κ) wurde die Interpretationsskala von Altman (1990) herangezogen:

$\kappa < 0,4$	unzureichende Übereinstimmung
$\kappa 0,41-0,60$	akzeptable Übereinstimmung
$\kappa 0,61-0,80$	gute Übereinstimmung
$\kappa > 0,80$	sehr gute Übereinstimmung

Ergebnisse

Im Folgenden werden die Einzelbeurteilungen für ausgewählte Röntgenbilder der jeweiligen anatomischen Region dargestellt.

Ergebnisse der Röntgenbefunde der Zehe (Zehe 01–Zehe 10)

Bei den latero-medialen Röntgenaufnahmen der Zehe von Vorder- und Hintergliedmaße (Zehe 01–09 von der Vordergliedmaße, Röntgenbild Zehe 10 von der Hintergliedmaße) wurden 5–35 (durchschnittlich 18) verschiedene Röntgenkennziffern von den Untersuchern angegeben. In Bezug auf

die RöKZ lag die höchste Übereinstimmung der Untersucher pro Röntgenaufnahme zwischen 45,95% und 95%. In Bezug auf den Endbefund (höchste RöKl je Röntgenaufnahme) lag die Übereinstimmung geringgradig höher (48,65%–97,5%). Exemplarisch werden die Röntgenaufnahmen mit der niedrigsten und der höchsten Übereinstimmung näher erläutert:

Die Röntgenaufnahme „Zehe03“ (Abb. 1) wurde von 37 Tierärzten/-innen beurteilt. Dabei wurden zumeist nur ein (37,8%) oder zwei (35,1%) verschiedene RöKZ angegeben, obwohl für diese Aufnahme insgesamt 20 verschiedene Röntgenkennziffern dokumentiert wurden. Lediglich bei 2 RöKZ wurde eine Übereinstimmung von mehr als 10 Untersuchern (n=17 bzw. 15) jedoch lediglich 45,95% der Untersucher festgestellt. Es handelte sich dabei um die Befunde •1.11.3 „Hufgelenk 90°/Strukturveränderungen im subchondralen Gelenkflächenbereich“ der RöKl III–IV und den Befund •1.2.8 „Zehennachse/deutlich gebrochen >5° im Hufgelenk als Beugung“ mit der RöKl II–III. Jedoch wurden vereinzelt auch die Befunde •1.2.9 „Zehennachse/deutlich gebrochen

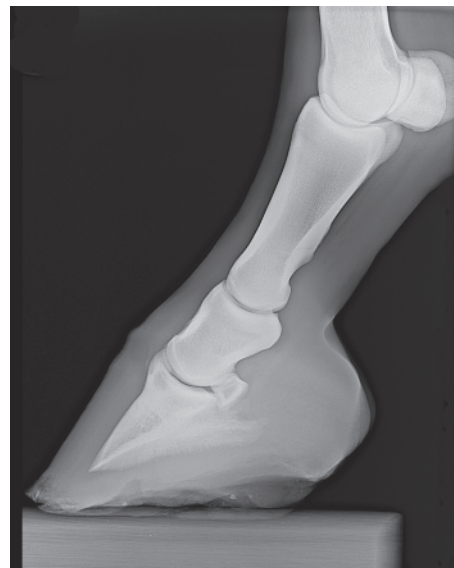


Abb. 1 Latero-mediale Röntgenaufnahme einer rechten Vorderzehe „Zehe03“. | Latero-medial view of a right front foot „Zehe03“

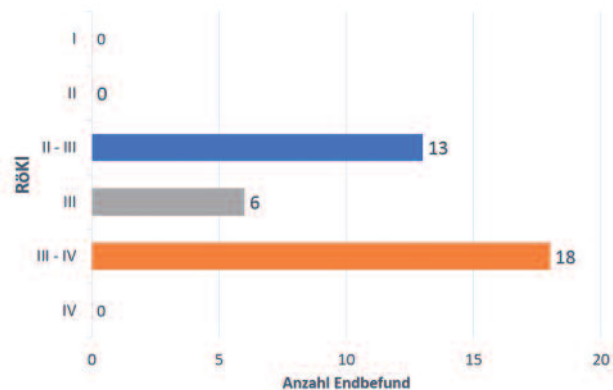


Abb. 2 Graphische Darstellung über die Anzahl der Nennungen des jeweiligen Endbefundes für die Aufnahme „Zehe 03“ (s. Abb. 1). X- Achse: Anzahl der Untersucher, Y-Achse: Röntgenklassen. Blauer Balken: Röntgenklasse II–III, grauer Balken: Röntgenklasse III, orange Balken: Röntgenklasse III–IV | Graph of number of radiological classes for the view „Zehe03“. Y-axis: radiological classification, X-axis: number of observers. Blue bar: class II–III, grey bar: class III, orange bar: class III–IV.

>5° im Hufgelenk als Überstreckung“ und •1.2.11 „Zehenchse / deutlich gebrochen >5° im Krongelenk als Überstreckung“ genannt (n=5). In Bezug auf das Hufgelenk wurde sechsmal der Befund •1.11.2 „Hufgelenk 90°/Konturveränderung an der Huf- oder Kronbeinfläche“ notiert, der eine hohe Einstufung in die RöKI III–IV hat.

In Bezug auf die Röntgenklassen wurde bei der Aufnahme „Zehe03“ in 48,65% der Fälle (n=18) der Endbefund „Röntgenklasse III–IV“ festgestellt, von 35% der Untersucher (n=13) der Endbefund der RöKI II–III und von 16% (n=6) der Endbefund der RöKI III (s. Abb. 2). Es wurden unterschiedliche Bemerkungen zu der Aufnahme von mehreren Untersuchern gemacht und vereinzelt Kontrollaufnahmen empfohlen. Die Befundungen spiegeln die geringe Übereinstimmung (unter 50%) der Untersucher bei dieser Aufnahme wider, so dass hier auch von einer Unsicherheit in Bezug auf die Befundung ausgegangen werden kann.

Auf der Röntgenaufnahme „Zehe07“ (s. Abb. 3) war die Zehe einer Vordergliedmaße abgebildet. Es wurden insgesamt 35 verschiedene RöKZ dokumentiert. Dabei wurden bis zu zehn



Abb. 3 „Zehe 07“, latero-mediale Röntgenaufnahme einer rechten Vorderzehe | Latero-medial view of a right front foot „Zehe07“.

