Abstract. Background/Aim: The aim of this study was to analyze the inter- and intra-observer variability regarding biopsy technique in bone and soft tissue sarcoma based on magnetic resonance imaging (MRI). Patients and Methods: Thirty-seven MRI scans of bone and soft tissue sarcomas treated in our clinic were randomly selected. Six observers with three different expertise levels were assigned to analyze the scans for suspected entity and preferred biopsy technique at 2 time points with a delay of 8 weeks. Results: The differentiation between bone and soft tissue sarcomas in MRI seemed closely related to the observer’s level of experience. Regarding biopsy technique, no inter-observer accordance could be identified in either group. Conclusion: We observed an association of inter- and intra-observer agreement regarding suspected tumor entity and the observer’s level of experience. The decision for either biopsy technique showed a low inter-observer but high intra-observer variability. These findings suggest that the decision for incisional or core needle biopsy is, even in the expert group, frequently based on personal predilection.

Diagnostic assessment includes clinical history and examination, proper imaging and tissue biopsy depicting the crucial step in the diagnostic cascade to set-up a histological diagnosis for the therapeutic strategy to be followed (3, 4).

Biopsy can be performed as open or percutaneous procedure. Both techniques are associated with specific advantages and disadvantages. However, the objective is to gain a representative tissue sample with minimized trauma considering the correct surgical approach for a later resection to facilitate limb-sparing procedures.

In this context, open biopsy has been considered as the gold-standard for diagnosis of malignant and uncertain musculoskeletal tumors for a long time. However, recent studies increasingly suggest similar diagnostic accuracy for core needle biopsy, especially in bone tumors (5-10).

The indication for either biopsy technique should be mainly based on the suspected entity (10-13). Interpretation of clinical and radiological findings in order to provide a tentative diagnosis can, however, be very demanding necessitating a high level of clinical experience.

The aim of the present study, therefore, was to analyze the inter- and intra-observer variability of junior, advanced and senior orthopedic oncologists regarding suspected tumor entity and favored biopsy technique at two time points in different bone and soft tissue sarcomas based on magnetic resonance imaging (MRI).

Materials and Methods

Thirty-seven MRI Scans of patients suffering from bone and soft tissue sarcoma treated in our clinic were randomly selected from a total of 338 scans presented during 2011 and 2013 at our clinic and anonymized.

All included patients underwent either a percutaneous or an open biopsy and received tumor resection in our musculoskeletal tumor center during the last 5 years. The selected patients included 20 males (m) and 17 females (fem). The age varied between 14 and 71 years. In total, 16 osteosarcomas (11 m, 5 fem), 5 Ewings sarcomas (3 m, 2 fem), 5 liposarcomas (2 m, 3 fem), 6 chondrosarcomas (3 m, 3 fem) and 4 pleomorphic sarcomas (1 m, 3 fem) were comprised.
Six observers with three experience levels (“beginner”: fellows, “intermediate”: shortly before their specialization as a consultant, “expert”: senior orthopedic tumor surgeon) had to analyze all 37 MRI scans for suspected entity and chose the preferred biopsy technique (open biopsy or core needle biopsy) in two separate review sessions 8 weeks apart. Except the MRI scans, the reviewers had no further patient information.

Data was recorded and the inter-observer, as well as the intra-observer, agreement for biopsy technique and suspected entity were calculated.

In order to assess intra-rater agreement of biopsy technique and suspected entity between two time points or inter-rater agreement, between two rater in each expertise group, The Cohen’s κ coefficient was computed and interpreted as stated by Landis and Koch (14) (Table I). Multi-rater agreement of all raters at each time point was assessed using the Fleiss’ κ variant. Bootstrap confidence intervals (95% CI) based on 500 replicates were computed for the κ coefficients. Differences in proportions of chosen methods or entities between two raters are presented with CI (95%). Percentage of correct specification was assessed per rater.

