



Assessing the inter- and intra-animal variability of in vivo OsteoProbe skeletal measures in untreated dogs



Erin M.B. McNerny^a, Jason M. Organ^a, Joseph M. Wallace^b, Christopher L. Newman^a,
Drew M. Brown^a, Matthew R. Allen^{a,b,*}

^a Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN, United States

^b Department of Biomedical Engineering, Indiana University Purdue University of Indianapolis, Indianapolis, IN, United States

ARTICLE INFO

Article history:

Received 25 April 2016

Received in revised form 17 July 2016

Accepted 8 August 2016

Available online 8 August 2016

Keywords:

Indentation
Bone mechanics
OsteoProbe
Bone quality
Variability

ABSTRACT

The OsteoProbe is a second-generation reference point indentation (RPI) device without a reference probe that is designed to simplify RPI testing for clinical use. Successful clinical implementation of the OsteoProbe would benefit from a better understanding of how its output, bone material strength index (BMSi), relates to the material properties of bone and under what conditions it reliably correlates with fracture risk. Large animal models have the potential to help fill this knowledge gap, as cadaveric studies are retrospective and limited by incomplete patient histories (including the potential use of bone matrix altering drugs such as bisphosphonates). The goal of this study was to assess the intra and inter-animal variability of OsteoProbe measures in untreated beagle dogs ($n = 12$), and to evaluate this variability in comparison to traditional mechanical testing. OsteoProbe measurements were performed in vivo on the left tibia of each dog and repeated 6 months later on the day of sacrifice. Within-animal variation of BMSi (CV of 5–10 indents) averaged 8.9 and 9.0% at the first and second timepoints, respectively. In contrast, inter-animal variation of BMSi increased from 5.3% to 9.1%. The group variation of BMSi was on par with that of traditional 3-point mechanical testing; inter-animal variation was 10% for ultimate force, 13% for stiffness, and 12% for total work as measured on the femur. There was no significant change in mean BMSi after 6 months, but the individual change with time across the 12 dogs was highly variable, ranging from -12.4% to $+21.7\%$ (mean 1.6%, SD 10.6%). No significant correlations were found between in vivo tibia BMSi and femur mechanical properties measured by ex vivo 3-pt bending, but this may be a limitation of sample size or the tests being performed on different bones. No relationship was found between BMSi and tissue mineral density, but a strong positive correlation was found between BMSi and tibia cortical thickness ($\rho = 0.706$, $p = 0.010$). This report shows that while the OsteoProbe device has inter-individual variability quite similar to that of traditional mechanical testing, the longitudinal changes show high levels of heterogeneity across subjects. We further highlight the need for standardization in post-testing data processing and further study of the relationships between OsteoProbe and traditional mechanical testing.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

A significant limitation in the quest to reduce fracture risk has been the inability to predict that risk on an individual basis. Bone mineral density (BMD) has long been the gold standard for assessing fracture risk, however the limitations of such imaging are well appreciated (Kanis, 2002; Sarkar et al., 2002). This limitation of BMD is not surprising given that the fracture resistance of a bone is multi-factorial – determined by not only bone mass (which can be estimated by BMD) but also its geometry/architecture and its material properties. Advances in

clinical imaging have made measures of geometry/architecture possible while assessing material properties has proven more difficult.

Indentation methods have long been used to assess bone's material properties (see reviews (Thurner, 2009; Zysset, 2009)). Indentation methods have recently been adapted into two devices that facilitate in vivo measurements of local mechanical properties. BioDent, an early generation device originally proposed for clinical use (Diez-Perez et al., 2010), utilizes a reference probe-based indentation method which cyclically indents the bone surface to provide various pieces of data linked to tissue mechanical properties (e.g. indentation depth, unloading slope, energy dissipation) (Bridges et al., 2012; Hansma et al., 2008). This technique has been shown to differentiate fracture/non-fracture patients (Diez-Perez et al., 2010) yet the majority of data have been generated from laboratory experiments, either on bone specimens or in vivo using animal models (Beutel and Kennedy, 2015;

* Corresponding author at: Dept. of Anatomy and Cell Biology, MS 5035, Indiana University School of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202, United States.
E-mail address: matalen@iupui.edu (M.R. Allen).

Granke et al., 2014; Gallant et al., 2013; Aref et al., 2013). A newer generation device, OsteoProbe, designed in part to make clinical use more straightforward (Bridges et al., 2012), uses similar yet distinct methods to assess indentation properties. It lacks a reference probe and instead relies on a pre-load trigger reference point followed by a single high-force impact from which the depth of penetration into the bone is determined (Bridges et al., 2012). This device has been used in several clinical studies and was able to differentiate bone properties in diabetic/non-diabetic patients (Farr et al., 2014; Furst et al., 2016), effects of glucocorticoids/treatments (Mellibovsky et al., 2015), fracture/non-fracture osteopenic patients (Malgo et al., 2015), and Norwegian/Spanish women (Duarte Sosa et al., 2015), but BMSi did not show an association with vertebral fractures in a cohort of older women (Rudäng et al., 2016). While the OsteoProbe has potential to fill a unique niche in our ability for in vivo assessment of bone, a number of basic, yet fundamentally essential, questions exist about the device such as the variability and reproducibility of the measures. The goal of this study was to assess variation in the OsteoProbe measure, bone material strength index (BMSi), within (using longitudinal measures over 6 months) and across a group of untreated beagle dogs in order to define properties of variability and reproducibility in a relatively homogenous setting of tissue properties.