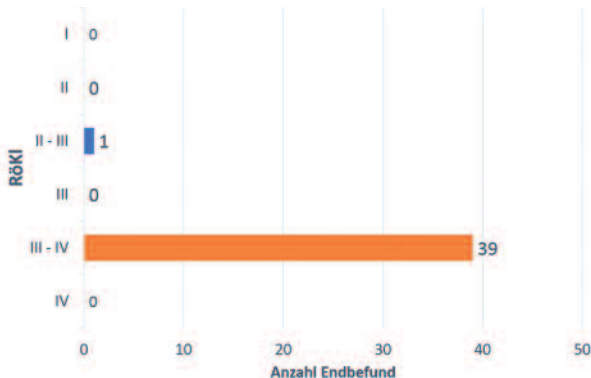


Abb. 4 Balkendiagramm über die Anzahl der Nennungen des jeweiligen Endbefundes für die Aufnahme „Zehe 07“ (s. Abb. 3). X-Achse: Anzahl der Untersucher, Y-Achse: Röntgenklassen. Blauer Balken: Röntgenklasse II-III, oranger Balken: Röntgenklasse III-IV | Graph distribution of radiological classes for the image „Zehe07“ (Fig. 3). Y-axis: radiological classification, X-axis: number of observers. Blue bar: class II-III, orange bar: class III-IV.

unterschiedliche RöKZ von einem Untersucher angeführt, und drei Untersucher stimmten bei der Kombination von 3 unterschiedlichen RöKZ überein. Die weiteren Kombinationen von mehreren RöKZ wurden zumeist nur von einem Untersucher ausgewählt. Es gab für den Befund „1.16.6 Gleichbeine 90°/Zubildungen an der palmaren/ plantaren Kontur (Fesselringband), mittel- bis hochgradig“ eine sehr gute Übereinstimmung von 90% (n=36). Dieser Befund ist in die RöKI III–IV eingestuft, so dass trotz der Vielzahl von ausgewählten Befunden für diese Aufnahme mit 97,5% der Untersucher (n=39) eine hohe Übereinstimmung für den Endbefund „Röntgenklasse III–IV“ festgestellt wurde. Bei dieser Aufnahme lag damit eine hohe Einigkeit bei der Erhebung des Endbefundes vor, der in eine hohe Röntgenklasse eingestuft wurde (RöKI III-IV, s. Abb. 4).



Abb. 5 „Oxspring06“, Röntgenaufnahme einer rechten Vorderzehe in der dorsoproximalen-palmarodistalen schrägen Aufnahmerichtung im Oxspring Klotz | Dorsoproximal-palmarodistal oblique view of a front foot named „Oxspring06“.

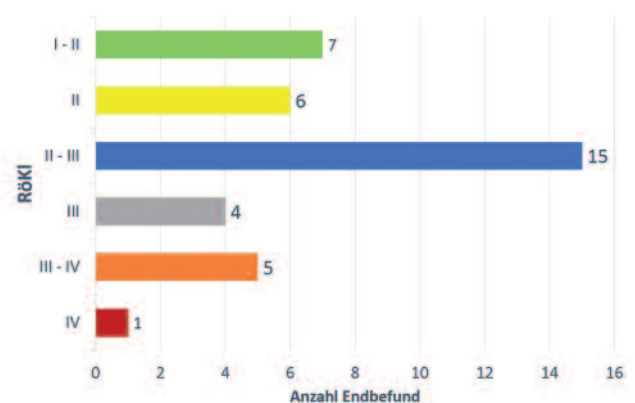


Abb. 6 Balkendiagramm über die Anzahl der Nennungen des jeweiligen Endbefundes für die Aufnahme „Oxspring06“ (s. Abb. 5). X-Achse: Anzahl der Untersucher, Y-Achse: Röntgenklassen. grüner Balken: Röntgenklasse I-II, gelber Balken: Röntgenklasse II, blauer Balken: Röntgenklasse II-III, grauer Balken: Röntgenklasse III, oranger Balken: Röntgenklasse III-IV und roter Balken: Röntgenklasse IV | Graph distribution of radiological classes for the view „Oxspring06“ (Fig. 5). Y-axis: radiological classification, X-axis: number of observers. Green bar: class I-II, yellow bar: class II, blue bar: class II-III, grey bar: class III, orange bar: class III-IV and red bar: class IV.

Ergebnisse der Röntgenaufnahme Oxspring 01–10

Für die Aufnahme „Oxspring06“ (s. Abb. 5) wurden 17 verschiedene Röntgenkennziffern notiert. Davon bezogen sich 10 auf die Region des Strahlbeines. Von 60,5% der Untersucher wurde die RökZ „2.1.10 „Strahlbein 0°/ Form der Canales schmal, spitz, breit, konisch oder zylindrisch“ der Röntgenklasse I–II gewählt. Die übrigen genannten Kennziffern wurden von ≤6 Untersuchern genannt, die zumeist zwei unterschiedliche RökZ für diese Aufnahme auswählten. Da eine hohe Übereinstimmung bei einem Röntgenbefund einer niedrigen RökI vorlag, ist die Übereinstimmung bei dem Endbefund des Röntgenbildes geringer. In Bezug auf den Endbefund dieses Röntgenbildes wurde fast jede der zur Verfügung stehenden Röntgenklasse genannt (s. Abb. 6). Viele Untersucher (38,5%) stellten einen Endbefund der RökI II–III.



Abb. 7 „Oxspring07“, Röntgenaufnahme einer Vorderzehe in der dorsoproximal-palmarodistalen schrägen Aufnahme im Oxspring Klotz | Dorsoproximal-palmarodistal oblique view of a front foot named „Oxspring07“.

Das Röntgenbild „Oxspring07“ (s. Abb. 7) wurde von 39 Untersuchern beurteilt, die 5 verschiedene RökZ nannten, die sich ausschließlich auf die Region des Hufbeines bzw. der Hufbeinknorpel bezogen. Eine hohe Übereinstimmung von 87,2% der Untersucher erhielt die RökZ •2.2.2 „Hufbein 0°/ Aufhellung (zystoider Defekt) Kontrollaufnahme empfohlen“ der Röntgenklasse IV. Zudem wurde von 33% der Untersucher der Befund •2.2.6 „Hufbein 0°/ Kontur Margo solearis zentral mit flacher Einziehung“ der Röntgenklasse II erhoben. Der Endbefund spiegelt diese hohe Übereinstimmung der RökZ direkt wider. Deshalb wurde auch in Bezug auf den Endbefund der RökI IV eine hohe Übereinstimmung von 87,2% der Untersucher festgestellt.

Tarsus

Bei der Röntgenaufnahme „Tarsus04“ (siehe Abb.8) wurden von 36 Untersuchern 13 verschiedene RökZ genannt. Zwei Untersucher forderten weitere Ebenen und notierten keine Befunde für diese Aufnahme. Zehn der aufgeführten Röntgenkennziffern bezogen sich auf die Intertarsalgelenke bzw. das

Tarsometatarsalgelenk und nur drei Befundkennziffern wurden von vier verschiedenen Untersuchern für das Talokruralgelenk dokumentiert. Es lag eine Häufung von 61% (n=22) der Untersucher für die RökZ •3.2.5 „Intertarsalgelenke und Tarsometatarsalgelenk / Randwülste und -zacken bis 2 mm (Osteophyten)“ der RökI II–III vor. Des Weiteren wurde auch der Befund mit der RökZ •3.2.3 „Intertarsalgelenke und Tarsometatarsalgelenk/verschwommene oder schmale Gelenkspalten“ von 12 Untersuchern genannt. Dieser Befund wird der RökI II zugeordnet. Die weiteren Befunde wurden jeweils nur einzeln beschrieben (bis zu drei Nennungen). Für den Endbefund lagen sowohl zwei bis drei Einzelnennungen für die RökI I, die RökI II, III, III–IV und die RökI IV vor, obwohl 60% der Untersucher den Endbefund der RökI II-III bei dieser Aufnahme erhoben. Zusammenfassend liegt für die Aufnahme „Tarsus04“ eine mehrheitliche Beurteilung für die mittlere



Abb. 8 „Tarsus04“, Röntgenaufnahme eines rechten Tarsus in der dorsolateral-plantaromedialen schrägen (45°) Aufnahme. | „Tarsus04“, dorsolateral-plantaromedial oblique view of a right tarsal joint.