Results

Tumor entity. In the expert group, the correct diagnosis could be identified in 90.5% in session 1 (91.9% and 89.2%) and 88.2% in session 2 (82.2% and 94.6%). The intermediate group attained 94.6% (91.9% and 97.3%) and 93.25% (89.2% and 97.5%), respectively. The beginner group achieved 86.5% (89.2% and 83.8%) and 83.8% (89.2% and 78.4%).

Intra-observer agreement. The Cohen’s κ coefficient calculated for the intra-observer agreement regarding biopsy technique between the two sessions varied form κ=0.17 to κ=1.00. The intra-observer agreement in the expert group was κ=0.94 (95% CI=[0.64; 1.00]) and κ=0.65 (95% CI=[0.21; 0.91]), the intermediate group achieved κ=1.00 and κ=0.61 (95% CI=[0.33; 0.81]), the κ values in the beginner group were 0.41 (95% CI=[−0.05; 0.77]) and 0.17 (95% CI=[0.18; 0.48]) (Figure 1).

Regarding the correct diagnosis, the agreement in the expert group at the two time points was κ=0.8 (95% CI=[0.42; 1.00]) and κ=0.76 (95% CI=[0.45; 0.94]), in the intermediate group κ=0.88 (95% CI=[0.59; 1.00]) and κ=1.00, in the beginner group κ=0.81 (95% CI=[0.43; 0.94]) and κ=0.38 (95% CI=[0.04; 0.66]).

Multi-observer agreement (biopsy technique). The multi-observer Fleiss’ κappa (κ) approach comparing all observers at the two time points relating to the biopsy technique was 0.0797 (95% CI=[−0.026; 0.238]) at the first time point and 0.103 (95% CI=[−0.025; 0.297]) in session 2.

Inter-observer agreement. Cohen’s κappa coefficient varied for the inter-observer agreement in each group for biopsy technique from –0.08 to 0.31. In the expert group, the calculated kappa values were –0.06 (95% CI=[−0.30; 0.33]) and 0.1 (95% CI=[−0.23; 0.41]), in the intermediate group kappa was –0.01 (95% CI=[−0.30; 0.25]) and –0.08 (95% CI=[−0.43; 0.22]). The beginner group achieved kappa values of 0.31 (95% CI=[0.05; 0.70]) and 0.07 (95% CI=[−0.23; 0.42]) (Figure 2). Regarding the inter-observer agreement of suspected diagnosis, high level kappa values were found in all groups of experience. For the expert group kappa values were 0.8 (95% CI=[0.39; 0.94]) (time point 1) and 0.76 (95% CI=[0.46; 0.94]) (time point 2). The intermediate group achieved 0.88 (95% CI=[0.61; 1.00]) at both time points and the beginner group 0.36 (95% CI=[0.04; 0.67]) and 0.94 (95% CI=[0.72; 1.00]), respectively.

The prevalence of biopsy and the differences in proportions of biopsy between two raters are summarized in Table II.

Discussion

The decision for either core needle or open biopsy of bone and soft tissue sarcomas can be difficult and requires experience. In literature, the choice of biopsy technique is discussed controversially. Strauss et al. define core needle biopsy as standard biopsy technique for suspected soft tissue
tumors (15). On the contrary, Pohlig et al. demonstrated almost equal results for open biopsy compared with core needle biopsy for bone sarcomas but inferior results for core needle biopsy in soft tissue sarcoma (10). Similar results were found by Kasraeian et al. when comparing open biopsy, core needle biopsy and fine needle aspiration in soft tissue tumors (16-19). These findings are probably due to the higher cellular heterogeneity of soft tissue sarcomas, particularly liposarcoma, synovial or pleomorphic sarcoma (3-5). Providing a proper tentative diagnosis based on the clinical and, especially, the radiological findings, is, therefore, indispensable for the surgeon in order to decide on either biopsy technique.