2. Methods

2.1. Experiment design

Twelve skeletally mature female beagle dogs (455 ± 59 days) were used in this study. Due to housing constraints, these animals were divided across two evenly sized cohorts, housed sequentially over two years. The first cohort was older than the second (starting age Cohort 1: 512 ± 7 days, Cohort 2: 399 ± 10 days, $p < 0.01$). These untreated animals comprised the vehicle control group for a larger study and received daily oral saline for 1 year. Six months into the study, the dogs were anesthetized via intravenous propofol to allow in vivo indentation (OsteoProbe, Active Life Scientific). At 12 months, in vivo indentations were repeated at the same site within 15 min following animal sacrifice (sodium pentobarbital overdose) and prior to tissue dissection. Right femora and left tibiae were collected, wrapped in saline-soaked gauze, and frozen at -20°C until use. Right tibiae were formalin-fixed and stored in ethanol. All animal procedures were conducted with prior approval of the Indiana University School of Medicine Animal Care and Use Committee.

2.2. OsteoProbe

OsteoProbe measurements were performed in vivo on the left tibia of each dog (6 month timepoint) and repeated 6 months later on the day of sacrifice (12 month timepoint). Both tests were done on the same leg, as the contralateral limb was dedicated to another assay. The location for testing was identified as the linear midpoint between the superomedial margin of the medial tibial condyle and the distomedial margin of the medial malleolus. This tibia mid-diaphyseal location has minimal soft tissue covering the bone in both canines and humans and is the preferred site in clinical tests (Farr et al., 2014; Furst et al., 2016; Mellibovsky et al., 2015; Malgo et al., 2015; Duarte Sosa et al., 2015; Rudäng et al., 2016; Guerri-Fernandez et al., 2014). The skin was shaved, the site was aseptically prepared, and local anesthetic (Bupivacaine + Lidocaine) was injected subcutaneously just proximal to the testing site. The test probe was carefully inserted through the lifted skin prior to the first indentation. The operator (JMO) used his opposite hand to move the skin with the probe for each indentation to prevent tension from the stretched skin from interfering with the probe (Fig. 1). The probe was positioned normal to the bone surface and the device was slowly lowered over 1–2 s to activate the indentation cycle, which monitors the indentation depth increase



Fig. 1. OsteoProbe in use in vivo.

resulting from an impact load of 30 N superimposed on a 10 N triggering preload (40 N total force) (Bridges et al., 2012). Each OsteoProbe measurement session consisted of 5 indentations located at least 2 mm apart along a line parallel to the long axis of the diaphysis and was performed without removing the probe from the skin between indents. The direction of this spacing was switched (moving proximally or distally from the midpoint) between 6 and 12 months to avoid indenting a previously tested site. In select cases, 1–5 additional measurements were made based on our assessment that one or more indents were questionable and warranted flagging for further evaluation. Five indents on the manufacturer-provided poly(methyl methacrylate) block were performed immediately following each bone test to allow calculation of the BMSi. A new probe tip was used for each animal at each timepoint. The probe tips are designated for single patient use and were used as provided by the manufacturer.

Raw, uncorrected indentation results were used to manually calculate BMSi using the published equation (Bridges et al., 2012). The reasons for manual calculations, rather than using the provided software's values, were twofold. First, because animals were split across two cohorts, the full dataset was collected at four time-points spanning 2 years. The OsteoProbe instrument was returned to the manufacturer between time-points, which allowed for calibration and inspection, but also resulted in an update to the supplied software mid-experiment. This update slightly changed the software-based calculation of BMSi (details of the manual calculations and software changes are provided in the Supplement). Though differences in the BMSi values following the software update were subtle (Supplement Table), we used manual calculations to ensure consistency across the full dataset. Manual calculation also allowed us the opportunity for post-testing decisions about the inclusion of individual indents, particularly PMMA indents used in the calculation of corrected BMSi. This was essential when an outlying (later identified as invalid) PMMA indent was overlooked during testing, resulting in a warning of a high PMMA standard deviation from the OsteoProbe software and a suggestion to repeat the test. We opted

to manually calculate the corrected BMSi with the invalid indent excluded rather than exclude the entire measurement and subject the animal to repeated testing.

Ex vivo OsteoProbe indentations were performed to test the effect of indentation location on BMSi and examine potential differences in variation ex vivo. Four fresh-frozen dog tibiae, saved from an unrelated experiment, were stripped of soft tissue and indented 15 times. The indentation sites were arranged in a grid of 3 indents across the anterior-medial circumference by 5 indents along the proximal length of the bone, ending at the mid-diaphysis. The OsteoProbe was positioned normal to the bone surface for all indentations, and the resulting BMSi measurements were averaged by indentation site.

2.3. pQCT

Right tibiae and femora were assessed ex vivo for volumetric bone density and geometry using a Norland Stratec XCT Research SA + pQCT (Stratec Electronics, Birkenfeld, Germany). A single slice at the bone's midpoint was scanned using a resolution of $0.07 \times 0.07 \times 0.50$ mm. Volumetric density and standard geometry parameters were obtained using standard scanner software with a segmentation threshold of 690 mg/cm^3 .