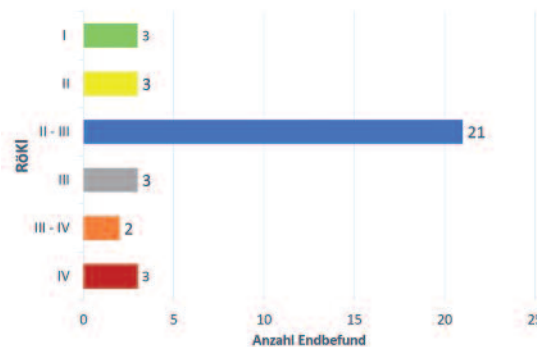


Abb. 9 Balkendiagramm über die Anzahl der Untersucher bezogen auf den Endbefund der Aufnahme „Tarsus04“ (s. Abb. 8). X-Achse: Anzahl der Untersucher, Y-Achse: Röntgenklassen. Grüner Balken: Röntgenklasse I, gelber Balken: Röntgenklasse II, blauer Balken: Röntgenklasse II-III, grauer Balken: Röntgenklasse III, oranger Balken: Röntgenklasse III-IV und roter Balken: Röntgenklasse IV | Bar diagram of radiological classes for the view „Tarsus04“ (Fig. 8). Y-axis: radiological classification, X-axis: number of observers. green bar: class I, yellow bar: class II, blue bar: class II-III, grey bar: class III, orange bar: class III-IV and red bar: class IV.



Abb. 10 „Tarsus08“, Röntgenaufnahme eines linken Tarsus in der latero-medialen Aufnahme-richtung, 90° | „Tarsus04“, latero-medial view of a left tarsal joint.



Abb. 11 „Tarsus13“, Röntgenaufnahme eines linken Tarsus in der dorso-plantaren Aufnahme-richtung, 0° | „Tarsus13“, dorso-plantar view of a left tarsal joint.



Abb. 12 Balkendiagramm der Untersucheranzahl bezogen auf den Endbefund der Aufnahme „Tarsus13“ (s. Abb. 11). X-Achse: Anzahl der Untersucher, Y-Achse: Röntgenklassen. gelber Balken: Röntgenklasse II, grauer Balken: Röntgenklasse III, oranger Balken: Röntgenklasse III-IV, roter Balken: Röntgenklasse IV | Bar diagram of radiological classes for the radiograph „Tarsus13“ (Fig. 11). Y-axis: radiological classification, X-axis: number of observers. yellow bar: class II, grey bar: class III, orange bar: class III-IV, red bar: class IV.

RöKL II-III und weitere Einzelbeurteilungen sowohl für höhere, als auch für niedrigere Röntgenklassen vor (s. Abb. 9).

Die Röntgenaufnahmen „Tarsus 06–10“ sind in der Projektionsrichtung 90–135° aufgenommen worden. Die folgende Aufnahme „Tarsus08“ (siehe Abb. 10) erhielt 15 unterschiedliche RöKZ, die sich zumeist auf die straffen Sprunggelenkreihen bezogen. Es lag eine sehr hohe Übereinstimmung von 95% für die •RöKZ 3.2.8 „Intertarsalgelenke und Tarsometatarsalgelenk/Randexostose am Mt III (Osteophyten größer 2 mm)“ der Röntgenklasse II–III vor. Die übrigen Befunde wurden jeweils von ≤10 der Untersucher erhoben. Obwohl in Bezug auf die Röntgenkennziffer in diesem Fall eine sehr hohe Übereinstimmung vorlag, spiegelt sich dies bei der Betrachtung des Endbefundes nicht in gleichem Maße wider. Es erhoben 60% der Untersucher den Endbefund der RöKL II–III und weitere 25% die RöKL III–IV für diese Aufnahme.

Die folgende Aufnahme „Tarsus13“ (s. Abb. 11) wurde von 15% der Untersucher als „obB“ bzw. in die Röntgenklasse I ohne eine weitere Nennung von Befunden eingestuft. Die übrigen Untersucher erhoben sieben verschiedene Befunde, davon stimmten 67,5% für die RöKZ •3.2.19 „unregelmäßige Sklerosierung proximal am Mt III, Insertionsdesmopathie M. interosseus medius“, ein Befund der Röntgenklasse III. Ein Drittel der Untersucher erhob den Befund mit der RöKZ •3.1.6 „Tibia Aufhellung (zystoider Defekt)“ der Röntgenklasse IV. Für den Endbefund lagen bei dieser Aufnahme folgende Nennungen vor: Röntgenklasse II (n=1), Röntgenklasse III (n=17) 53,1%, Röntgenklasse III-IV (n=1) Röntgenklasse IV (n=13) 40,6% vor, wobei die Hälfte der Untersucher für dieselbe RöKL gestimmt hat (s. Abb. 12). Von zwei Untersuchern wurde die nicht orthograde Darstellung der Intertarsalgelenke angemerkt und entsprechende Kontrollaufnahmen zur abschließenden Beurteilung gefordert.

Für die fünf verschiedenen latero-medialen Röntgenaufnahmen des Kniegelenkes (Knie 01–05) wurden zwischen 10 und 18 verschiedene RöKZ (im Durchschnitt 13,6) für jeweils eine Röntgenaufnahme dokumentiert. Im RÖLF07 sind insgesamt

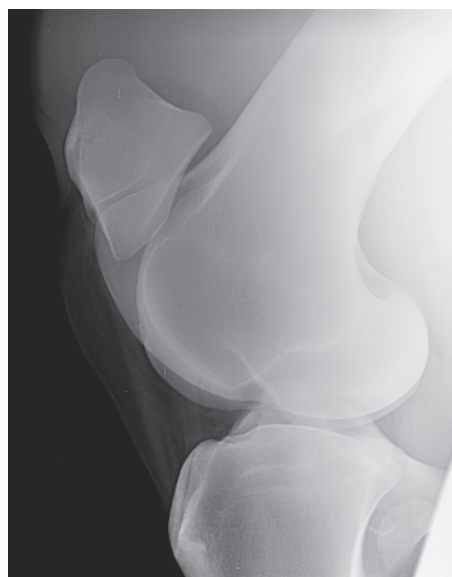


Abb. 13 „Knie01“, Röntgenaufnahme eines rechten Kniegelenkes in der latero-medialen Projektionsrichtung, ca. 90° | „Knie01“, latero-medial view of a right stifle joint.

37 verschiedene Röntgenkennziffern für diese Aufnahme aufgeführt. Dabei sind von einzelnen Untersuchern eine Kombination von bis zu sieben verschiedenen RöKZ genannt worden. Bei der Aufnahme „Knie01“ (siehe Abb. 13) haben 57,5% der Untersucher für die RöKZ •4.2.8 „Trochlea femoris (Knie 90–115°)/isolierte Verschattungen zwei oder mehr“ der RöKL III–IV gestimmt. Ähnliche Befundbeschreibungen, die ebenfalls die Dokumentation einer isolierten Verschattung berücksichtigen, wurden ebenfalls genannt. Es haben 30% der Untersucher die RöKZ •4.2.6 „Trochlea femoris (Knie 90–115°)/isolierte Verschattung ohne Abflachung der Kontur, größer als 2 mm“ der RöKL III gewählt und weitere 20% der Untersucher die RöKZ •4.2.7 „Trochlea femoris (Knie 90–115°)/isolierte Verschattung mit Abflachung der Kontur“ der RöKL III–IV. Wenige Nennungen (10%) lagen für die RöKZ •4.2.5 „Trochlea femoris (Knie 90–115°)/isolierte Verschattung ohne Abflachung der Kontur, kleiner als 2 mm“ vor, die in die Röntgenklasse II–III eingeordnet ist. Endbefunde lagen bei der Aufnahme „Knie01“ zwischen RöKL I–II und IV, wobei 67,5% der Untersucher diese Aufnahme in die RöKL III–IV einstuften, 20% die RöKL III und 2,5% die Röntgenklasse IV dokumentierten.

