**Title**

**Agreement of Grading Radiographs of Thoracolumbar Spinous Processes in Healthy Horses on Pre-Purchase Examination**

Running head: Agreement of back radiographs in healthy horses

M. G. P. Looijen1, K. Hanousek1, J.T.M. Maree2, M.C. Visser2, D. Berner1

1 Equine Referral Hospital, Royal Veterinary College, University of London, Hawskhead Lane, Hatfield, AL9 7TA, United Kingdom

2 Evidensia Veterinair Centrum Honselersdijk, Harteveldlaan 3, Honselersdijk, 2675 LE, the Netherlands

\*Corresponding author e-mail: [mlooijen@rvc.ac.uk](mailto:mlooijen@rvc.ac.uk), telephone number: +44 1707 666297

Keywords**:** equine, back, without clinical symptoms, inter-observer, intra-observer, x-rays.

Word count: 3706

**ABSTRACT:**

**Background**: There is wide variability in radiographic grading systems in thoracolumbar spinous processes in horses. The aim of this study was to determine agreement of grading different parameters, and combinations of those, in radiographs of the spinous processes of the equine thoracolumbar spine in the scope of prepurchase examination. We hypothesized that agreement is variable and interpretation of these radiographs is difficult.

**Methods**: Radiographs of the thoracolumbar spine of healthy horses (N=100) were assessed by three observers. Spinous processes were separately graded for interspinous space width, modelling, radiopacities, radiolucencies and isolated opacities dorsally. Inter- and intra-observer agreement was assessed for individual and combinations of parameters.

**Results**: Agreement (inter- and intra-observer) was good for the parameters; interspinous space width, isolated opacities dorsally, beak-shaped formations craniodorsally and modelling cranioventrally. Agreement increased slightly for sum of specific parameters such as radiopacities, modelling, dorsal abnormalities and related abnormalities involving increased opacity, modelling and osseous cyst like lesions. Agreement for the total radiographic abnormalities per back was moderate.

**Conclusions**: Grading of thoracolumbar radiographs in horses without back pain showed good inter- and intra-observer agreement for specific parameters and these should be used in future grading of spinous processes. Limitations should be considered in pre-purchase examinations.

**INTRODUCTION**

Recently there is much discussion about the clinical benefit of radiographic examination of the thoracolumbar spinous processes (SP) in healthy horses during pre-purchase examination (PPE). The FEEVA stated in 2018 that radiographs of equine thoracolumbar SP as part of the PPE are not recommended, they consider that there is no evidence of a correlation between the radiographic appearance of SP and future risk of back pain in asymptomatic horses (1) These findings are supported by multiple studies, which reveal a lack of clinical significance of radiographic findings of SP.(2–6) However, this is in contrast to other studies which showed a good correlation between clinical symptoms and radiographic abnormalities in horses with signs of back pain.(7–10) In addition, Zimmerman et al., concluded that in horses with combined back and sacroiliac pain with or without hind limb lameness or sacroiliac pain with hind limb lameness significantly more severe radiographic abnormalities were found compared to horses with primary back pain, front limb lameness and clinical normal horses. Horses with primary back pain only did not show significant differences in radiographic changes compared to the other groups. In the group of clinically normal horses, 40% had medium graded abnormalities (grade III-IV). (7) A general scientific consensus of the clinical relevance of radiographic abnormalities in SP has not been established.

Furthermore, there are several externally dependent factors may which influence image interpretation of the equine thoracolumbar spine. Major differences in horse conformation, limitations in radiographic technique such as current exposure limits, the infeasibility for orthogonal views and magnification due to increased film-object distance result in a variable quality of radiographs. One study showed that different head and neck positioning resulted in inconsistent measurements of the interspinous space width (ISW) with lower positions leading to increasing intervertebral distances. (11) In addition, measuring the ISW on radiographs in comparison to computed tomographic sagittal reconstructions resulted in an effect likely associated with geometric distortion because of x-ray beam angulation and led to misinterpretation of interspinous space (12). All of these factors can contribute to the variability of radiographic grading.

Only few studies have investigated radiographic alterations of the thoracolumbar spine in horses without clinical signs of back pain during PPE, all using different grading scales.(13,14) Previous used scoring systems all vary substantially and they have not fully described all possible radiographic abnormalities per SP; results were generalized and rather focussed on the sum of all abnormalities in the thoracolumbar spine.(14–18) A limited amount of studies described a grading scale originating from a sum of grades per SP.(4,19) Although, these studies have described prevalence and common anatomical locations of radiographic alterations, none of them investigated agreement among different observers. Pre-purchase examinations are performed by equine clinicians all over the world with different degrees of experience. Therefore, a need for a single reliable and repeatable scoring system is of high priority. A consensus of different parties involved in a pre-purchase examination as well as reproducible availability of comparison of follow-up of radiographic alterations is important. Due to the previously described limitations using such different scoring systems on PPE, interpretation might be variable, non-repeatable and therefore at the risk of being unreliable. This is shown to be similar to what was found for agreement of radiographic abnormalities of other areas in PPE.(20)

Therefore, the aim of this study was to investigate the inter- and intra-observer agreement of grading radiographs of thoracolumbar SP from horses without clinical symptoms of back pain on PPE. The hypothesis was that there is variable agreement between grading different single radiographic variables and agreement might improve for combinations of radiographic alterations.