2.4. Whole bone mechanical testing

Three-point bending to failure was performed on the right femora on a servohydraulic testing system (MTS 858 MiniBionix II, MTS Systems Corporation). Whole bones were placed on a 5 cm support span, with the anterior surface tested in tension and the mid-diaphysis placed directly below the center loading point. Hydration of the bones was maintained with saline irrigation. Bones were loaded under displacement control (1 mm/s) until failure. Recorded load and displacement data were analyzed with a custom Matlab script and converted to estimated stress-strain curves using engineering beam theory and each bone's cross-sectional geometry determined by pQCT. Yield was defined using the 0.2% strain offset method. Both whole bone (structural) and estimated tissue level (material) properties were determined. Primary properties of interest were ultimate force, stiffness, displacement (pre-yield, post-yield and total), and work (pre-yield, post-yield, and total).

2.5. Statistical analysis

The intra-animal coefficient of variation (CV) of BMSi was calculated and averaged by cohort (1 or 2) and time point (6 and 12 months). The inter-animal (group) CV was calculated for mean BMSi, final bodyweight, and ex vivo measures including bone mineral density, cortical geometry, and mechanical outcomes from whole bone mechanical testing. Group

CVs were examined within each cohort (all measures) and time-point (BMSi only). Change in BMSi between 6 and 12 months was tested with a paired *t*-test. The relationships between 12 month BMSi and measures from pQCT and traditional mechanical testing were tested by Spearman's nonparametric correlation analysis. This nonparametric test was chosen over Pearson's correlation to allow for the possibility of non-linear relationships.

3. Results

Endpoint bodyweight and morphological characteristics of the tibia and femur showed no significant differences between cohorts (Table 1). Overall inter-dog CVs for pQCT measures ranged from 0.6% for femur tissue mineral density (TMD) to 14.4% for tibia cortical thickness (Ct. Th). Tibia measures were consistently more variable than corresponding measures on the femur.

The individual indentation results, plotted by animal and time-point, are presented in Fig. 2. Analysis of the indentation curves for the full dataset revealed that 4 indents were the result of abnormal tests based on the shape of the curve having distinct deviation from the typical curve (Fig. 3). These indentations were excluded from further analysis. Subsequent to the removal of these values, Dixon Q tests (Rorabacher, 1991; Dixon, 1950) were used to test for outlier measurements within dog and time point. One indent met the Q test criteria with >95% confidence and was discarded from further analysis.

The BMSi intra-animal CV was 8.9 and 9.0% for the first and second time-points, respectively, and ranged from 7.8 to 10.0% when considered separately for each cohort (Table 2). Intra-animal CV was not significantly different as a function of timepoint or cohort (Repeated measures 2-factor ANOVA; Timepoint $p = 0.958$, Cohort $p = 0.724$, Interaction $p = 0.512$). Inter-animal variation of BMSi ranged from 5.1 to 9.8% across cohort and time (Table 3) with greater variation observed at the second time-point.

One animal (animal #2 in Fig. 2) at the 12 month timepoint was re-measured immediately following the first test due to machine prompting (suggesting the measurements were unstable). The second set of indents was not used for the primary analysis, but it serves as an example of measure repeatability (Fig. 4). In this case, immediate retesting of the same bone resulted in a BMSi of 76.8 (compared to the initial 69.7).

Percent changes from the first to the second BMSi measurement (6 months apart) were calculated to evaluate the variability of BMSi measures in these untreated animals over time. As a group, there was no significant change in mean BMSi (Table 4, paired *t*-test, $p = 0.85$), but the individual change with time across the 12 dogs was highly variable, ranging from -12.4% to $+21.7\%$ (mean 1.6%, SD 10.6%) (Fig. 5).

Table 1
Animal weight and pQCT measures of bone size and density.

	Dog weight (kg)	Femur length (mm)	Tibia TMD (mg/cm^3)	Femur TMD (mg/cm^3)	Tibia Ct.Area (mm^2)	Femur Ct.Area (mm^2)	Tibia Ct.Th (mm)	Femur Ct.Th (mm)
Overall (n = 12)								
Mean	9.1	100.3	1294.5	1339.0	48.4	45.0	2.5	1.9
SD	0.8	5.1	23.8	8.3	6.5	3.7	0.4	0.2
CV (%)	9.0	5.1	1.8	0.6	13.5	8.2	14.4	9.6
Cohort 1 (n = 6)								
Mean	9.3	101.2	1306.6	1343.3	46.5	45.2	2.5	1.9
SD	0.9	5.4	21.2	6.2	7.0	4.0	0.4	0.2
CV (%)	9.7	5.3	1.6	0.5	15.1	8.8	18.0	12.8
Cohort 2 (n = 6)								
Mean	8.9	99.4	1282.4	1334.8	50.3	44.9	2.6	1.9
SD	0.7	5.2	21.1	8.4	6.0	3.8	0.3	0.1
CV (%)	8.4	5.2	1.6	0.6	11.9	8.4	10.9	6.4
Cohort comparison (p-values)								
Student's 2-tailed <i>t</i> -test	0.402	0.568	0.077	0.071	0.337	0.891	0.435	0.775

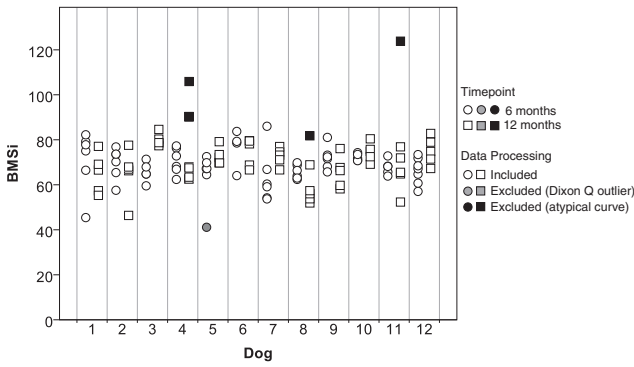


Fig. 2. Individual BMSi values of indentations performed for each animal at 6 and 12 months. Shaded data points were excluded from further analysis due to abnormal indentation curve shape (black) or outlier identification by Dixon Q testing (gray).