Die Aufnahmen „Knie 06–10“ wurden in der kaudo-kranialen Projektionsrichtung (0°) angefertigt. Bei dem Röntgenbild „Knie07“ (siehe Abb. 14) wurden sieben verschiedene RöKZ genannt und es stimmten fast alle Untersucher (97,6%) für den Befund der RöKZ •4.5.6 „Knie/Kniegelenk/medialer Femurkondylus Aufhellung (Zystoide Defekt)“ der RöKL IV. Eine weitere deutliche Übereinstimmung von 87,8% wurde bei der RöKZ •4.5.12 „Knie/Kniegelenk/Tibia Zubildung medialer Kondylus“ der RöKL II–III festgestellt. Entsprechend hoch war deshalb auch die Übereinstimmung von 97,6% bei dem Endbefund der Röntgenklasse IV bei dieser Aufnahme „Knie07“, die aufgrund der zugeordneten Röntgenklasse exakt der am häufigsten genannten RöKZ entspricht.

Für die fünf verschiedenen Röntgenaufnahmen der Dornfortsätze des Rückens (Rücken 01–05) wurden zwischen 8 und 10 verschiedene RöKZ für jeweils eine Röntgenaufnahme



Abb. 14 „Knie07“, Röntgenaufnahme eines rechten Kniegelenkes in der kaudo-kranialen Projektionsrichtung, ca. 0° | „Knie07“, caudo-cranial view of a right stifle joint.

dokumentiert. Dabei sind von einzelnen Untersuchern eine Kombination von bis zu vier verschiedenen RöKZ genannt worden.

Bei der Aufnahme „Rücken03“ (Abb. 15) wurde von 90% der Untersucher dieselbe RöKZ benannt: •5.2.6 „Berühren der Dornfortsätze mit starker Sklerosierung und / oder Zubildung“ der Röntgenklasse III–IV. Entsprechend lag für diese Aufnahme deshalb auch eine sehr hohe Übereinstimmung in Bezug auf den Endbefund von 97,5% für die RöKL III–IV vor. Das Beispiel „Rücken05“ (Abb. 16) erhielt für die RöKZ •5.2.7 „Berühren der Dornfortsätze mit zystoiden Defekten“ der RöKL III–IV eine hohe Übereinstimmung von 87,8% der Untersucher. Als zweithäufigster Befund wurde von 29,3% die RöKZ •5.2.4 „Zwischenräume 2 bis 8 mm mit reaktiven Veränderungen (Sklerosierungssaum, Zubildung)“ der RöKL II–III genannt. Ähnlich dem erst genannten Befund wurde die RöKZ •5.2.6 „Berühren der Dornfortsätze mit starker Sklerosierung und/oder Zubildung“, die ebenfalls der RöKL III–IV zugeordnet ist, von 22% der Untersucher dokumentiert.

Von jedem Untersucher wurden eine oder mehrere der folgenden RöKZ angegeben, die ähnliche Befunde beschreiben und alle in die RöKL III–IV eingestuft wurden: •5.2.6 Berühren der Dornfortsätze mit starker Sklerosierung und/oder Zubildung, •5.2.7 Berühren der Dornfortsätze mit zystoiden Defekten oder •5.2.8 Dornfortsatz mit zystoidem Defekt. Deshalb ist bei dieser Röntgenaufnahme eine 100%ige Übereinstimmung der Untersucher für den Endbefund der RöKL III–IV festgestellt worden.

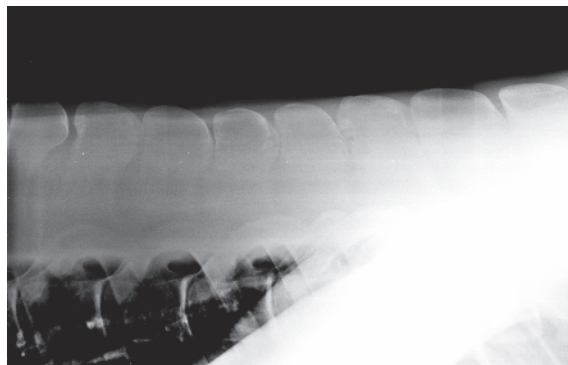


Abb. 15 „Rücken03“, Röntgenaufnahme der Dornfortsätze der caudalen Brustwirbelsäule | „Rücken03“, latero-lateral view of the spinous processes of the caudal thoracic spine.

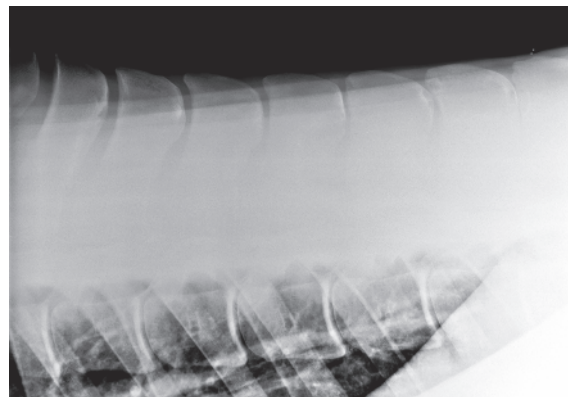


Abb. 16 „Rücken05“, Röntgenaufnahme der Dornfortsätze der Brustwirbelsäule | „Rücken05“, latero-lateral view of the spinous processes of the mid thoracic spine.