**MATERIAL AND METHODS**

**Study population**

This study was approved by the Social Science Research Ethical Review Board of the Royal Veterinary College (URN SR2020-0015). Radiographs of one hundred horses presented between July 2019 and December 2019 for PPE at Veterinair Centrum Honselersdijk were retrospectively and randomly selected. Information of animals included age, sex, breed, purpose of use, level of training and height. A thorough clinical and orthopaedic examination was performed in all selected horses by two experienced clinicians (M.L., J.M.), with one having a clinical experience for more than 30 years and the other more than five years. Examination of the thoracolumbar region at rest involved palpation and observation of muscle tension or wastage, evidence of kyphosis, lordosis, pain response to palpation, reflexes and movement of the thoracolumbar spine with manipulative tests. Horses were excluded if there were orthopaedic-related remarks after the clinical part of the pre-purchase examination such as lameness or positive flexion tests.

Horses were sedated with a combination of detomidine (10 µg/kg BW i.v., Detogesica) and butorphanol (10 µg/kg BW i.v., Torbogesica). Left lateral-right lateral radiographs were acquired with the same radiographic equipment (APR VET SHF-535 x-ray generatorb) and computer software (XDR3 and XDR 5 veterinary DR x-ray systemc) using a digital radiography (DR) flat panel detector plate (Cuattro 1417 HD 384mmx460mmc). Radiographs were obtained from the withers to the lumbar region with a minimum of at least one SP overlapping in between consecutive images, resulting in at least three views per thoracolumbar spine. Horses were standing squarely on all four limbs with a neutral head position. The detector plate was positioned touching the horse and a film-focus distance of 100 cm was maintained. Exposures varied depending on the location, size and weight of horses (range: 10 – 100 kV, 25 – 40 mAs).

**Radiographic grading**

All images were anonymised, randomized and at least one SP per radiograph was numbered using established anatomical landmarks.(21) Three observers, one diplomate (D.B.) and one resident in large animal diagnostic imaging (with more than 7 years of clinical experience) (M.L.) and one recently graduated intern working in an equine referral hospital (K.H.) graded images independently once and one observer twice (M.L.). DICOM images were reviewed using image viewer software (Osirixd) with cranial to the left. In order to keep an individual unaltered judgement no training set was created nor training sessions were performed.

Each single spinous process involved in the radiographic study in every horse was evaluated separately, with individual exceptions of those which were not included in the radiographic study and therefore not graded. A grading scale was designed involving structural changes leading to radiographic abnormalities described in previous used scoring systems.(14–19) For each single SP 16 different radiographic variables were graded, this involved ISW, modelling, radiolucencies and radiopacities along its margins and isolated opacities dorsal to the SP. These parameters were individually graded for their abnormalities which varied from 0 (normal) to I-IV (abnormal), see Table 1 for the details of the used radiographic parameters. Examples of the aforementioned radiographic abnormalities of the SP is shown in Figure 1 and 2.

**TABLE 1:** Radiographic scoring system for alteration of the thoracolumbar spinous processes. Definitions of single parameters used for determining the radiographic grade.

|  |  |
| --- | --- |
| **Interspinous space width (ISW)**  0 = > 4 mm  I = 1-4 mm (narrowing)  II = 0 mm (impinging)  III = < 0 mm (overlapping)  Graded according to the narrowest point and caudally to the spinous process | **Isolated linear radiopacity dorsal margin (IsolDo)**  0 = no  I = focal (<25% of the width)  II = moderate (25-50% of the width)  III = marked (>50% of the width) |
| **Modelling dorsal margin (DoMod)**  0 = normal  I = mild regular modelling  II = moderate regular or mild irregular modelling  III = marked regular or moderate irregular modelling  IV = marked irregular modelling | **Modelling craniodorsal margin/beak-shaped formation (DoNose)**  0 = normal  I = mild elongation  II = moderate elongation  III = marked elongation and/or fragment |
| **Intensity of increased opacity cranial (InOpCr) / caudal margin (InOpCa)**  0 = normal, no increased opacity  I = mild increased opacity  II = moderate increased opacity  III = marked increased opacity  Separately recorded but similarly graded | **Width of increased opacity cranial (WidthOpCr) / caudal margin (WidthOpCa)**  0 = 0 mm  I = less than 2 mm  II = in between 2-4 mm  III = more than 4 mm  Separately recorded but similarly graded |
| **Radiolucency cranially (LucencyCr or OCCLCr) / caudally (LucencyCa or OCCLCa)**  0 = no radiolucencies  I = radiolucency < 5 mm in diameter  II = radiolucency 5-10 mm in diameter  III = radiolucency > 10 mm in diameter  The largest abnormality is recorded. Separately recorded and graded. OCCL are radiolucencies with presence of a circumferential rim of radiopacity. | **Modelling craniodorsal third (CrDMod) / cranioventral two thirds (CrVMod) / caudodorsal third (CaDMod) / caudoventral two thirds (CaVMod)**  0 = normal  I = mild regular modelling  II = moderate regular or mild irregular modelling  III = marked regular (misshapen) or moderate irregular modelling  IV = marked irregular modelling (misshapen)  V = fused spinous processes or congenital deformities/bony bridges  Separately recorded but similarly graded |