Ex vivo tests on 4 unique tibiae showed significant spatial variation in BMSi on the anterior-medial surface both proximally to distally and anteriorly to medially (Fig. 6). The instrument failed to register the test at the 15th indent site (Fig. 6) on all bones tested and the 14th indent site on 2 bones. The variability within each bone averaged $10.2\% \pm 1.3$. Averaging the indents for each bone to a single value, the 4 bones had an average BMSi of 72.3 ± 3.7 (5.1% CV).

Traditional 3-point mechanical testing on the femur determined whole bone mechanical properties including ultimate strength, stiffness, and work to failure (Table 5). There were no mean differences between the cohorts for any mechanical properties. Overall, inter-animal CV was 10% for ultimate force, 13% for stiffness, and 12% for total work. The greatest variability was observed for post-yield measures, with post-yield displacement and post-yield work both having a CV of 17%.

No significant correlations (Table 6) were found between in vivo tibia BMSi at 12 months and femur mechanical properties measured by ex vivo 3-pt bending. However, the trends suggest positive relationships between BMSi and measures of stiffness and ultimate strength and negative relationships between BMSi and measures of deformation (displacement) and energy absorption (work). These trends were mirrored for measures of tissue-level strength (stress) and energy absorption (toughness) normalized for bone size, but these relationships were also not statistically significant in this small sample. No significant relationship was found between BMSi and tissue mineral density, but a strong positive correlation was found between BMSi and tibia cortical thickness ($\rho = 0.706, p = 0.010$).

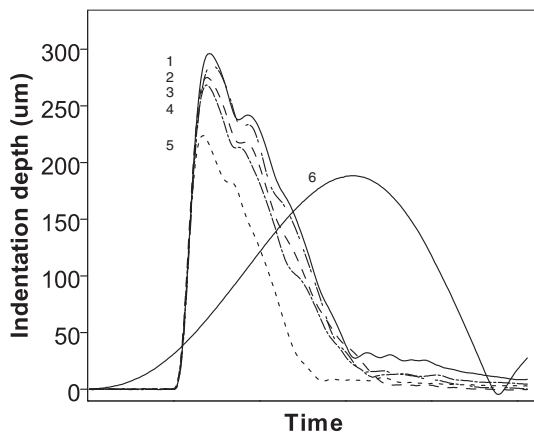


Fig. 3. The plotted indentation data from animal #8 at 12 months shows an example of an atypical test curve (line 6), clearly different from the test curve shape usually observed for bone (lines 1–5). Indentations with abnormal curves were excluded from analysis (Fig. 2).

Table 2
BMSi intra-animal coefficient of variation. Presented as mean (SD).

	Cohort 1 (n = 6)	Cohort 2 (n = 6)	All animals (n = 12)
6 months	10.0 (5.1)	7.8 (5.9)	8.9 (5.4)
12 months	8.8 (5.7)	9.2 (3.4)	9.0 (4.5)

4. Discussion

The study of bone mechanical properties has, until recently, been limited to ex vivo measurements. The reference point indentation (RPI) devices, BioDent and OsteoProbe, were developed to provide an in vivo mechanical assessment of bone tissue properties (Bridges et al., 2012; Hansma et al., 2008). Although BioDent was originally used in patients, this device has been supplanted by the OsteoProbe which has been used in several recent clinical studies (Farr et al., 2014; Furst et al., 2016; Mellibovsky et al., 2015; Malgo et al., 2015; Duarte Sosa et al., 2015; Rudäng et al., 2016). As outlined in a recent review (Allen et al., 2015), although the OsteoProbe may hold promise as a clinical tool, this instrument outputs a novel parameter, BMSi, which requires additional study to help in the interpretation of its meaning. The goal of this study was to contribute to this understanding by assessing the variability of the OsteoProbe machine in a cohort of untreated dogs. We also aimed to examine relationships between BMSi and bone geometry and mechanical properties. The relationship between BMSi and apparent mechanical properties is complicated by bone's heterogeneity, anisotropy, and viscoelasticity, which can be expected to affect the OsteoProbe and whole bone bending tests differently, but the presence of a correlation would be helpful to understanding and interpreting the relatively new BMSi metric.