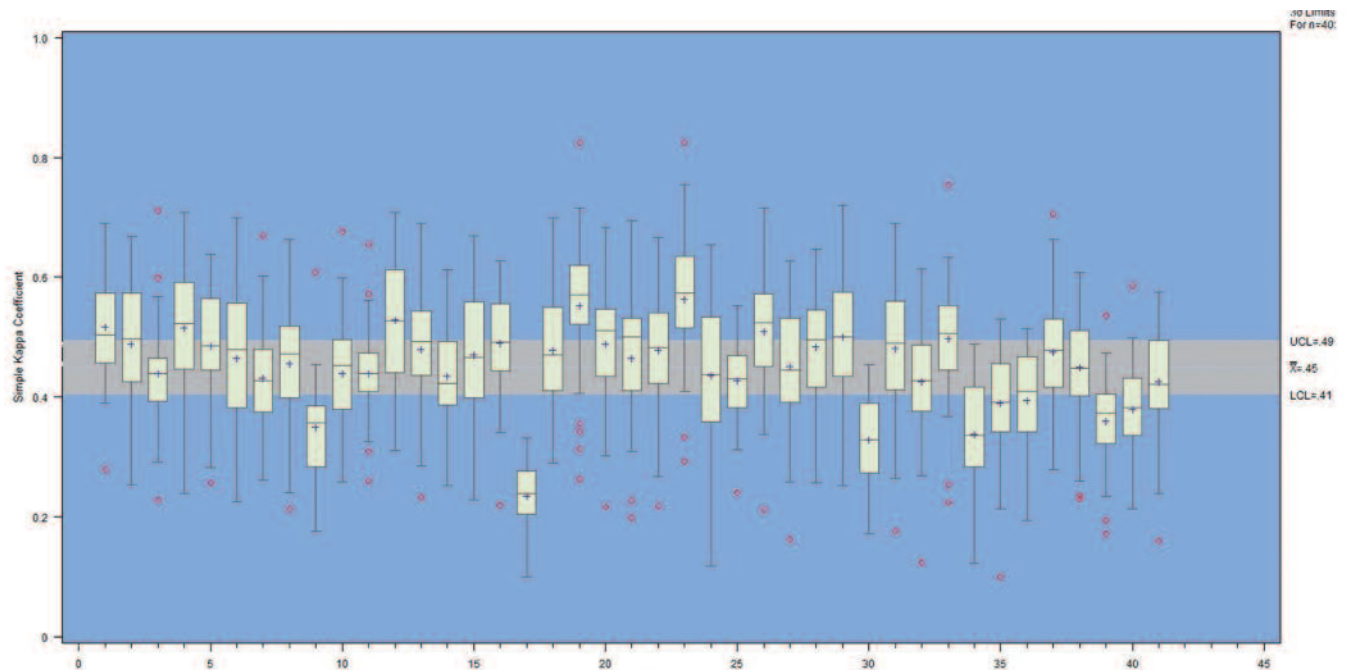


Abb. 17 Box plot des Fleiss-Choen Kappa-Wertes eines jeden Untersuchers im Vergleich zu allen übrigen Untersuchern bezogen auf den Endbefund der fünfzig Röntgenbilder | Box plot of Fleiss-Chohen weighted kappas for each rater for final classification of fifty radiographic images.

Ergebnis der Übereinstimmung aller Untersucher in Bezug auf den Endbefund aller Röntgenbilder

Für die Berechnung der PABAK-Werte wurden die Röntgenklassen des Endbefundes aller fünfzig Röntgenaufnahmen jedes einzelnen Untersuchers paarweise mit den Befunden der übrigen Untersucher verglichen, um zu ermitteln, ob bzw. in welchem Maß eine Übereinstimmung der 41 Untersucher zutrifft. Diese 40 Werte jedes einzelnen Untersuchers in Bezug auf die anderen Untersucher. Zusammenfassend wurde in der vorliegenden Arbeit eine gute Übereinstimmung (PABAK: 0,72) der vierzig Untersucher in Bezug auf die Röntgenklasse (Endbefund) der fünfzig Aufnahmen festgestellt.

Diskussion

Die vorliegende Studie dokumentiert erstmalig die Befundung mehrerer Röntgenbilder von verschiedenen Pferden nach dem RÖLF07 durch viele erfahrene Untersucher. Entgegen einer üblichen Beurteilung von Röntgenbildern im Rahmen einer Kaufuntersuchung wurde den Untersuchern in der vorliegenden Studie lediglich eine Aufnahme pro Pferd zur Verfügung gestellt. Deshalb war es nicht möglich, bei fraglichen Befunden eine andere Ebene desselben Gelenkes (z.B. beim Tarsus) zum Vergleich heran zu ziehen. Bei der Aufnahme „Tarsus 13“ lag die Übereinstimmung in Bezug auf den Endbefund bei < 50% und es wurden jegliche Röntgenklassen als Endbefund für diese Aufnahme angeführt. Mehrere Projektionsrichtungen des Tarsus erhöhen hingegen die diagnostische Sicherheit (Eksell et al. 1999) so dass dadurch eine höhere Übereinstimmung zwischen mehreren Untersuchern zu erwarten wäre.

Ebenfalls wurden unterschiedliche Bemerkungen zu der Aufnahme „Zehe03“ von mehreren Untersuchern gemacht und vereinzelt Kontrollaufnahmen empfohlen. Dies spiegelt die geringe Übereinstimmung (< 50%) der Untersucher bei die-

ser Aufnahme wider, so dass hier auch von einer Unsicherheit in Bezug auf die Befundung ausgegangen werden kann.

Grundsätzlich wurden den Untersuchern jedoch ähnliche Bedingungen bei der Interpretation der Röntgenbilder geschaffen, die auch bei der üblichen Beurteilung in der täglichen Praxis vorliegen. Es wurde auf ein Zeitlimit verzichtet, um keine prüfungs-ähnliche Situation zu schaffen und Stress zu vermeiden. In einer humanmedizinischen Untersuchung führte eine überlange Untersuchungszeit zu erhöhten falsch positiv wie falsch negativen Ergebnissen (Kundel et al. 1989). In der vorliegenden Studie wurde auf eine Auswertung der Untersuchungszeit verzichtet, da es in diesem Studiendesign keinen „gold standard“ gab und aufgrund dessen keine Untersuchung in Hinblick auf falsch positive oder falsch negative Ergebnisse durchgeführt werden konnte.

Die vorliegende Studie sollte einer Röntgenbeurteilung nach dem RÖLF07 ohne Wissen über mögliche klinische Befunde entsprechen. Diese Voraussetzung ist auch in der Präambel des RÖLF verankert, nachdem die „Einteilung in die Röntgenklassen nur an die röntgenologischen Befunde gebunden“ (RÖLF07) ist. Deshalb ging der Untersucher unvoreingenommen an die Beurteilung der einzelnen Röntgenaufnahmen heran.

Ein Begriff in der radiologischen Befundinterpretation ist die „satisfaction of search“ (Berbaum et al. 1990), die eine Erklärung für das Übersehen von Befunden liefert: wenn Röntgenbilder mit dem Hintergrundwissen über das Krankheitsbild des Patienten beurteilt werden, ist der Untersucher häufiger mit einem Hauptbefund, der die klinische Symptomatik erklärt, zufrieden und erhebt bzw. übersieht mögliche weitere Befunde. Dieses Phänomen wurde in der Humanmedizin auch bei der Interpretation von Röntgenbildern des Skelettsystems festgestellt, insbesondere, wenn zwei bzw. drei Befunde pro Röntgenaufnahme vorlagen (Ashman et al. 2000). Für die Röntgenbeurteilung nach dem RÖLF kann dieses Kriterium

nur bedingt bzw. abgewandelt auf die Untersucher angewandt werden. Da der Röntgenbefund mit der höchsten Röntgenklasse den Endbefund der Röntgenaufnahme darstellt, ist die „satisfaction of search“ womöglich erreicht, wenn ein Befund erhoben wird, der einer hohen Röntgenklasse zugeordnet ist. Die Relevanz von anderen Befunden tritt ab diesem Moment in den Hintergrund.

In der vorliegenden Studie ist es retrospektiv nicht möglich zu ergründen, ob eine solche Motivation bzw. Ausrichtung ausschlaggebend für die Beurteilung der Röntgenbilder war. Es ist anzunehmen, dass jeder einzelne Untersucher jedoch motiviert war, jeden vorhandenen Befund zu benennen. Diese Frage muss jedoch offenbleiben.