**Data analysis**

Data were recorded in a spreadsheet (Microsoft Excel Version 1640) and statistical analyses were performed using a statistical program (SPSS Statistics version .26, IBM). Inter- and intra-observer agreement between two experienced observers (M.L., D.B.) was evaluated using Kappa weighted (Kintra, Kinter) calculations. Results of the predetermined multiple grade scoring system (Table 1) were compared. In addition, two different grading systems, extrapolated from the aforementioned grading system were also used to search for agreement. The first included either presence or absence of a radiographic abnormality per parameter per SP. This meant that any abnormality found per parameter resulted in the spinous process graded as abnormal (grade I) and only those without radiographic abnormalities were found normal (grade 0). The latter compared radiographic abnormalities per parameter per SP from being graded as either normal to mild abnormal (grade 0-I) as grade 0 to the moderate to severe abnormalities (grade II-V) together as grade I.

A Spearman's correlation (Sintra, Sinter) was performed for the total of radiographic abnormalities of all parameters of all SP per horse (Total Back radiographic abnormalities). In addition, combinations of different variables were summed to seek for a corresponding correlation coefficient. The results graded by the diplomate and the experienced resident were used for this and an average was calculated out of those sums and used for comparison. Correlation coefficients for combinations were done for the following: Modelling Cranial SP, Modelling Caudal SP (including the cranial or caudal summed dorsal and ventral margins), Modelling total SP (modelling cranial and caudal margins) and Modelling total SP including Dorsal (modelling cranial, caudal and dorsal margins); the sum of radiopacities (intensity and width) cranially, caudally and both combined (Radiopacity Cranial SP, Radiopacity Caudal SP and Radiopacity Total SP); the sum of the different radiographic abnormalities graded dorsally, cranially or caudally (Total Dorsal radiographic abnormalities, Total Cranial radiographic abnormalities and Total Caudal radiographic abnormalities). Finally, the radiographic abnormalities per SP (Total SP radiographic abnormalities) as used in other grading systems were analysed.(7,19) The latter involved the combination of cranial and caudal radiographic alternations involving width and intensity of radiopacities, modelling craniodorsally and caudodorsally and the presence of osseous cyst like lesions per SP. The Spearman correlation coefficients and the Kappa weighted calculations were considered excellent when greater than 0.8, good when between 0.6 and 0.8, moderate when between 0.4 and 0.6, fair when between 0.2-0.4 and poor when less than 0.2. Values of *P*< 0.05 were considered statistically significant. (22–24)

Finally, inter-observer agreement for all observers, involving the less experienced equine clinician, were evaluated using Fleiss’ Kappa (Fk). Results of the predetermined multiple grade scoring system (Table 1) as well as the two-grade grading system for presence or absence of radiographic alterations were compared as described similarly by the Kappa weighted calculations.

**RESULTS**

**Signalment**

The study population comprised of 98 warmblood horses involving Dutch (85), German (9), Danish (2) and Belgian warmbloods (2) and 2 welsh ponies. Their age ranged from two to 14 years (mean: 5.17 years) and the study involved 25 mares, 23 stallions and 52 geldings. All horses were sport horses used for dressage, show jumping or all-round purposes. The height of all horses ranged from 145 to 187 centimetres measured at the highest point of the withers (mean: 169 centimetres). Overall, 1596 SPs were evaluated and individually graded separately. A mean of 15 SP were graded per back which ranged from the most cranial SP being T5 and the most caudal SP of L6. In all horses at least the SP of T7 to L1 were included in the radiographic study.

**Graded radiographic abnormalities**

Most common structural changes were graded in the range of grade 0, I and II. Changes involving the highest grades (grade III, IV or V) involved the minority of the findings in ISW, radiolucencies, osseous cyst like lesions and modelling along the cranial and caudal margins of the SP. More commonly higher graded abnormalities were found in structural changes of the dorsal margin as well as radiopacities along the cranial and caudal margin. Grade V modelling was not present in this group for all investigated margins. Osseous cyst like lesions of all grades were present in 0.63% and 2.38% of all SP for the first observer (ML) and second observer (DB), retrospectively. An overview of the percentages of radiographic abnormalities above grade III are shown in Table 2.

