



Published in final edited form as:

*J Bone Miner Res.* 2015 September ; 30(9): 1539–1550. doi:10.1002/jbmr.2603.

## True gold or pyrite: A review of reference point indentation for assessing bone mechanical properties *in vivo*

Matthew R. Allen<sup>1,2</sup>, Erin McNerny<sup>1</sup>, Jason M. Organ<sup>1</sup>, and Joseph M. Wallace<sup>3</sup>

<sup>1</sup>Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN, United States

<sup>2</sup>Department of Medicine-Nephrology, Indiana University School of Medicine, Indianapolis, IN, United States

<sup>3</sup>Department of Biomedical Engineering, Indiana University Purdue University of Indianapolis, Indianapolis, IN, United States

### Abstract

Although the gold standard for determining bones' mechanical integrity is the direct measure of mechanical properties, clinical evaluation has long relied on surrogates of mechanical properties for assessment of fracture risk. Nearly a decade ago, reference point indentation (RPI) emerged as an innovative way to potentially assess mechanical properties of bone *in vivo*. Beginning with the BioDent device, and then followed by the newer generation OsteoProbe, this RPI technology has emerged in several papers. In this review we overview the technology and some important details about the two devices. We also highlight select key studies, focused specifically on the *in vivo* application of these devices, as a way of synthesizing where the technology stands in 2015. The BioDent machine has been shown, in two clinical reports, to be able to differentiate fracture versus non-fracture patient populations and in preclinical studies to detect treatment effects that are consistent with those quantified using traditional mechanical tests. The OsteoProbe appears able to separate clinical cohorts yet there exists a lack of clarity regarding details of testing, which suggests more rigorous work needs to be undertaken with this machine. Taken together, RPI technology has shown promising results, yet much more work is needed to determine if its theoretical potential to assess mechanical properties *in vivo* can be realized.

### Keywords

bone; RPI; microindentation; mechanical properties; OsteoProbe; BioDent

### Introduction

A large fraction of skeletal biology research is aimed at reducing the risk of clinical fracture. Although fractures are complex events, it is generally thought that increases in bone mechanical properties translate to increased fracture resistance. Mechanical properties of

bone have been measured since the late 19<sup>th</sup> century and great strides have been made in understanding factors that contribute to the outcome variables. The field has also come a long way in our knowledge of how interventions (altered loads, pharmacological agents, genetic signaling) alter bone properties. Although the gold standard for determining mechanical integrity is the direct measure of whole bone mechanical properties, there are certainly ethical (and recruitment) roadblocks to performing whole bone bending/compression tests on patients. We have therefore, as a field, relied on 1) surrogates to bone mechanical properties for clinical assessment of fracture risk and 2) *ex vivo* mechanical assessment of cadaver tissue or preclinical animal models. Surrogate measures of bone strength have advantages and do work for classifying fracture risk<sup>(1)</sup>, yet the development tools that can directly test mechanical properties of patient bone would most certainly increase our ability to better predict fracture on an individual basis.

In 2006 the first report emerged describing an innovative device designed to measure mechanical properties of bone *in vivo*<sup>(2)</sup>. The device introduced a concept called reference point indentation (RPI), a modification of the accepted technology of indentation for assessing mechanical properties of bone<sup>(3,4)</sup>. In this original RPI device, a reference probe was placed on the bone surface to determine the reference point and then an inner test probe was mechanically driven into the bone. Among other outcomes, the indent distance was suggested to reflect mechanical properties of the underlying tissue. Beyond the novelty of being designed for *in vivo* use, this device also utilized cyclic indentation testing. Although cyclic indentation protocols do exist for assessing bone<sup>(5-7)</sup>, the gold standard indentation techniques use a single indent test<sup>(3,4)</sup>. This is an important distinction because whereas single indentations can provide information on pre-yield properties of bone (e.g. hardness and elastic modulus), cyclic indentation provides additional measurements as the material yields. In addition, the device created a stress sufficient to produce small-scale cracks with morphologies similar to fracture-associated cracks<sup>(8)</sup>. Since its initial development, the device has undergone several hardware and name alterations (now referred to as BioDent) and soon began to be used by several laboratories for both clinical and preclinical work. More recently, a newer generation, single indent device (referred to as OsteoProbe) was designed to operate without a reference probe<sup>(9)</sup>, (the reference point here is the position of the probe after a 10N force is achieved) with the goal of removing many of the technical challenges implicit with the BioDent device<sup>(10)</sup> and thus having a greater chance of being used in clinical research.