Based on the manufacturers recommendation and provided software, multiple OsteoProbe indentations are performed and averaged into a single BMSi measurement for a given patient or sample. (Bridges et al., 2012) This is good practice given a bone's heterogeneity and the multiple sources of error possible for any single indentation (e.g. soft tissue interference, angle between the probe and bone surface, probe slippage). The in vivo intra-measure CV of 9% for BMSi in the dogs is on par with the variability reported in two in vivo human studies (9–10%) (Malgo et al., 2015; Duarte Sosa et al., 2015), although a third reports achieving a precision of 1.65%. (Farr et al., 2014) Our ex vivo tests performed on dog tibiae demonstrated significant patterns of spatial variability for BMSi (Fig. 6), with greater differences around the circumference of the bone than down its length. This finding guided our decision to collect in vivo indents in a single line, roughly parallel to the long axis. Despite our best efforts, high spatial variability is a plausible explanation for the 10% change in BMSi found when one measure was immediately repeated (Fig. 4). In the ex vivo tests, the range of BMSi measured for each of the four tibia covered a range of about 20 (~60–80), clustered by location. An early review of the OsteoProbe (Randall et al., 2013) contrasted the difference in variability between a human in vivo test and the plastic standard to demonstrate the effects of bone heterogeneity on BMSi. In this human test using 7 indents, the BMSi ranged from ~80–100, indicating the spread observed in the dog tibia is not unique to the dog model or test operator. It would be relevant to explore whether alterations in bone heterogeneity, often associated with aging and bone disease (Bala and Seeman, 2015), are reflected in OsteoProbe measurement variability and affect test repeatability and statistical power. Alternatively, assessment of BMSi intra-measure

Table 3
BMSi inter-animal coefficient of variation.

	Cohort 1 (n = 6)	Cohort 2 (n = 6)	All animals (n = 12)
6 months	5.1	5.3	5.3
12 months	9.0	9.8	9.1

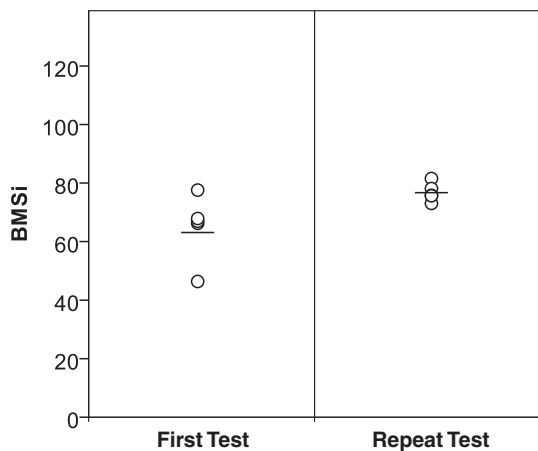


Fig. 4. Individual BMSi values of indentations from initial and repeated OsteoProbe measurements performed on animal #2 at 12 months. Retesting resulted in a mean BMSi of 76.8 compared to the initial mean BMSi of 69.7 ($p = 0.065$, paired t -test).

variation could potentially prove to be a useful way to assess bone heterogeneity.

Overall, the clinical studies report mean *in vivo* BMSi values ranging from approximately 65 to 85, inclusive of individuals with known fragility fractures, (Farr et al., 2014; Furst et al., 2016; Mellibovsky et al., 2015; Malgo et al., 2015; Duarte Sosa et al., 2015; Rudäng et al., 2016; Guerri-Fernandez et al., 2014). The BMSi in these healthy, untreated dogs averaged 69. Our whole bone mechanical test results (Table 3), as well as the health and age of the dogs, do not support the presence of fragility in these animals. This highlights the fact that differences in BMSi between populations or individuals may reflect factors other than compromised bone quality; any characteristic that allows greater probe penetration, such as lower mineralization or osteoid at the bone periosteal surface, could contribute to a reduced BMSi. In order for the OsteoProbe to reach its maximum utility in research or clinical use, additional studies are needed that explore the properties of bone contributing to the BMSi measurement and how their influences differ across species, age and disease. In other words, the relationship between a given BMSi value and bone fragility almost certainly differs across patient populations. Even across comparable populations, very different absolute BMSi values have been reported; two studies examining patients with Type II diabetes in post-menopausal women reported mean diabetic BMSi values of 63.7 (Furst et al., 2016) and 77.2 (Farr et al., 2014), though both reports found reductions in BMSi on the order of 10% compared to their respective matched controls.

One main goal of this study was to assess the longitudinal measurement potential of the OsteoProbe. As opposed to human data, where changes in material properties are most certainly occurring over time in treated or control patients, these untreated dogs likely have modest changes over the 6-month period. The mean BMSi did not significantly change between 6 and 12 months, but the individual animal response was quite variable. Also, the inter-animal CV increased with time, from 5 to 9%, an observation consistent for both cohorts. It is possible these patterns are a result of the tests being repeated on the same bone at the same location, although our approach was designed to prevent direct overlap of indentations. Damage induced by the first round of testing could conceivably affect the second test either directly, if the damage was not fully repaired in the intervening 6 months, or

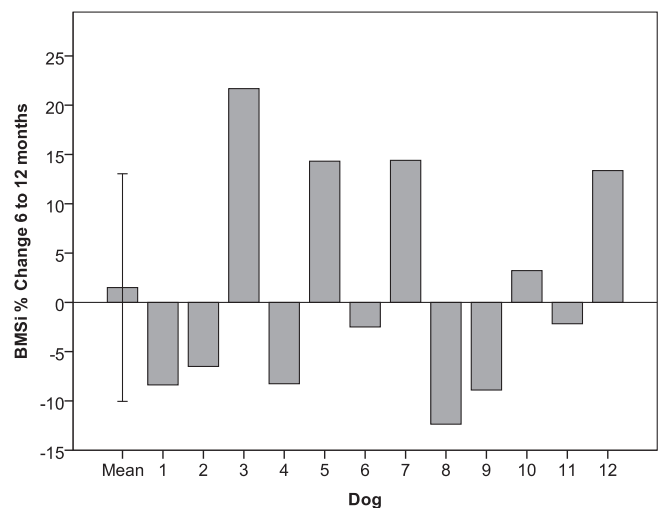


Fig. 5. Percent change in BMSi between 6 and 12 months is presented for each animal (single value) and for the average of all animals (mean \pm standard deviation).

indirectly, if the first round induced an inflammatory or turnover response at the testing site. Future work should explore the damage induced by *in vivo* testing, how this heals over time, and what effect this has on BMSi measures.