TABLE 2: Percentage of maximum radiographic grades of all horses calculated per observer. These involved radiographic abnormalities with grade III or higher for interspinous space width (ISW), radiopacities (InOpCr+InOpCa+WidthOpCr+WidthOpCa), radiolucencies (LucencyCr+LucencyCa), osseous cyst like lesions (OCCL), (OCCLCr+OCCLCa), dorsal abnormalities (IsolDo, DoMod, DoNose) and radiographic abnormalities with grade IV or higher for modelling (CrDMod+CaDMod+CrVMod+CaVMod). For definition of abbreviations see Table 1.

|  |  |  |
| --- | --- | --- |
|  | **Observer 1** | **Observer 2** |
| **ISW** | 1.06% | 1.75% |
| **Radiopacities** | 3.76% | 10.53% |
| **Radiolucencies** | 1.19% | 0.94% |
| **OCCL** | 0.00% | 0.56% |
| **Dorsal abnormalities** | 7.83% | 8.96% |
| **Modelling** | 0.06% | 0.69% |

**Agreement for single parameters**

An overview of all kappa weighted calculations is shown in Table 3. Agreement was excellent for ISW (Kintra=0.84, Kinter=0.82) and good for isolated opacities of the dorsal margin (Kintra=0.78, Kinter=0.76). For presence and size of beak-shaped formation of the craniodorsal margin of the SP moderate to good agreement was found (Kintra=0.72, Kinter=0.54). For modelling of the cranioventral two thirds of the margin moderate agreement was found (Kintra=0.57, Kinter=0.56). The agreement of the remaining parameters was fair to poor. No sustainable improvement was observed using both different two-grade grading systems.

**TABLE 3:** Inter- and intra-observer agreement of single parameters. Kappa weighted values for two observers and Fleiss Kappa calculation for multiple observers for all used parameters. Kappa weighted 0-1 defines agreement when graded for presence or absence of abnormalities. Kappa weighted new grade defines agreement when graded for absent (grade 0) to mild (grade I) radiographic abnormalities vs moderate (grade II-III) to marked (III-V) abnormalities. Fleiss Kappa 0-1 defines agreement when graded for presence or absence of abnormalities. (P<0.05), n.s. = not significant†.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Kappa weighted** | | **Kappa weighted 0-1** | | | | **Kappa weighted normal/mild vs. moderate/severe** | | **Fleiss Kappa** | | | **Fleiss Kappa 0-1** |
|  | Intra | Inter | | Intra | Inter | Intra | | Inter | |  |  | |
| **Interspinous space width** | 0.8399 | 0.8205 | | 0.821 | 0.807 | 0.758 | | 0.725 | | 0.269 | 0.624 | |
| **Isolated radiopacity dorsally** | 0.7845 | 0.763 | | 0.772 | 0.717 | 0.747 | | 0.705 | | 0.237 | 0.626 | |
| **Modelling dorsal margin** | 0.5762 | 0.1484 | | 0.425 | 0.194 | 0.62 | | 0.108 | | 0.107 | 0.232 | |
| **Beak-shaped formations craniodorsally** | 0.7153 | 0.5414 | | 0.687 | 0.458 | 0.545 | | 0.42 | | 0.157 | 0.332 | |
| **Modelling cranioventral two thirds** | 0.5735 | 0.5629 | | 0.395 | 0.43 | 0.51 | | 0.416 | | 0.139 | 0.385 | |
| **Modelling caudoventral two thirds** | 0.4504 | 0.1859 | | 0.384 | 0.196 | 0.363 | | 0.109 | | 0.067 | 0.216 | |
| **Modelling craniodorsal third** | 0.3827 | 0.3792 | | 0.297 | 0.258 | 0.352 | | 0.257 | | 0.062 | 0.186 | |
| **Modelling caudodorsal third** | 0.4317 | 0.2373 | | 0.351 | 0.205 | 0.315 | | 0.175 | | 0.036 | 0.165 | |
| **Intensity radiopacity cranial margin** | 0.5008 | 0.3307 | | 0.48 | 0.296 | 0.368 | | 0.206 | | 0.095 | 0.297 | |
| **Width radiopacity cranial margin** | 0.4857 | 0.2663 | | 0.477 | 0.296 | 0.418 | | 0.21 | | 0.092 | 0.296 | |
| **Intensity radiopacity caudal margin** | 0.4858 | 0.3938 | | 0.386 | 0.321 | 0.403 | | 0.266 | | 0.119 | 0.286 | |
| **Width radiopacity caudal margin** | 0.4113 | 0.3144 | | 0.384 | 0.319 | 0.308 | | 0.205 | | 0.119 | 0.286 | |
| **Osseous-cyst-like-lesion cranial margin** | 0.3229 | 0.1949 | | 0.312 | 0.127 | 0.198 | | 0.186 | | n.s. † | 0.075 | |
| **Osseous-cyst-like-lesion caudal margin** | 0.4362 | 0.1672 | | 0.499 | 0.108 | 0.399 | | 0.181 | | n.s. † | n.s. † | |
| **Radiolucency Cranial margin** | 0.2771 | 0.2087 | | 0.245 | 0.215 | 0.225 | | 0.127 | | 0.079 | 0.157 | |
| **Radiolucency Caudal margin** | 0.273 | 0.107 | | 0.262 | 0.029 | 0.262 | | 0.117 | | n.s. † | 0.047 | |

**Agreement of combined parameters**

An overview of all Spearman correlation coefficients is shown in Table 4. Correlation of the sum of modelling of craniodorsal and cranioventral part of the SP (Modelling Cranial SP) was moderate to good (Sintra=0.60, Sinter=0.45). This was slightly less for Modelling Caudal SP (Sintra=0.56, Sinter=0.27) and Modelling Total SP (Sintra=0.63 Sinter=0.37), which were both moderate to fair. Correlations for agreement of the sum of intensity and width of radiopacities cranially and caudally (Radiopacity total SP) was fair to moderate (Sintra=0.54 and Sinter=0.35) and almost similar to agreements when separately graded (Radiopacity Cranial SP: Sintra=0.52 and Sinter=0.31, Radiopacity Caudal SP: Sintra=0.51 and Sinter=0.38).