Bei auffälligen Befunden (z.B. 4.5.6 „Knie/Kniegelenk/medialer Femurkondylus Aufhellung – zystoider Defekt“), die in einer hohen RöKI resultieren (z.B. RöKI IV) ist häufig eine hohe Übereinstimmung sowohl in Bezug auf die RöKZ als auch auf den Endbefund vorhanden. Sehr hohe Übereinstimmungen für den Befund einer RöKZ von 90% wurden auch bei der Aufnahme „Rücken 03“ erzielt. Der Befund „5.2.6 „Berühren der Dornfortsätze mit starker Sklerosierung und/oder Zubildung“ wird in die RöKI III–IV eingeordnet. Anhand dieser Beispiele zeigt sich, dass offensichtlichere Röntgenbefunde und die dazu gehörige Röntgenkennziffer häufiger mit einer besseren Übereinstimmung erhoben werden. Im Allgemeinen werden Fälle, bei denen deutliche röntgenologische Befunde vorliegen, mit einer großen Sicherheit beurteilt und als „einfache Fälle“ empfunden. Deshalb herrscht bei den „einfachen Fällen“ meist eine höhere Übereinstimmung als bei Befunden, die als „schwierige Fälle“ eingestuft werden und bei denen auch die Experten unterschiedlicher Meinung sind (Robinson 1997). Dies wurde in der vorliegenden Studie betätigt. Die prozentuale Übereinstimmung auf einen Endbefund war umso größer, wenn die entsprechende Röntgenaufnahme in eine hohe Röntgenklasse eingestuft wurde. Bei der Röntgenaufnahme „Zehe 07“ wurde z.B. eine Vielzahl von Einzelbefunden aufgeführt, jedoch stellten auch 90% denselben Befund (Gleichbeine 90°/Zubildungen an der palmaren/plantaren Kontur (Fesselringband), mittel- bis hochgradig) fest, der aufgrund seiner Einstufung in die RöKI III–IV als deutlicher und relevanter Befund gelten kann und deshalb möglicherweise auch als gut bzw. einfach erkennbar von den Untersuchern empfunden wurde. Dieser Befund stellte bei dieser Aufnahme den Befund der höchsten RöKI dar und führt zu der sehr hohen Übereinstimmung von 97,5% bei dem Endbefund dieser Aufnahme. Lediglich ein Untersucher vergab für diese Aufnahme den Endbefund der RöKI II-III. Dieses Beispiel zeigt, dass auch bei scheinbar offensichtlichen und „einfachen“ Befunden keine vollständige Übereinstimmung zwischen vielen erfahrenen Untersuchern auftritt. Das entspricht den Erfahrungen aus in der Praxis und in gerichtlichen Auseinandersetzungen, die häufig ausschließlich wegen eines strittigen Röntgenbefundes geführt werden, ohne dass das streitgegenständliche Pferd klinisch auffällig ist.

Eine hohe Übereinstimmung zeigte sich bei der Befundung von isolierten Verschattungen im Fesselgelenk (sowohl dorsal als auch palmar) auf der seitlichen Aufnahme der Zehen. Hier erhoben 60 bis 95% der Untersucher diesen Befund. Möglicherweise hängt dies mit der Bedeutung der isolierten Fragmente für die Kaufuntersuchung bzw. das besondere Augen-

merk bzw. das geschulte Auge des Untersuchers für diese Befunde zusammen, da ein isoliertes Fragment im Fesselgelenk sowohl im Rahmen einer orthopädischen Untersuchung als auch im chirurgischen Bereich einen häufigen Befund darstellt. Auch für den Laien (Käufer bzw. Verkäufer) ist ein solcher Befund nachvollziehbar und hat einen entsprechenden Wiedererkennungswert.

In der vorliegenden Studie wurden für die Aufnahmen des Tarsalgelenkes zwischen 5 bis 18 (ø 10 RöKZ) verschiedene RöKZ angegeben. Bei einer anderen Untersuchung (Bach et al. 2012) wurden bei der Befundung von Aufnahmen der Tarsalgelenke von 139 Pferden nach dem RøLF07 nur 8 verschiedene RöKZ erhoben. Die Untersuchung von Bach et al. erwähnt als häufigsten Befund im Talokruralgelenk die RöKZ 3.1.12 „Talus distal gerichtete Zubildung“ und bei den Intertarsalgelenken 7 von 22 möglichen Befunden (RöKZ), am häufigsten den Befund 3.2.7 (28,8%) „Randexostose am Mtl-II-Osteophyten bis 2 mm“.

In der vorliegenden Studie wurde für die Region des Talokruralgelenkes am häufigsten die RöKZ 3.1.5 „Tibia isolierte Verschattung, malazisches Dissekat oder mehrere“ und bei den straffen Sprunggelenkreihen die RöKZ 3.2.9 „Strukturauflösung und Usuren in den Gelenkspalten“ gefolgt von der RöKZ 3.2.3 „verschwommene oder schmale Gelenkspalten“ am häufigsten als Befundbeschreibung genannt.

Diese Diskrepanzen könnten darin begründet sein, dass in der vorliegenden Studie Röntgenbilder verwendet wurden, die nicht nur im Rahmen von Kaufuntersuchungen, sondern auch von klinischen Patienten im Rahmen einer Lahmheitsuntersuchung angefertigt wurden und deshalb auch Befunde von klinischer Relevanz beinhalteten.

Dennoch ist die Übereinstimmung der Untersucher für den Befund „Tarsometatarsalgelenk/Randexostose am Mt III (Osteophyt größer 2 mm)“ der Aufnahme „Tarsus 08“ in der vorliegenden Studie deutlich höher, als die lediglich leichte Übereinstimmung zwischen vier Untersuchern ($\kappa=0,28$), die in der Literatur für diesen Befund beschrieben ist (Jackson et al. 2013).

In Bezug auf die Interpretation von röntgenologischen Befunden im Bereich der straffen Sprunggelenke stellten Labens et al. (2007) eine deutlich bessere Zuverlässigkeit bei der Befundung von geringgradigen und hochgradigen gegenüber mittelgradigen Befunden fest. Dieses Ergebnis wird von der vorliegenden Studie bestätigt, in der die Übereinstimmung bei hochgradigen Befunden der straffen Sprunggelenkreihen (z.B. Tarsus 11) sehr hoch war.

Auch bei der Befundung von OCD des Tarsus wurde in der vorliegenden Studie bei der dorsolateral – plantaromedialen schrägen Projektionsrichtung (45°) der Aufnahme „Tarsus 03“ von 100% der Untersucher eine exzellente Übereinstimmung für den Befund der isolierten Verschattung festgestellt, hingegen zeigt eine andere Untersuchung eine mittelmäßige Übereinstimmung für diesen Befund durch vier Untersucher (Jackson et al. 2013).