To help place the variability of OsteoProbe in context, we compared the 12-month inter-animal variability of the OsteoProbe with that of traditional mechanical properties (Table 3) and morphological parameters including bodyweight, bone density, and bone size (Table 4). Notably, the OsteoProbe's inter-animal variability was less than that of ultimate force, the parameter commonly used when performing power analyses for experiments testing bone mechanics. This is promising for the adoption of OsteoProbe in experiments already powered to detect differences in mechanical properties using traditional *ex vivo* approaches, assuming BMSi mean differences are of a comparable size.

Although variability of the OsteoProbe was on par with that of traditional mechanical testing, there were no significant correlations between tibia BMSi and femur mechanical properties. This may be due to the relatively small group size or the tests being performed on different bones. Despite this, a significant positive correlation was observed between cortical thickness and BMSi. This could potentially be explained by overall deflection or ring deformation of the cortex, resulting in an artifactual increase in the measured indentation increase from impact during OsteoProbe testing. Additional study is needed to understand the limits and behavior of the OsteoProbe with respect to bone size. Since the device was originally designed for clinical use, with human tibia cortical thickness on the order of 4–10 mm at the mid-diaphysis (Capozza et al., 2010), it is possible that the force of the impact is larger than is ideal for use in canines and similarly sized animals.

One aspect of the OsteoProbe device that became an important issue to consider during our study was the handling of irregular/bad tests. The goal of the multiple-indent test session is to achieve a fair representation of the overall bone properties with the final BMSi. However, as with any small sample size, extreme values have disproportionate influence on the mean, whether they stem from measurement error or natural variation. The OsteoProbe software gives the user the ability to flag measurements and thereby remove them from the calculated BMSi. Unfortunately, there is no guidance in the literature for how to standardize this practice. Our laboratory chose to plot indentation distance vs time for each curve during our post-testing analysis. In some cases, we found that outlying values could be identified as invalid measurements by the shape of the indentation curves (Fig. 3). Although these points appeared to be possible outliers when viewed in the context of the sample in which they appeared, the invalid indentations resulted in BMSi values within the range of values from other animals in the experiment

Table 4
Average BMSi by cohort and timepoint. Presented as mean (SD).

	Cohort 1 (n = 6)	Cohort 2 (n = 6)	All animals (n = 12)
6 months	70.3 (3.6)	67.9 (3.6)	69.1 (3.6)
12 months	70.4 (6.4)	68.6 (6.7)	69.5 (6.3)

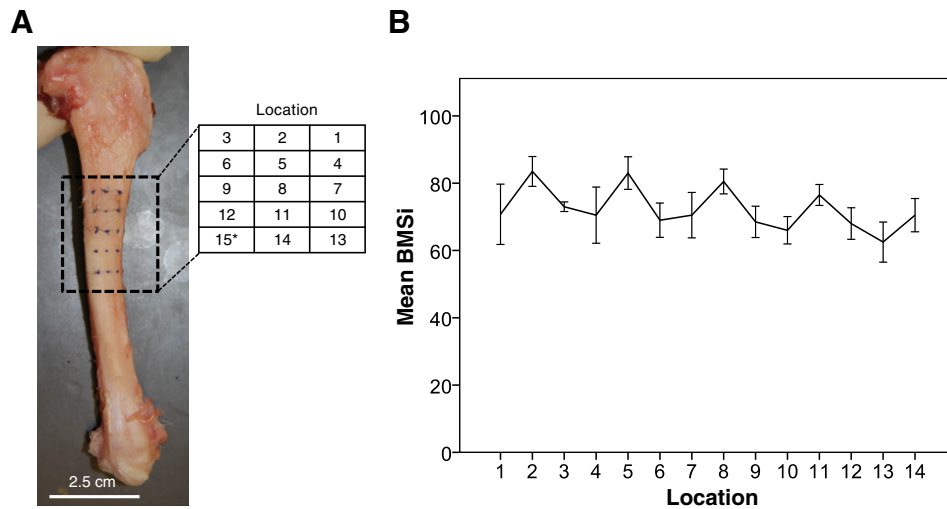


Fig. 6. (A) Ex vivo indentations were performed on dog tibiae ($n = 4$) at 15 locations in a grid on the anterior-medial surface. (B) BMSi (mean \pm st dev) is plotted as a function of indent location. Patterns of spatial variability were observed along both the length and around the circumference of the bone. *Indentations fired but failed to register at site 15 for all bones tested.

(Fig. 2). We propose that until clear guidelines are developed for including/excluding data, users should plot and examine the curves for all indentations, and not just suspected outliers, as a first screen of data quality. In our hands, even after this curve-by-curve inspection, other outlying indentations with valid test curves and no user observations (e.g. the probe slipped) to indicate testing error existed. We used the Dixon Q statistic to objectively identify which of these values were discardable outliers with 95% confidence. The Dixon Q is specifically designed for small sample sizes and is commonly used to detect outliers in replicate measurements. Development of standardized ways to evaluate individual indents will help to assure uniformity in data handling across laboratories and within groups between studies.