For combination of parameters per SP, all dorsal abnormalities (Total Dorsal radiographic abnormalities) revealed moderate to good agreement (Sintra=0.68 and Sinter=0.46). Cranial and caudal abnormalities revealed good intra-observer correlations but only fair correlations on inter-observer agreement (Total Cranial radiographic abnormalities: Sintra 0.61 and Sinter=0.37, Total Caudal radiographic abnormalities: Sintra=0.57 and Sinter=0.28). Combining different clinically related parameters in a single SP (Total SP radiographic abnormalities), as previously explained, showed moderate to fair results (Sintra=0.58 and Sinter=0.36). Finally, the Total Back radiographic abnormalities per horse (involving the sums of all SP), was found to be moderate (Sintra=0.49 and Sinter=0.39).

**TABLE 4:** Inter- and intra-observer agreement of combined parameters. Spearman correlation coefficient for agreement of the average of the sum of certain specific radiographic alterations. (P<0.05). SP= Spinous process†. Total Dorsal = DoMod, IsolDo and DoNose, Total Cranial/Caudal = CrDMod/CaDMod, CrVMod/CaVMod, InOpCr/InOpCa and LucencyCr/LucencyCa, Total SP = CrDMod, CaDMod, InOpCr, InOpCa, WidthOpCr, WidthOpCa, OCCLCr, OCCLCa (for definition of abbreviations see table 1).

|  |  |  |
| --- | --- | --- |
|  | **Intra-Spearman** | **Inter-Spearman** |
| **Modelling Cranial SP**† **(dorsal & ventral)** | 0.603 | 0.45 |
| **Modelling Caudal SP**† **(dorsal & ventral)** | 0.562 | 0.274 |
| **Modelling Total SP**† **(cranial and caudal)** | 0.631 | 0.367 |
| **Modelling Total SP**† **incl. Dorsal** | 0.606 | 0.285 |
|  |  |  |
| **Radiopacity Cranial SP**† **(intensity and width)** | 0.518 | 0.309 |
| **Radiopacity Caudal SP**† **(intensity and width)** | 0.514 | 0.38 |
| **Radiopacity Total SP**† **(cranial & caudal)** | 0.544 | 0.35 |
|  |  |  |
| **Total Dorsal radiographic abnormalities** | 0.681 | 0.463 |
| **Total Cranial radiographic abnormalities** | 0.613 | 0.365 |
| **Total Caudal radiographic abnormalities** | 0.572 | 0.275 |
|  |  |  |
| **Total SP radiographic abnormalities** | 0.583 | 0.358 |
| **Total Back radiographic abnormalities** | 0.489 | 0.389 |

**Agreement between multiple observers**

The Fleiss Kappa agreement for multiple observers was overall fair to poor (Table 3). Only ISW and isolated opacities dorsally were found to have good agreement for the presence or absence of abnormalities, Fk=0.62 and 0.63 respectively.

**DISCUSSION**

The current study is the first to describe inter- and intra-observer agreement of different radiographic alterations found in equine thoracolumbar SP on PPE. Previous published scoring systems have not been validated and tested for reliability and these studies either graded their radiographic alterations by only one experienced observer or two observers in consensuses.(4,14–19)

Research involving inter- and intra-observer repeatability of interpretation of radiographs of other areas in horses state that agreement varies substantially. Agreement varied depending on severity of findings, anatomical area, training of observers and different evaluated parameters used or the approach of grading.(20,25–27) For grading of equine distal intertarsal joint only poor agreement was observed, agreement was improved when descriptive terms were used rather than a grading scale.(28) In the same study the author also advised that in whatever type of grading scale is used, it should not provide more than five choices. Next to their results they also state that human literature revealed that radiographic changes of osteoarthritis consistent with mild and severe changes were more reliable assessed than moderate changes .(28) This corresponds to the findings in the current study, where modelling along the margins consisted of five choices resulted in poorer agreement in comparison other parameters with less choice. Other studies showed inter- and intra-observer agreement is depending on the evaluated radiographic parameters, i.e. of the navicular bone. For instance, agreement on presence of distal border fragments or sclerosis of the spongiosa was poor whereas enthesophyte or osteophyte formation along the navicular bone margin was almost perfect.(29) This variance is similar to what was found in the current study and is associated with the subjective interpretation rather than a measured interpretation of some of those parameters.