In general, the skeletal research community seems to have met RPI technology with mixed opinions. On the one hand, this represents a novel, innovative technology that might have the potential to allow clinical assessment of mechanical properties. On the other hand, with so many questions surrounding the devices, most notably questions about what exactly they are measuring and how these parameters relate to traditional mechanical properties, some are quite skeptical. The goal of this review is to overview key papers that have used BioDent and OsteoProbe *in vivo*, highlighting their findings (and in some cases limitations) as a way of synthesizing where the technology stands in 2015.

## BioDent and OsteoProbe: Related devices with distinct and important differences

The specifics of both BioDent and OsteoProbe are explained in great detail in the original scientific instrumentation reports<sup>(2,9,12)</sup>. There are several important machine features that researchers need to be aware of in order to place the various reports into proper context. The largest – and perhaps most important difference is that the BioDent utilizes cyclic indentation tests at relatively low forces (2–10N) applied over several seconds while the OsteoProbe is a single impulse indentation to a higher force (40 N) within a much shorter time interval (0.25 ms). This results in the OsteoProbe having 3-orders of magnitude higher loading rate compared to BioDent (120,000 N/sec versus 40 N/sec). It is also worth noting that the BioDent system indents using load control feedback while the OsteoProbe system generates a standard force (30N following the initial 10N preload) using a spring-based mechanism<sup>(9)</sup>. Other differences between the machines are summarized in Table 1. In the only direct comparison of the two devices to date, measurement of *ex vivo* cadaver bones showed no significant correlation between parameters of the two machines<sup>(11)</sup>. Thus not only are the intrinsic characteristics of the machines different, they measure different properties of bone tissue.

A large part of the confusion between these devices is in their names. What is now commonly called BioDent was originally termed OsteoProbe<sup>(2)</sup> (also later referred to as Tissue Diagnostic instrument<sup>(13)</sup>, OsteoProbe II<sup>(12)</sup>, Bone Diagnostic Instrument<sup>(14,15)</sup>, and eventually BioDent<sup>(16)</sup>). This device is primarily used in preclinical work although the first clinical data were obtained using this device<sup>(8)</sup>. The BioDent is generally considered a mounted device in that it is attached to a base unit during the test (it can be handheld but this is not recommended<sup>(17)</sup>). The system utilizes one of three 700  $\mu\text{m}$  diameter reference probes that differ in their tip morphology: a tri-beveled surface (BP1), a beveled surface with a blunted end (BP2) or a flat concentric surface (BP3). Based on the manufacturer recommendations, BP1 probes are ideal for samples with intact soft tissue as the probe can be used to scrape the soft tissue away from the test site (important for *in vivo* studies). BP2 probes are suggested for *ex vivo* work on large bones, and BP3 probes are suggested for small animal work. Within each of the three reference probes is a similar test probe (a 375  $\mu\text{m}$  diameter, 90 degree cono-spherical, 2.5  $\mu\text{m}$  tip radius). It is essential to note the type of reference probe being used in a study, as the potential exists for it to affect the results. For example, the sharp reference probe associated with the BP1 probe produces a fairly standard linear microcrack in the tissue, both during *in vivo* and *ex vivo* tests (Figure 1A and 1B). It is unclear if BP2 probes produce these cracks, but BP3 probes do not seem to based on qualitative observations from the authors' experiments (Figure 1C). The reference probe crack is likely due to the reference force, as it exists even if the test is not run (Figure 1D). More concerning is that it appears, based on limited qualitative observations by our laboratory, that the reference probe crack can interact with the test probe crack and form additional damage between the two regions (Figure 1E–F). This is supported by data showing that the damage generated doesn't just penetrate down into the bone but also expands radially (Figure 1G)<sup>(18)</sup>. More work is necessary to understand the damage field

associated with these tests and, more importantly, how this field might be affected by various changes in the bone matrix and how this affects the data produced by the machine.