Die Aufnahme „Knie 01“ zeigt die latero-mediale Projektionsrichtung eines Kniegelenkes, welches eine OCD Läsion an

der lateralen Trochlea aufweist. Eine isolierte Verschattung wurde insgesamt von 35 der 40 Untersucher (87,5%) erkannt, jedoch verschiedenen RökZ zugeordnet. Im RölF07 sind vier verschiedene RökZ (4.2.5–4.2.8) aufgeführt, die eine isolierte Verschattung des lateralen Rollkammes beschreiben, jedoch unterschiedliche Begleiterscheinungen (mit oder ohne Abflachung der Kontur, Größe und Anzahl der isolierten Verschattungen) aufweisen. So reicht auch die Klassifizierung dieser unterschiedlichen Befundinterpretationen von der RökI II–III bis zur RökI III–IV. Von einigen Untersuchern wurden weitere Befunde (Abflachung, Einkerbung bzw. Strukturveränderungen) an den Rollkammern erhoben, die ebenfalls als Hinweis auf eine OC gelten. Zusammenfassend wurde von 97,5% der Untersucher Hinweise auf OC bzw. OCD bei dieser Aufnahme dokumentiert, die sich jedoch nicht in der Gesamtklassifizierung der Röntgenaufnahme widerspiegelt. Diese sehr gute Übereinstimmung der vierzig teilnehmenden Untersucher deckt sich mit einer Angabe aus der Literatur, die bei der Beurteilung von Röntgenaufnahmen des Kniegelenkes auf der latero-medialen Aufnahme zwischen zwei Untersuchern eine sehr gute Übereinstimmung ($\kappa = 0,86$) in Bezug auf die Erkennung von OC Läsionen feststellten (Beccati et al. 2013). In einer weiteren Studie wurde lediglich eine mittelmäßige Übereinstimmung von vier Untersuchern bei der Erhebung des Befundes „OCD an der Trochlea des Femurs“ festgestellt (Jackson et al. 2013).

Die latero-mediale Aufnahme „Knie04“ wurde von 44% der Untersucher in die RökI IV eingeordnet und von 38,9% in die RökI I. Der Befund, der in diesem Fall zu der Klassifizierung in die hohe RökI IV führte, war ein zystoider Defekt im Femurkondylus. Dieser Befund ist auf der latero-medialen Aufnahme des Kniegelenkes nicht offensichtlich, sondern erst nach Veränderung von Kontrast und Helligkeit als solcher erkennbar. Dies könnte dazu geführt haben, dass dieser Befund von 38,9% der Untersucher nicht als solcher erkannt wurde. Lediglich 28% haben den Befund des zystoiden Defektes im Femurkondylus auf der latero-medialen Aufnahme erhoben und der entsprechenden RökZ zugeordnet. Es ist deshalb möglich, dass diese „Fehlinterpretation“ aufgrund der kontrastarmen Röntgenaufnahme und damit technisch bedingt oder aufgrund einer „falschen Wahrnehmung“ (Robinson 1997) beruht.

Jedem Untersucher war es möglich, die einzelnen Röntgenaufnahmen in Helligkeit und Kontrast zu verändern, da diese in einem DICOM Format vorlagen und mit einem DICOM Viewer präsentiert wurden. Es bleibt offen, ob jeder Untersucher für die genannte Aufnahme diese Möglichkeit entsprechend genutzt hat.

Auf der kaudo-kranialen Projektionsrichtung wurde in der vorliegenden Studie auf der Röntgenaufnahme „Knie07“ hingegen von 97,6% der Untersucher ein zystoider Defekt im medialen Kondylus des Femurs festgestellt. Die Übereinstimmung der vierzig Untersucher in der vorliegenden Studie bei der Befundung der zystoiden Läsion im medialen Femurkondylus ist deutlich besser als die mittelmäßige bis gute Übereinstimmung, die in der Literatur durch vier spezialisierte veterinärmedizinische Radiologen für diesen Befund vorliegen (Jackson et al. 2013).

Dieses Ergebnis macht deutlich, dass es insbesondere bei der röntgenologischen Darstellung des Kniegelenkes ratsam ist,

die zweite Aufnahme in der kaudo-kranialen Aufnahmerichtung anzufertigen, um mögliche zystoide Defekte im medialen Femurkondylus zweifelsfrei darstellen zu können. Der gravierende Befund eines zystoiden Defektes im medialen Femurkondylus, der in die Röntgenklasse IV eingestuft ist, kann kaufentscheidend sein. Eine Kontrollaufnahme bzw. eine Aufnahme in der kaudo-kranialen Projektionsrichtung wurde jedoch nur von zwei Untersuchern in dem Freitextfeld „Bemerkungen“ der latero-medialen Aufnahme „Knie04“ gefordert.

Eine vollständige Übereinstimmung aller Untersucher in Bezug auf den Endbefund lag bei der Aufnahme „Rücken05“ vor. Dies könnte damit begründet sein, dass für die Röntgenaufnahmen des Rückens bzw. der Dornfortsätze (Unterpunkt 5. des RölF) insgesamt nur 22 verschiedene Röntgenkennziffern angegeben sind. Im Gegensatz dazu sind der Aufnahme „Zehe 90°“ allein 117 verschiedene Röntgenkennziffern zugeordnet. Deshalb führt möglicherweise eine reduziertere Auswahl an Befunden in der vorliegenden Studie zu einer besseren Übereinstimmung der Untersucher.

Eine weitere Erklärung für die sehr gute Übereinstimmung der Untersucher bei dem Endbefund der Aufnahme „Rücken05“ könnte jedoch auch darin begründet sein, dass das Augenmerk lediglich auf die Dornfortsätze und die Wirbelkörper der Brustwirbelsäule konzentriert ist.

Entgegen der übrigen Aufnahmen, bei denen unterschiedliche knöcherne Strukturen abgebildet sind, wiederholt sich bei den Rücken-aufnahmen die Darstellung der Dornfortsätze und die Beurteilung ist primär auf den Abstand bzw. die Kontur und Struktur der Dornfortsätze gerichtet.

Diese sehr gute Übereinstimmung der Untersucher in Bezug auf die Befundung der Dornfortsätze wurde auch in einer anderen Studie festgestellt, an der jedoch lediglich zwei Untersucher teilnahmen (Cousty et al. 2010). Die Befundung von sich berührenden Dornfortsätzen (kissing spines) scheint auf der Aufnahme „Rücken05“ ein „einfacher“ Befund (Robinson 1997) zu sein, weil dieser von allen Untersuchern mit hoher Sicherheit erkannt wurde.

Eine gute „inter-observer“ Übereinstimmung von drei Untersuchern ($\kappa 0,7$ bis $0,85$) wurde für den Befund „Osteophyt proximaler Gelenkrand“ in der Literatur dokumentiert. Keine gute Übereinstimmung lag hingegen bei Canales im proximalen Strahlbeinrand zwischen drei Untersuchern ($\kappa 0,34$ und $0,43$) vor. Die Bewertung einzelner Befunde wurde deshalb als sehr variabel beschrieben (Groth et al. 2009). Beispielhaft dafür steht die Aufnahme „Oxspring06“ in der vorliegenden Studie, für die 17 verschiedene Röntgenkennziffern angegeben wurden.

Für den Befund einer isolierten Verschattung am distalen Strahlbeinrand wurde in der Studie von Groth et al. (2009) durch lediglich drei Untersuchern eine leichte Übereinstimmung ($\kappa 0,38$) festgestellt. Dieser Befund ist in der vorliegenden Studie von 58,5% der 40 Untersucher erhoben worden.

Bei der Aufnahme „Oxspring06“ lagen Nennungen für jede RökI bei dem Endbefund vor. Diese breite Streuung verdeutlicht, dass insbesondere die Oxspringaufnahme Schwierigkeiten bei der Befundinterpretation birgt, dieses aber von den

Untersuchern im vorliegenden Fall nicht in diesem Maße empfunden wurde. Denn gerade bei diesen Aufnahmen sind in dem Freitextfeld keine Bemerkungen oder Angaben zum Schwierigkeitsgrad gemacht worden.