Our study suggests an association between experience level and correct diagnosis. The observer in the expert group correctly identified the tumor in 90.5%, whereas the beginner group only achieved 85.1%. Although the 5% difference may seem insignificant, it can lead to a wrong decision regarding the biopsy technique and accordingly yield incorrect histological results. In these cases, the tumor might be treated inadequately or a re-biopsy has to be performed accounting for unfavorable outcomes (15, 20-24).

Figure 2. Inter-observer variability regarding biopsy technique (95% CI).
A further aspect is the reproducibility of choosing either biopsy technique within each observer. In our study we could also see an association of the intra-observer variability with the level of experience. In this regard, beginners showed a low intra-observer agreement. On the other hand, experts exhibited a high agreement of their decisions regarding biopsy technique between the 2 sessions. This indicates that experts and intermediates probably follow a certain and reproducible pattern of decision finding compared to the beginner group. Kappa values for experts were 0.94 and 0.65 and 1.00 and 0.61 for intermediates. These values are considered to have strong accordance up to complete accordance for the two time points.

Despite the highly consistent intra-observer results in the expert and intermediate group, we found a low inter-observer agreement regarding biopsy technique. No accordance between the observers of each group could be shown in the intermediate group (kappa of –0.01 and –0.08) and in the expert group (–0.06 and 0.1). Furthermore, no objective criteria regarding the decision like tumor location, homogenous or heterogeneous structure, bone or soft tissue sarcoma could be identified in either group. Multi-rater agreement of all raters at each time

Figure 1. Intra-observer variability regarding biopsy technique (95% CI).
point was assessed using the Fleiss’ kappa coefficient. Bootstrap confidence intervals (95%) based on 500 replicates were computed for the kappa coefficients. This showed accordance by coincidence in both sessions (95% CI). These findings can be possibly explained by the development of biopsy in bone and soft tissue sarcomas as open biopsy was regarded the gold standard for many years. In recent years, percutaneous techniques like core needle biopsy (CNB) or fine-needle aspiration (FNA) were applied with increasing accuracy (25, 26). However, recent studies suggest slightly inferior results of CNB and FNA in soft tissue masses. In this regard, older, more experienced surgeons might tend to open biopsies as this was their standard for many years. In the other hand younger surgeons might prefer CNB as it yields at least similar results (10, 16, 27-29).

However, we have to note some limitations of our study. Although our Center is specialized in sarcoma therapy as a sarcoma center, bone and soft tissue sarcomas are rare diseases and even in a center like ours just part of the professional stuff is qualified to participate in studies of tumor diagnosis and treatment. Therefore, the good results concerning the differentiation between soft tissue and bone sarcoma can be easily explained. Even the beginner group has, compared to an average orthopedic surgeon, a lot of experience in evaluating MRI scans of suspected sarcomas. A study with higher number of participants could only be accomplished in a multicenter study. Secondly, we investigated a quite homogenous population in terms of tumor entities (osteosarcoma, Ewing’s sarcoma, chondrosarcoma, liposarcoma, pleomorphic sarcoma) and more rarely tumor entities were not included.

Conclusion

In our study we observed an association of inter- and intra-observer agreement regarding the suspected tumor entity and the observer’s level of experience. The decision for either biopsy technique showed a low inter-observer but high intra-observer agreement. These findings suggest that the decision for incisional or core needle biopsy is, even in the expert group, frequently based on personal predilection emphasizing the need for biopsy guidelines. Inter disciplinary tumor boards attending orthopedic surgeons, radiologists, oncologists pathologists could improve diagnosis and therapy by making a decision by consensus regarding biopsy technique. We accentuate that patients with suspected bone or soft tissue sarcoma have to be transferred to dedicated Centers for diagnostic workup and integrated therapy.

Acknowledgements

This work was supported by the Wilhelm-Sander-Foundation for Cancer Research.

References


Received October 16, 2014
Revised November 11, 2014
Accepted November 14, 2014