Conducting a test with the BioDent is relatively similar *in vivo* and *ex vivo*, the main difference being in the setup. For *in vivo* tests, in order to measure the reference force, the limb must be extended over a force plate/scale, either the one supplied with the unit or a separate one (the latter necessary for large samples such as dogs<sup>(19)</sup>). It is essential to assure that the bone being tested is secure so that it will not move during the test and the bone surface to be tested is perpendicular to the test probe. Once the reference probe is in contact with the bone, the recommended protocol involves periosteal scraping, where the beveled BP1 reference probe is moved back and forth to scrape the periosteal soft tissue away from the test site. This scraping has been utilized in both human studies<sup>(8,20)</sup>, and in the large animal *in vivo* study<sup>(19)</sup>, but not in the rodent preclinical studies<sup>(21,22)</sup> where it was determined that the bones were too small to reliably scrape in a controlled fashion. *Ex vivo* tests are somewhat simpler to set up, as the bone can be more easily secured (any number of techniques can be used, the key being that the bone is secured to a rigid surface and the testing surface is perpendicular to the probe). Once the bone is secured, the probe is placed on the sample (typically without any scraping as the periosteum can be removed during specimen preparation).

In both *in vivo* and *ex vivo* tests, once the setup is complete, the subsequent processes are the same. A stabilizing reference force is applied to the reference probe to prevent the probe from lifting off the surface and then the test is run. There exists a series of experiments reporting good practices for *ex vivo* testing which provide useful data in determining some of the test settings and showing how they affect the generated properties<sup>(17,23)</sup>. It is also important to recognize when doing *ex vivo* testing that BioDent properties are sensitive to tissue anisotropy so orientation of the specimen is important to control and report<sup>(24)</sup>. However, because far fewer studies have used BioDent for *in vivo* assessment of bone mechanical properties, no such best practices currently exist.

Whether performing *in vivo* or *ex vivo* tests, the system provides a force versus distance plot allowing the user to visualize the data during the test. This is important as it can provide very useful, real time, insight into the quality of the test (Figure 2A–C). Whereas some tests can be identified clearly as good or bad, others are more of a judgment call for inclusion/exclusion. Having trained personnel, who have run many tests and thus can more easily differentiate good/bad tests, or more importantly make the difficult judgment calls, is important. There are plenty of potential explanations for outwardly bad tests (which in our lab are identified as having a combination of high IDI/high TID and awkward looking curves such as in Figure 2C), such as insufficient removal of soft tissue (*in vivo*), irregular bone surface, and porosity<sup>(24)</sup>. However there does exist some number of tests for which the plots fall in a gray area of inclusion/exclusion, potential features of such tests being a long toe-in region on the first cycle or awkwardly shaped subsequent curves (Figure 2B). In our laboratory, the standard operating procedure is to take a replacement measure for any ‘questionable’ curves, with the goal of acquiring an *a priori* number of ‘good’ tests (typically 3–5 in the various *in vivo* studies conducted by our laboratory).

Although the actual testing with BioDent is straightforward, there exist a number of testing parameters that can be customized, such as the number of preconditioning-cycles (studies either don't use these or use 3–5 cycles at 2N force), maximum indentation force (2–10 N; early systems went higher), and cycle number (up to 20). Each of these parameters can be modified by the user and studies have shown such modifications affect the results<sup>(17,24,25)</sup>. In addition, the number of tests run on a particular sample can be varied, although typically 3–5 tests of a sample are used and then averaged to get representative values for each bone.

BioDent tests provide several outcome variables (Table 2) although most of them are interrelated<sup>(11,24)</sup>. Total indentation distance (TID – total distance the probe indents during the test) and indentation distance increase (IDI, difference between the first and last indent distance) have been the most often reported/highlighted variables from these tests. The proposed concept is that TID and IDI are inversely related to traditional mechanical properties (higher TID and IDI equates to lower mechanical properties). In addition to the variables that are generated by the system, other parameters have been investigated using customized data analysis codes<sup>(19,24)</sup>. The value of these codes is that they each have been developed to output cycle-by-cycle data, in addition to the data produced by the manufacture software. This provides the user more freedom to explore/evaluate the output data beyond those provided by the manufacturer. For example, the manufacturer software excludes the first several cycles in its calculation of slopes and energy, the rationale for which is not clear or well validated experimentally.