There are several limitations to our study. Although some of our variation data are comparable to those published on humans, whether or not the high heterogeneity in change over time in these dogs is more/less than humans remains unknown. It is possible that the higher bone mass of an adult human bone, compared to these dogs, might alter the reproducibility. Our correlation to traditional mechanical properties necessitates using different bones (tibia vs femur) and thus it's possible that stronger relationships would exist if the same bone were used for both measures. Whole bone mechanical tests were performed

on the femur to avoid the possible influence of damage from the indentation tests on the results. While we chose to perform whole bone tests to focus on the correlation of BMSi with apparent level mechanical properties, another approach would be to machine beams of controlled size to more specifically measure tissue level mechanical properties closer to the site of OsteoProbe testing. Mechanical properties measured on these beams might better correlate with BMSi than our whole bone tests, but it would be important to avoid inclusion of the actual indentation sites within the testing span of the beam. It is also possible that stronger correlations would exist to other more clinically relevant skeletal sites, such as the femoral neck or vertebra. Finally, although the fact that we used healthy dogs was an advantage for answering the questions about reproducibility, it is possible that detecting changes over time would have less variability in situations where bone deterioration is occurring.

In conclusion, this report shows that while the OsteoProbe device has inter-individual variability quite similar to that of traditional mechanical testing, the longitudinal measures show high levels of heterogeneity across subjects. We further highlight the need for standardization in post-testing data processing to help assure similarities across centers using this device.

Table 5
Whole bone mechanical properties of the femur at 12 months ($n = 12$).

	Ultimate force (N)	Displacement to yield (μm)	Postyield displacement (μm)	Total displacement (μm)	Stiffness (N/mm)	Work to yield (mj)	Postyield work (mj)	Total work (mj)
Overall								
Mean	1438	1036	1848	2885	1268	642	2517	3159
SD	145	89	322	331	159	64	418	390
CV (%)	10	9	17	11	13	10	17	12
Cohort 1								
Mean	1448	1050	1892	2942	1249	647	2567	3214
SD	210	124	435	442	229	87	556	506
CV (%)	15	12	23	15	18	14	22	16
Cohort 2								
Mean	1428	1023	1805	2828	1286	638	2467	3105
SD	53	47	200	206	56	39	279	282
CV (%)	4	5	11	7	4	6	11	9
Cohort comparison (p-values)								
Student's 2-tailed t-test	0.847	0.658	0.699	0.614	0.744	0.838	0.733	0.689

Table 6
Correlation with 12-month BMSi.

	Spearman's rho	p value
Femur 3-point bending measures		
Ultimate force	0.248	0.489
Displacement to yield	−0.358	0.310
Post-yield displacement	−0.455	0.187
Total displacement	−0.430	0.214
Stiffness	0.358	0.310
Work to yield	0.079	0.829
Post-yield work	−0.236	0.511
Total work	−0.285	0.425
Femur 3-pt estimated tissue level properties		
Ultimate stress	0.442	0.200
Modulus	0.515	0.128
Strain to yield	−0.564	0.090
Resilience	0.018	0.960
Toughness	−0.176	0.627
Tibia pQCT		
TMD	−0.490	0.106
Ct.Area	0.566	0.055
Ct.Th	0.706	0.010

Acknowledgements

This work was supported by the National Institutes of Health (AR 62002). Active Life Scientific provided the OsteoProbe instrument and probes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bonr.2016.08.002>.

References

- Allen, M.R., McNerny, E.M., Organ, J.M., Wallace, J.M., 2015. True gold or pyrite: a review of reference point indentation for assessing bone mechanical properties in vivo. *J. Bone Miner. Res.* 30, 1539–1550. <http://dx.doi.org/10.1002/jbmr.2603>.
- Aref, M., Gallant, M.A., Organ, J.M., Wallace, J.M., Newman, C.L., Burr, D.B., et al., 2013. In vivo reference point indentation reveals positive effects of raloxifene on mechanical properties following 6 months of treatment in skeletally mature beagle dogs. *Bone* 56, 449–453. <http://dx.doi.org/10.1016/j.bone.2013.07.009>.
- Bala, Y., Seeman, E., 2015. Bone's material constituents and their contribution to bone strength in health, disease, and treatment. *Calcif. Tissue Int.* 97, 308–326. <http://dx.doi.org/10.1007/s00223-015-9971-y>.
- Beutel, B.G., Kennedy, O.D., 2015. Characterization of damage mechanisms associated with reference point indentation in human bone. *Bone* 75, 1–7. <http://dx.doi.org/10.1016/j.bone.2015.01.019>.
- Bridges, D., Randall, C., Hansma, P.K., 2012. A new device for performing reference point indentation without a reference probe. *Rev. Sci. Instrum.* 83, 044301. <http://dx.doi.org/10.1063/1.3693085>.