Ultimately, a reliable scoring system should comprise a grading scale which could be used by equine veterinarians with different levels of experience. It should be appropriate for PPE, for use in scientific studies and for clinical follow-up. The multigrade single parameter scoring system in the current study resulted in moderate agreement amongst more experienced observers but in a poor agreement for less experienced veterinarians. This finding is unsurprising in the opinion of the authors and is supported by a study performed on additional radiographs of the navicular bone and grading the cervical spine.(27,30) The wide variety of radiographic abnormalities in association with difficulty of obtaining standardized radiographs (due to major differences in radiographic techniques and horses’ conformation leading to compositional artefacts), image distortion and no possibility for orthogonal projections may results in different interpretations of alterations as similarly found in other studies.(27,30). Therefore, a new simple and more repeatable scoring system with training before usage in the field needs to be developed for future use. In order to be representative of veterinarians performing PPE in many different situations, no training set was created nor training sessions were performed. Whether agreement would have been improved after training was beyond the aim of the current study.

Grading systems which do not grade the SP separately but use a combined and generalized grading system for the thoracolumbar spine, as done in previous grading systems (14–18), was found to be of moderate to fair agreement in the current study for the most experienced observers. However, usage of multiple individual parameters per anatomical region of the SP also revealed some unreliable results. Therefore, a different approach according to these results appears necessary. Inter- and intra-observer agreement for single parameters such as ISW appears to be a reliable parameter for the usage of a future grading system. Similarly, the presence of size of isolated opacites along the dorsal margin and beak-shaped formation of the craniodorsal margin of the SP was found to be reliable. Whether all of the single parameters were graded as present or absent or in according to the multigrade or two-grade grading system did not reveal major differences in agreement.

The combination of intensity and width of radiopacities per SP and combinations of modelling along margins of the SP showed better agreement in comparison to a single approach of each parameter per anatomical region. One exception was the modelling of the cranioventral margin of the SP which revealed good agreement. The cause of this finding is not clear in relationship to the poor agreement of the other margins. Grading osseous cyst like lesions in comparison to grading radiolucencies was more reliable. This finding is likely due to the unspecified definition of radiolucencies in comparison to osseous cyst like lesions. Osseous cyst like lesions were specifically those radiolucencies surrounded by a rim of sclerosis whereas radiolucencies lacked this and were defined by decreased radiopacity within the SP. Lack of histopathology deprives us from differentiating both, however a radiolucency without a rim of opacity may still have represented a radiographic abnormality and were therefore separately graded in this study. A lack of a clear sclerotic rim may have caused confusion interpretating them in the current radiographic examination. However, one should consider that both results were still disappointing.

Combining different radiographic abnormalities along the cortical margins dorsally, caudally and cranially, as done in other studies revealed similar to better agreement to the values found for single parameters in the current study.(4) Combining different possibly associated abnormalities within the complete SP, as found in other studies using scoring systems for grading the spinous processes, led to moderate agreement.(7,19,31) Therefore, the combination of certain parameters, in addition to the usage of some individual reliable variables in a future grading system might be advisable.

Whether the usage of a combined or summed grading system is scientifically as accurate as a single parameter grading system as originally chosen in this study was beyond the scope of this study. Future studies which incorporate a comparison of clinically affected and non-affected horses should reveal if a more detailed grading system is necessary to radiographically differentiate between the both groups. On the contrary, one should also bear in mind that the true incidence of pain arising from abnormal SP is not known, anatomical variations are however extremely common.(31) One study found a large widespread of radiographic and scintigraphic changes in riding horses without clinical signs of back problems which led to the conclusion that their clinical significance is questionable.(4) Unfortunately, no studies have been performed on the degree and the variance of single structural changes of radiographic abnormalities of the SP in accordance to the change of developing clinical findings of back pain in the future. This, in association with our findings of moderate agreement of scoring SP of unaffected horses on PPE and the statement of the FEEVA indicates that inclusion of radiographs of the thoracolumbar SP in a PPE is questionable. Follow-up studies are required to establish cut-off values for severity in radiographic abnormalities of the SP.

Limitations of the study relate to the chosen study cohort. By including only horses without clinical signs of back pain only less severe radiographic abnormalities and a lower prevalence of those abnormalities were observed. Secondly, the mean age of this group was five years and it can be expected that radiographic changes might be of less extent. This is similar to what was found in another study where results showed that the only group where no radiographic abnormalities were found in SP in the scope of PPE was the youngster group (age between 3-7 years). (5) Therefore, poor agreement of certain radiographic features might be associated with the prevalence of these abnormalities. This most likely led to the poor agreement in grading radiolucencies as well as osseous cyst like lesions in the current study, which could be explained by a similar effect due to low prevalence which was found in another study.(29) Follow-up of this cohort was not obtained since the aim of the study only investigated inter- and intra-observer agreement of different radiographic alterations amongst observers and therefore stating healthy horses is only true for that aforementioned time frame in which the PPE was performed.