OsteoProbe is a much simpler system compared to the BioDent<sup>(9,10)</sup>. The handheld device has a single type of test probe which has a 90 degree conical tip with a diameter of about 375 microns and a tip radius of < 10 microns. The device utilizes reference point indentation but does so without a reference probe. There is no preconditioning, no variable force, and no variable number of indentation cycles. Rather, the machine is depressed on the bone surface until a force of 10N (equivalent to the maximum load with the BioDent) is achieved, at which time the machine engages to generate an additional 30N impulse force. The depth of the probe at the time 10N is reached is used as the “reference point”, and the distance the probe moves between 10N and 40N is measured as an IDI. This IDI, measured in microns, is normalized to a PMMA (plastic) standard and the machine outputs a variable called bone material strength index (BMSi). It is important to note that despite its name, BMSi does not measure strength (force), per se, but rather a distance. More specifically, BMSi is a ratio of distances, defined as 100 times the ratio of PMMA IDI/sample IDI. Higher BMSi is suggested to be indicative of higher tissue mechanical properties (note this is opposite of BioDent, where properties such as IDI and TID appear to be inversely related to mechanics). There have been no rigorous studies directly comparing BMSi to traditional mechanical test properties.

As with BioDent, using the OsteoProbe device is relatively straightforward. The key aspect that the test specimen needs to be secured and positioned perpendicular (+/- 10 degrees) to the testing device<sup>(9)</sup>. Once the skin is pierced, the apparatus is rested on the bone surface, depressed slowly (to achieve the 10N trigger force) and then the system engages to a 40N force and the calculated dimensionless BMSi value is plotted on the screen. This BMSi is considered a raw (non-normalized) value in the sense that it is normalizing against an

assumed PMMA indentation value of 150  $\mu\text{m}$  (150  $\mu\text{m}/\text{IDI}$  of sample \* 100). Once all tests on a specimen are completed (typically 5–10 tests are undertaken on each specimen), a PMMA standard is tested 5 times and the average of this IDI is used to calculate the final dimensionless BMSi (with a standard deviation) of the sample.

After each individual test with the OsteoProbe, the software provides the option to flag the data point. A flagged test is not considered in the calculation of the final BMSi, but the data are stored in the system. As opposed to the BioDent, the OsteoProbe provides no visual plot of the test curve. The x-y data from the test (force and time) are saved and thus individual tests plots can be generated by the user – but not in real time. Thus, assessing whether a test should be flagged/removed is a significant challenge, requiring one to rely on the ‘feel of the test’ and on the BMSi value. In our experience, we find this to be a major limitation because it is possible to have odd BMSi values (high/low) that have normal plots or, conversely, to have tests with normal looking BMSi values but very odd plots (Figure 3). Thus, basing a decision to include/exclude data on BMSi alone seems like a slippery slope. But with no other option, this is what most users are likely doing. The most unfortunate aspect of this is that so little information is provided in the published OsteoProbe articles (covered in a later section) concerning flagging/excluding points. Given the influence this could have on the data, more details need to be provided in future studies.

One of the key take home messages of this review is that although there are some similarities between BioDent and OsteoProbe, they are distinctly different in several important ways and thus should be considered unique entities. One should not interpret the results from BioDent to reflect those of OsteoProbe (and vice versa). Unfortunately, the literature is becoming full of such cross-generalizations and for those not familiar with the details, this adds considerable confusion. The clearest illustration of differences between the two machines comes from a study in which they were directly compared in a cadaveric sample<sup>(11)</sup>. Using a standard BioDent protocol (BP2 reference probe), 5 tests were averaged and compared to eight tests using OsteoProbe (some tests were removed if the machine slipped or was not held normal to the surface). The summary statistics showed mean BMSi values of ~83, first cycle ID of 80  $\mu\text{m}$ , TID of 88  $\mu\text{m}$ , and IDI of ~13  $\mu\text{m}$ , all values that are consistent with others in the literature. Most noteworthy, the authors found no significant correlations between BMSi from the OsteoProbe and any variables of the BioDent or between age and parameters from either machine. Based on these data the authors concluded that the two machines are testing different properties of bone tissue.

Perhaps the greatest unknown/potential limitation of both machines is that they are limited, *in vivo*, to testing the outer 100–200 microns of tibial bone. Thus, their ability to inform on the properties at fracture sites (femoral neck/spine) assumes there is a relationship between these fundamentally different bone locations. Recently, this concept was addressed in a cadaver study in which BioDent properties of the tibia were compared to whole bone femoral neck mechanics<sup>(26)</sup>. The results showed a significant negative correlation between IDI and femoral neck ultimate load. Although femoral neck BMD had a greater correlation than IDI to ultimate load, combining the two parameters together significantly improved the relationship. This suggests that there may be some utility in testing tibial periosteal tissue properties for inferring properties at distant sites. However, it does not discount that there

could be changes in trabecular or intracortical bone that affect fracture risk and cannot be picked up by indentation. It is also not clear if the above results are achievable with the OsteoProbe device.