- Capozza, R.F., Feldman, S., Mortarino, P., Reina, P.S., Schiessl, H., Rittweger, J., et al., 2010. Structural analysis of the human tibia by tomographic (pQCT) serial scans. *J. Anat.* 216, 470–481. <http://dx.doi.org/10.1111/j.1469-7580.2009.01201.x>.
- Diez-Perez, A., Guerri, R., Nogués, X., Caceres, E., Jesus Pena, M., Mellibovsky, L., et al., 2010. Microindentation for in vivo measurement of bone tissue mechanical properties in humans. *J. Bone Miner. Res.* 25, 1877–1885. <http://dx.doi.org/10.1002/jbmr.73>.
- Dixon, W.J., 1950. Analysis of extreme values. *Ann. Math. Stat.* 21, 488–506. <http://dx.doi.org/10.1214/aoms/1177729747>.
- Duarte Sosa, D., Vilaplana, L., Guerri, R., Nogués, X., Wang-Fagerland, M., Diez-Perez, A., et al., 2015. Are the high hip fracture rates among Norwegian women explained by impaired bone material properties? *J. Bone Miner. Res.* 30, 1784–1789. <http://dx.doi.org/10.1002/jbmr.2537>.
- Farr, J.N., Drake, M.T., Amin, S., Melton, L.J., McCready, L.K., Khosla, S., 2014. In vivo assessment of bone quality in postmenopausal women with type 2 diabetes. *J. Bone Miner. Res.* 29, 787–795. <http://dx.doi.org/10.1002/jbmr.2106>.
- Furst, J.R., Bandeira, L.C., Fan, W.-W., Agarwal, S., Nishiyama, K.K., McMahon, D.J., et al., 2016. Advanced glycation endproducts and bone material strength in type 2 diabetes. *J. Clin. Endocrinol. Metab.* 101, 2502–2510. <http://dx.doi.org/10.1210/jc.2016-1437>.
- Gallant, M.A., Brown, D.M., Organ, J.M., Allen, M.R., Burr, D.B., 2013. Reference-point indentation correlates with bone toughness assessed using whole-bone traditional mechanical testing. *Bone* 53, 301–305. <http://dx.doi.org/10.1016/j.bone.2012.12.015>.
- Granke, M., Coulmier, A., Uppuganti, S., Gaddy, J.A., Does, M.D., Nyman, J.S., 2014. Insights into reference point indentation involving human cortical bone: sensitivity to tissue anisotropy and mechanical behavior. *J. Mech. Behav. Biomed.* 37, 174–185. <http://dx.doi.org/10.1016/j.jmbbm.2014.05.016>.
- Guerri-Fernandez, R., Villar-García, J., Molina-Morant, D., Torres-del-Pliego, E., Garcia-Giralt, N., Vilaplana-Marz, L., et al., 2014. HIV patients have deteriorated bone material properties assessed by in vivo microindentation. *Bone Abstr.* 3, PP213. <http://dx.doi.org/10.1530/boneabs.3.PP213>.
- Hansma, P., Turner, P., Drake, B., Yurtsev, E., Proctor, A., Mathews, P., et al., 2008. The bone diagnostic instrument II: indentation distance increase. *Rev. Sci. Instrum.* 79. <http://dx.doi.org/10.1063/1.2937199>.
- Kanis, J.A., 2002. Diagnosis of osteoporosis and assessment of fracture risk. *Lancet* 359, 1929–1936. [http://dx.doi.org/10.1016/S0140-6736\(02\)08761-5](http://dx.doi.org/10.1016/S0140-6736(02)08761-5).
- Malgo, F., Hamdy, N.A.T., Papapoulos, S.E., Appelman-Dijkstra, N.M., 2015. Bone material strength as measured by microindentation in vivo is decreased in patients with fragility fractures independently of bone mineral density. *J. Clin. Endocrinol. Metab.* 100, 2039–2045. <http://dx.doi.org/10.1210/jc.2014-4346>.
- Mellibovsky, L., Prieto Alhambra, D., Mellibovsky, F., Güerri Fernández, R., Nogués, X., Randall, C., et al., 2015. Bone tissue properties measurement by reference point indentation in glucocorticoid-induced osteoporosis. *J. Bone Miner. Res.* 30, 1651–1656. <http://dx.doi.org/10.1002/jbmr.2497>.
- Randall, C., Bridges, D., Guerri, R., Nogués, X., Puig, L., Torres, E., et al., 2013. Applications of a new handheld reference point indentation instrument measuring bone material strength. *J. Med. Devices* 7, 410051–410056. <http://dx.doi.org/10.1115/1.4024829>.
- Rorabacher, D.B., 1991. Statistical treatment for rejection of deviant values: critical values of Dixon's 'Q' parameter and related subrange ratios at the 95% confidence level. *Anal. Chem.* 63, 139–146. <http://dx.doi.org/10.1021/ac00002a010>.
- Rudäng, R., Zoulakis, M., Sundh, D., Brisby, H., Diez-Pérez, A., Johansson, L., et al., 2016. Bone material strength is associated with areal BMD but not with prevalent fractures in older women. *Osteoporos. Int.* 27, 1585–1592. <http://dx.doi.org/10.1007/s00198-015-3419-0>.
- Sarkar, S., Mital, B.H., Wong, M., Stock, J.L., Black, D.M., Harper, K.D., 2002. Relationships between bone mineral density and incident vertebral fracture risk with raloxifene therapy. *J. Bone Miner. Res.* 17, 1–10. <http://dx.doi.org/10.1359/jbmr.2002.17.1.1>.
- Turner, P.J., 2009. Atomic force microscopy and indentation force measurement of bone. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 1, 624–649. <http://dx.doi.org/10.1002/wnan.56>.
- Zysset, P.K., 2009. Indentation of bone tissue: a short review. *Osteoporos. Int.* 20, 1049–1055. <http://dx.doi.org/10.1007/s00198-009-0854-9>.