Inter- and intra-observer agreement for grading thoracolumbar SP of horses without back pain at PPE showed variable levels of agreement. Several single parameters such as ISW and the presence, size of isolated opacities dorsal to the SP, craniodorsal beak-shaped formations and cranioventral modelling process appeared reliable. However, structural changes such as modelling along the other cortical margins of the SP, radiolucencies, radiopacities or OCCLs are less precise. Combining certain features of the SP such as the sum of different sides of modelling, different features of abnormal radiopacity, or certain different associated radiographic abnormalities either per anatomical region of the SP or for the complete SP show promising results. Parameters advised for a future grading system would involve: Interspinous space width, Intensity and width of increased opacity cranial & caudal margin of the spinous process, Modelling cranial & caudal margins of the spinous process, Osseous cyst like lesions cranially and caudally of the spinous process, Isolated linear radiopacity dorsal margin spinous processes, Modelling craniodorsal margin of the spinous process, Number of affected spinous processes in total and Location(s) of affected spinous processes. However, in light of the possible clinical insignificance of radiographic alterations on horses without clinical symptoms of back pain, the relevance of these radiographs might be questionable. Future research will need to incorporate such combinations to establish a single reliable and repeatable scoring system tested on horses with and without back pain.

Manufacturers’ addresses

aZoetis Belgian, Louvain-la-Neuve, Belgium

bSedecal, Madrid, Spain

cVetZ, Sliedrecht, The Netherlands

dPixmeo SARL, Bernex, Switzerland

Presented at: Parts of this study were presented as oral presentation at the ECVDI Online Conference 2020, September 17-18.

Author contributorship

All authors contributed to study design. Data collection and analysis was performed by all authors. Statistical analysis was performed by M.G.P. Looijen and D. Berner. M.G.P. Looijen drafted the manuscript and all authors contributed to and approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Competing interests

No competing interests have been declared.

Ethical approval

Ethics approval has been given by the Social Science Research Ethical Review Board at the Royal Veterinary College (URN SR2020-0015).

Patient consent

Client confidentiality has been maintained throughout the study. Explicit owner informed consent for participation in the study was not obtained.

Data sharing

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The authors thank the staff of Evidensia Veterinair Centrum Honselersdijk, Dr. Poen and O.K. Kortlang for their assistance in collecting all the data.

**REFERENCES**

1. FEEVA Position Statement of pre-purchase radiography on dorsal spinous processes. FEEVA. 2018. Available from: https://feeva.fve.org/working\_documents/ppe\_rx/

2. Rieland E. Radiological findings in spinous processes in clinically sound horses. Proc 4th Maastricht International Conference Equine Medicine, Maastricht, the Netherlands, 6-7th December, 2002.

3. Ranner W, Gerhards H. Vorkommen und Bedeutung von Rückenerkrankungen-Insbesondere des “Kissing Spine”-Syndroms-Bei Pferden in Süddeutschland. Pferdeheilkunde. 2002;18(1):21-3

4. Erichsen C, Eksell P, Roethlisberger Holm K, Lord P, Johnston C. Relationship between scintigraphic and radiographic evaluations of spinous processes in the thoracolumbar spine in riding horses without clinical signs of back problems. Equine Vet J. 2004;36(6):458–65.

5. Holmer M, Wollanke B, Stadtbäumer G. X-ray alterations on spinal processes of 295 warmblood horses without clinical findings. Pferdeheilkunde. 2007;23(5):507–11.

6. Geiger CP, Gerhards H. Radiologische Befunderhebung an der Brustwirbelsäule des Pferdes gemäß des Röntgenleitfadens 2007 unter Berücksichtigung der klinischen Relevanz. Pferdeheilkunde. 2015;31(1):39–48.

7. Zimmerman M, Dyson S, Murray R. Close, impinging and overriding spinous processes in the thoracolumbar spine: The relationship between radiological and scintigraphic findings and clinical signs. Equine Vet J. 2012;44(2):178–84.

8. Turner TA. Overriding Spinous Processes (“Kissing Spines”) in Horses: Diagnosis, Treatment, and Outcome in 212 Cases. Proc 57th Annual Convention of the American Association of Equine Practitioners. San Antonio, Texas, USA, 18-22 November, 2011.

9. Cousty M, Retureau C, Tricaud C, Geffroy O, Caure S. Location of radiological lesions of the thoracolumbar column in French trotters with and without signs of back pain. Vet Rec. 2010;166(2):41–5.

10. Hendrickson D. The thoracolumbar spine/Back. In: Baxter GM, editor. Adams’ Lameness in Horses. 7th ed. Lippincott Williams & Wilkins; 2020. p. 784–95.

11. Berner D, Winter K, Brehm W, Gerlach K. Influence of head and neck position on radiographic measurement of intervertebral distances between thoracic dorsal spinous processes in clinically sound horses. Equine Vet J. 2012;44(SUPPL. 43):21–6.

12. Djernæs JD, Nielsen J V., Berg LC. Effects of X-Ray Beam Angle and Geometric Distortion on Width of Equine Thoracolumbar Interspinous Spaces Using Radiography and Computed Tomography—a Cadaveric Study. Vet Radiol Ultrasound. 2017;58(2):169–75.

13. Holmer M. Röntgenbefunde an den Dornfortsätzen klinisch rückengesunder Warmblutpferde. Doctoral dissertation, Ludwig-Maximilians-University, Munich, Germany, 2005.