Zusammenfassend ist festzustellen, dass in Bezug auf den Endbefund der Röntgenklasse eine gute Übereinstimmung von den vierzig erfahrenen Untersuchern für fünfzig verschiedene Röntgenaufnahmen in der vorliegenden Studie vorlag. Für einige Einzelbefunde wurde sowohl bei Aufnahmen des Kniegelenkes als auch des Tarsus bessere Übereinstimmungen durch die vierzig Untersucher der vorliegenden Studie erzielt, als durch wenige Untersucher anderer Studien. Die Übereinstimmungen waren umso besser, je höher die jeweilige Röntgenklasse war, so dass deutliche Befunde, bei denen eine klinische Relevanz wahrscheinlich ist, selten zu Rechtsstreitigkeiten führen sollten. Bei undeutlichen Befunden muss jedoch weiterhin mit einer individuellen Interpretation und deren Zuordnung zu verschiedenen RÖKZ gerechnet werden.

Eine Reduktion der Röntgenklassen würde zu einer höheren Übereinstimmung führen und der Unsicherheit in Bezug auf die Definition der Röntgenklassen gerechter werden.

Danksagung

Allen Kolleginnen und Kollegen, die als Untersucher an dieser Studie beteiligt waren, sei ganz herzlich für Ihre Mitarbeit gedankt.

Literatur

- Ashman C. J., Yu J. S., Wolfman D. (2000) Satisfaction of Search in Osteoradiology. *Am. J. Rad.* 175, 541- 544; DOI 10.2214/ajr.175.2.1750541
- Altman D. G. (1990) Inter-rater agreement In: *Practical statistics for medical research.* Altman D. G., Chapman and Hall, London, 1st Ed, 403-406
- Bach S., Plückhahn N., Stadtbäumer G., Hertsch B. W. (2012) Röntgenologische Untersuchungen und Nachuntersuchungen der Tarsi von 139 Warmblutpferden in Zeitabständen von zehn bis 40 Monaten und Klassifizierung der jeweiligen Befunde nach dem Röntgenleitfaden 2007 (RöLF 07). *Prakt. Tierarzt* 93, 130-139
- Beccati F., Chalmers H. J., Dante S., Lotto E., Pepe M. (2013) Diagnostic sensitivity and interobserver agreement of radiography and ultrasonography for detecting trochlear ridge osteochondrosis lesions in the equine stifle. *Vet. Radiol. Ultrasound* 54, 176- 184; DOI:10.1111/vru.12004.
- Bemmann K., Stadler P., Schüle E. (2008) Die Änderung des Röntgenleitfadens und ihre Bedeutung für die tierärztliche Praxis und Forensik. *Recht d. Landwirtschaft* 60, 141-148
- Berbaum K. S., Franken E. A., Dorfman D. D., Rooholamini S. A., Kathol M. H., Barloon T. J., Behlke F. M., Sato Y., Lu C. H., El-Khoury G. Y., Flickinger F. W., Montgomery W. J. (1990) Satisfaction of Search in Diagnostic Radiology. *Invest. Radiol.* 25, 2, 133-140

- Bladon B. M., Main J. P. M. (2003) Clinical evidence in the evaluation of presale radiography: are we in a desert on a horse with no name? *Equine Vet. J.* 35, 341-342; DOI: 10.2746/042516403776014154
- Brady A., Laoide R. O., McCarthy P., McDermott R. (2012) Discrepancy and Error in Radiology: Concepts, Causes and Consequences. *Ulster Med. J.* 81, 3-9; PMID: PMC3609674
- Byrt T., Bishop J., Carlin J. B. (1993): Bias, prevalence and kappa. *J. Clin. Epidemiol.* 46, 423-429; DOI: 10.1016/0895-4356(93) 90018-V
- Cohen J. (1960) A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* 20, 37-46
- Cousty M., Retureau C., Tricaud C., Geffroy O., Caure S. (2010) Location of radiological lesions of the thoracolumbar column in French trotters with and without signs of back pain. *Vet. Rec.* 166, 41-45; DOI: 10.1136/vr.c70
- Eksell P., Uhlhorn H., Carlsten J. (1999) Evaluation of different projections for radiographic detection of tarsal degenerative joint disease in Icelandic Horses. *Vet. Radiol. Ultrasound* 40, 228-232
- Engl C. A., Sychterz C. J., Young A. M., Pollock D. C., Toomey S. D., Engl C. A. (2002) Interobserver and intraobserver variability in radiographic assessment of osteolysis. *J. Arthroplasty* 17, 752-759; DOI:10.1054/arth.2002.33554
- Fleiss J. L., Levin B., Paik M. C. (2003) The measurement of interrater agreement In: *Statistical Methods for Rates and Proportions*, Wiley, Hoboken, 3rd Ed., 598-626
- Flight L., Julious S. A. (2015): The disagreeable behavior of the kappa statistic. *Pharmaceutical Statistics* 14, 1, 74-78; DOI: 10.1002/pst.1659
- Gerhards H., Hertsch B. W., Jahn W., Brunken G. (2008) Leitfaden für die röntgenologische Beurteilung bei der Kaufuntersuchung des Pferdes (RöLF07). *Deutsch Tierärzteblatt* 56, 292-303
- Groth A. M., May S. A., Weaver M. P., Weller R. (2009) Intra- and interobserver agreement in the interpretation of navicular bones on radiographs and computed tomography scans. *Equine Vet. J.* 41, 124-129; DOI: 0.2746/042516408X345125
- Jackson M. A., Vizard A. L., Anderson G. A., Mattoon J. S., Lavelle R. B., Smithenson B. T., Lester N. V., Clarke A. F., Whitton R. C. (2013) An assessment of intra-and interobserver agreement of reporting orthopaedic findings on presale radiographs of Thoroughbred yearlings. *Equine Vet. J.* 46, 567-574; DOI: 10.1111/ evj.12150
- Kundel H. L., Nodine C. F., Krupinski E. A. (1989) Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions. *Invest. Radiol.* 24, 472-478
- Labens R., Innocent G. T., Voute L. C. (2007) Reliability of a quantitative rating scale for assessment of horses with distal tarsal osteoarthritis. *Vet. Radiol. Ultrasound* 48, 204-211; DOI: 10.1111/ j.1740-8261.2007.00230.x
- Lindenhovius A., Karanicolas P. J., Bhandari M., Ring D. (2012) Radiographic Arthrosis After Elbow Trauma: Interobserver Reliability. *J. Hand Surg.* 37, 755-759; DOI: 10.1016/j.jhsa.2011.12.043
- Plewa D. (2014) Dritthaftung des Tierarztes bei der vom Verkäufer in Auftrag gegebenen Kaufuntersuchung. *Pferdeheilkunde* 30, 188-190; DOI: 10.21836/PEM20140206
- Robinson P. J. A. (1997) Radiology s Achilles heel: error and variation in the interpretation of the Röntgen image. *Br. J. Radiol.* 70, 1085-1098; DOI: 10.1259/bjr.70.839.9536897
- Stadler P. (2005) Röntgenbefundung bei Ankaufuntersuchungen – Anspruch und Wirklichkeit. *Pferdespiegel* 1, 51-53
- Zöttl B., Rohn K., Stadler P. (2009) Röntgenleitfaden 07 – Einheitlichkeit in der tierärztlichen Befundung? 18. Tagung über Pferdekrankheiten im Rahmen der EQUITANA, wak Verlag, Gescher, 81-86