## Clinical studies using BioDent

The first report of *in vivo* assessment using RPI used the BioDent machine<sup>(8)</sup>. The clinical cohort included 27 patients that had experienced a fracture and 8 age-matched controls. Using a BP1 probe, the periosteum of the anterior tibia midshaft was scraped away and then a protocol of pre-conditioning cycles (unspecified number, at 4Hz and 2.5N) followed by 20 cycles to 11N (more recent versions of this device have a 10N maximum force). This was repeated at 5 (or more) locations, although the paper is vague as how many patients had more than 5 tests and why additional tests were conducted in some patients and not others. Using an analysis of covariance, age-adjusted means were compared between the two groups. The results showed that fracture patients had significantly higher IDI (12.3  $\mu\text{m}$  versus 18.1  $\mu\text{m}$ ; +47%) and TID (31.7  $\mu\text{m}$  versus 46  $\mu\text{m}$ ; +45%) compared to non-fracture controls. The plots of individual patient measurements suggest a fair amount of variability in both IDI and TID, with patient values (which were the mean of several tests, not the individual tests) ranging from 10–35  $\mu\text{m}$  for IDI and 28 – 85  $\mu\text{m}$  for TID (Figure 4 A&B). The paper also presented inter-observer variability data, generated by assessing 14 different patients twice: IDI variation was 8.7% while TID was 15.5%. Finally, experiments were conducted on cadaveric material to examine relationships between BioDent properties and traditional mechanical tests. These results, stated to be preliminary by the authors, showed a correlation between IDI and crack growth toughness – although the results are clearly influenced by a single data point with high IDI and low toughness (Figure 4C).

A second clinical report with the BioDent machine aimed to determine if patients having experienced a bisphosphonate-associated atypical femoral fracture (AFF), thought to be due to deficient tissue-level properties<sup>(27)</sup>, could be distinguished from controls<sup>(20)</sup>. Specifically, the study included controls without fracture, controls with fracture, bisphosphonate-treated patients that had not experienced a fracture during treatment, and bisphosphonate-treated patients that had experienced an AFF. A testing protocol similar to the one used in the study above was employed<sup>(8)</sup>, with the exception that only 5 measurements were taken in each subject. Once again, an analysis of covariance was used to obtain and compare age-adjusted means among the groups. IDI group means were between 13–18  $\mu\text{m}$  while TID means were between 36–47  $\mu\text{m}$ . Based on box plot presentation, the range of IDI was about 10–25  $\mu\text{m}$  while TID ranged from ~28–75  $\mu\text{m}$ , both within the range of the study by Diez-Perez<sup>(8)</sup>. Statistical differences in the means of IDI and TID were noted for both typical fracture and atypical fracture groups compared to untreated controls and non-fracture bisphosphonate patients. This suggests that the device was able to differentiate fracture from non-fracture patients, but not able to differentiate bisphosphonate effects. Given the known effects of bisphosphonates on tissue properties such as mineralization, collagen crosslinking, microdamage, and tissue-level mechanical properties<sup>(28)</sup>, this is somewhat surprising and suggests that large changes in properties (or larger numbers of subjects) are needed in order for detection by BioDent testing.

Although not explicitly stated, it is assumed both clinical studies utilized BP1 probes as periosteal scraping was conducted. Reference forces were likely applied, but to what level is unknown. Thus, it can be assumed that reference probe cracks were produced and, in theory these could have influenced the results (Figure 1E–F). It is also unclear in these two reports how variable the *in vivo* tests were (the intra-subject variation) and if any tests were excluded in the analysis. Thus sort of information will be important to have in future reports.