14. De Graaf K, Enzerink E, Van Oijen P, Smeenk A, Dik KJ. The radiographic frequency of impingement of the dorsal spinous processes at purchase examination and its clinical significance in 220 warmblood sporthorses. Pferdeheilkunde. 2015;31(5):461–8.

15. Pettersson, H., B. Strömberg and IM. Das thorakolumbale, interspinale Syndrom (TLI) des Reitpferdes-Retrospektiver Vergleich konservativ und chirurgisch behandelter Fälle. Pferdeheilkunde. 1987;3:313–9.

16. Sager J. Die Erkrankungen der Processus spinosi der Brust-und Lendenwirbelsäule des Pferdes: eine klinische und röntgenologische Studie. Doctoral dissertation, Free University of Berlin, Berlin, Germany, 1997.

17. Denoix, Jean-Marie M., Dyson SJ. Thoracolumbar spine. In: Diagnosis and Management of Lameness in the Horse 2nd ed. Eds Ross MW, Dyson SJ, WB Saunders; 2011. p. 592–605.

18. Gerhards, H., Hertsch, B., Jahn, P., & Brunken G. Leitfaden für die röntgenologische Beurteilung bei der Kaufuntersuchung des Pferdes (RoLF07). Deutsches Tierärzteblatt. 2007;57:292–303.

19. Zimmerman M, Dyson S, Murray R. Comparison of radiographic and scintigraphic findings of the spinous processes in the equine thoracolumbar region. Vet Radiol Ultrasound. 2011;52(6):661–71.

20. Hellige M, Rohn K, Buschkamp L, Stadler P. Interobserver agreement in interpreting radiographs of horses using the German Radiology-Guidelines 2007 (“Röntgenleitfaden 2007”). Pferdeheilkunde. 2018;34(3):212–22.

21. Butler JA, Colles CM, Dyson SJ, Kold SE, Poulos PW. The vertebral column. In: Clinical radiology of the horse 4th ed. Eds Butler JA, Colles CM, Dyson SJ, Kold SE, Poulos PW. Blackwell Scientific Publications Ltd; 2017. p. 569–601.

22. McHugh ML. Lessons in biostatistics interrater reliability : the kappa statistic. Biochem Medica. 2012;22(3):276–82.

23. Schober P, Schwarte LA. Correlation coefficients: Appropriate use and interpretation. Anesth Analg. 2018;126(5):1763–8.

24. Akoglu H. User’s guide to correlation coefficients. Turkish J Emerg Med. 2018;18(3):91–3.

25. Lepeule J, Robert C, Bareille N, Valette JP, Jacquet S, Seegers H, et al. A reliable severity scoring system for radiographic findings in the limbs of young horses. Vet J. 2013;197(1):52–7.

26. Hauspie S. The proximal aspect of the dorsal condylar sagittal ridge and the adjacent soft tissues in the fetlock joint of the Warmblood horse. Doctoral dissertation, Ghent University, Belgium, 2012.

27. Johnson SA, Barrett MF, Frisbie DD. Additional palmaroproximal–palmarodistal oblique radiographic projections improve accuracy of detection and characterization of equine flexor cortical lysis. Vet Radiol Ultrasound. 2018;59(4):387–95.

28. Labens R, Innocent GT, Voûte LC. Reliability of a quantitative rating scale for assessment of horses with distal tarsal osteoarthritis. Vet Radiol Ultrasound. 2007;48(3):204–11.

29. Groth AM, May SA, Weaver MP, Weller R. Intra- and interobserver agreement in the interpretation of navicular bones on radiographs and computed tomography scans. Equine Vet J. 2009;41(2):124–9.

30. Espinosa‐Mur P, Phillips KL, Galuppo LD, DeRouen A, Benoit P, Anderson E, et al. Radiological prevalence of osteoarthritis of the cervical region in 104 performing warmblood jumpers. Equine Vet J. 2020;(March):1–7.

31. Henson, F. M.D.Kidd JA. Overriding Dorsal Spinous Processes. In: Equine back pathology: Diagnosis and treatment. Eds Henson FMD. Wiley-Blackwell; 2009. p. 147–56.

**Figure legends**

Figure 1: Laterolateral radiographs of the thoracolumbar spine. Cranial is to the left. a: interspinous space width grade II (impinging) (square). b: small osseous cyst like lesion (OCCL) grade I (circle). Interspinous space width grade I (oblique horizontal rectangular boxes). Dorsal modelling grade I (horizontal square). c: Grade III isolated opacities dorsal to the spinous processes (rectangular boxes). Medium sized area of radiolucency, grade II (circle).

Figure 2: Laterolateral radiographs of the thoracolumbar spine. Cranial is to the left. a: Grade II beak-shaped formation of the craniodorsal aspect of the spinous process (circle). Modelling of the craniodorsal third of the spinous process, grade II (left oblique rectangular box), grade I intensity and grade II width of increased opacity along the margin of the spinous process (right oblique rectangular box). b: Large OCCL in the dorsal third of the spinous process, grade III, this lesion was graded cranially (circle). Grade III width and intensity of increased opacity and grade II modelling of the craniodorsal third of the spinous process (vertical oblique box).