Introduction

Agreement between veterinarians in determining the edges of a tumor are important to allow accurate and consistent patient staging, assessment of response to therapy, as well as ensure a uniform “surgical dose” between surgeons and studies.

The objective of this prospective, blinded, randomized clinical pilot study was to evaluate agreement and reliability of veterinarians in determining the gross edge of locally invasive subcutaneous malignant tumors in dogs.

Materials and Methods

Tumor measurements were modeled after the recommendations from the RECIST criteria. (Nguyen et al., Vet Comp Oncol 2015)

Four raters were instructed to use calipers to measure the longest diameter of cytologically-confirmed, treatment naïve, mast cell tumors or soft tissue sarcomas in client-owned dogs.

Three randomized measurement trials were performed, both pre- and post-sedation. Each rater was blinded to other raters measurements, but were aware of their own previous measurements. An oncology technician photodocumented each measurement trial.

Inter- and intra-rater reliability was evaluated using intra-class correlation coefficient and agreement was evaluated using Bland-Altman analysis. An a priori limit of agreement was set at 10 mm, as it was reasoned that a difference of this magnitude could carry clinically-relevant consequences for patient staging, measuring response to treatment, and determining surgical margins during tumor excision.

Results

Twelve tumors (seven soft tissue sarcomas and five mast cell tumors) were measured for the study with raw measurement values presented in Figure 1 (directly below).

Inter- and intra-rater reliability were good to excellent (ICC range 0.7263 - 0.9966).

Inter- and intra-rater agreement were considered unacceptable (Bland-Altman limits of agreement range 11.9 mm - 55.6 mm).

There were no significant differences between pre- and post-sedation reliability or agreement.

Review of measurement trial photographs revealed that caliper orientation changes were frequent, occurring in 9/12 (75%) and 8/12 (67%) pre- and post-sedation cases. (Figure 2; lower left)

No significant correlation was found between inter-rater measurement standard deviations and caliper orientation changes or dog BCS.

Discussion

The four raters in this pilot study demonstrated high reliability but poor agreement in determining the gross edge of MCTs and STSs. Interpreted in concert, this indicates that within-rater measurements are consistent but raters do not actually measure the same point when compared to raters other than themselves.

This lack of agreement may affect surgical margin planning and resection (i.e. “surgical dose”), tumor staging, and assessment of response to therapy—all of which may influence clinical decision making and quality of life for veterinary cancer patients.

Because of the potential implications of these findings, repetition of this experiment, ideally with a larger number of raters and dogs, would be a worthwhile effort.