The development of OsteoProbe, specifically designed for *in vivo* human testing, has likely reduced the significance of these early BioDent *in vivo* data as the latter device is not being marketed for clinical use. Yet, significant preclinical work has been undertaken with the BioDent and thus its clinical data provide a nice reference. Contrasting *in vivo* data to the wealth of *ex vivo* data with the BioDent, one thing to note is that the *in vivo* tests appear to consistently generate smaller TIDs. Several measures on *ex vivo* cadaver specimens, using similar protocols to the *in vivo* measures with the exception of using a 10N maximum force and no pre-cycles, show TID values of ~85  $\mu\text{m}$  (range 75 to 95)<sup>(24)</sup> and ~88  $\mu\text{m}$  (range 64–135)<sup>(11)</sup>. It is unclear why *in vivo* measures would be lower, but it could have to do with soft tissue interference that exists despite the best efforts to scrape it away and to penetrate through using preconditioning prior to testing. This is supported by the fact that *in vivo* tests appear to have more low values based on the two published reports. It is also possible that tissue preservation or testing conditions of *ex vivo* samples could affect the tissue properties – this has not been rigorously evaluated. The important point here, which probably also translates to the OsteoProbe, is that the environment of *in vivo* tests is unique, and has the potential to yield quite different results than what are achieved in *ex vivo* experiments. If BioDent is going to be used in pre-clinical studies *in vivo*, it will be important to tease out what explains *in vivo* and *ex vivo* differences or to at least assure that relative differences between groups *in vivo* match those observed *ex vivo* and vice versa.

## Clinical studies using OsteoProbe

To date, four papers have reported OsteoProbe data from clinical studies. The goal of this section is to provide key details about the measurement methods, results, and interpretation. There are a lot of details presented here, but the details are important to understanding and bringing together this early work. It is important to keep in mind that although differences in BMSi are noted in each of the four studies, what this parameter actually represents is unclear, as no studies have rigorously compared it to mechanical properties measured with traditional mechanical methods.

The initial clinical report using OsteoProbe was designed to assess the bone material properties of patients with type 2 diabetes<sup>(29)</sup>. Preclinical studies have shown reductions in tissue-level mechanical properties in diabetic animals using traditional mechanical tests<sup>(30–32)</sup>, making this a logical clinical cohort to perform initial tests of the device. Sixty postmenopausal women, thirty with and thirty without type 2 diabetes, were subjected to OsteoProbe testing on the anterior tibia mid-shaft. Ten separate indentations were made on each subject; there was no mention of whether or not tests were flagged. Patients with diabetes had a significantly lower BMSi compared to controls (–11.7% for raw data; –10.5% after adjustment for differences in BMI) (Figure 5A). The paper also reported a



within-subject variability of 1.65% for BMSi. The authors concluded, in accordance with preclinical work, that patients with type 2 diabetes have reduced bone material-level mechanical properties. The interpretation of results in the discussion relied heavily on the *in vivo/ex vivo* relationships that had been published between BioDent and mechanical properties. As outlined above, the two RPI devices are fundamentally different and thus connecting BMSi to changes in bone mechanical properties was a bit premature, as pointed out in the commentary that accompanied the article<sup>(33)</sup>. The conclusion that BMSi is different between the two clinical cohorts in this study is clear, but what exactly that means from a mechanical standpoint is not.

A second paper using the OsteoProbe technology took an important step forward with the device by making longitudinal measures in each patient<sup>(34)</sup>. The clinical cohort included a complex set of four patient populations, all of which were on glucocorticoid treatment, and then segregated into one of four secondary treatment groups: calcium + vitamin D, teriparatide, risedronate, or denosumab. The cohort was a mixture of males/females, and the groups appeared to differ in many important variables known to influence bone (previous fracture, BMD, age, dose of glucocorticoids) thus some of the group effects should be cautiously interpreted. OsteoProbe measures were made at baseline, 7 weeks, and 20 weeks. Eight indentations were taken on each patient at each time point, patterned as two vertical rows of 4 indents each at the mid-shaft of the tibia. The first of these 8 measures was systematically disregarded, as the authors stated that there was an influence of inserting the probe through the skin on the first test. This is somewhat confusing as the authors also state that the probe was inserted through the skin and periosteum prior to taking a measure. Related to this point, there is some discussion of aberrant tests and measurement flagging by the authors, but the details of this are quite vague and in some cases incorrect (the statement is made that the system detects outliers, it does not – the user defines outliers). Thus it is quite unclear what data were used in the final analysis.

The results of the study show that at baseline, BMSi values ranged from 64 to 89.6. These encompass the entire range of values for control and diabetes patients from Farr et al<sup>(29)</sup>. Seven weeks later, patients treated with teriparatide and denosumab had higher median BMSi values compared to baseline, those treated with risedronate were unchanged, and those treated with calcium and vitamin D had significantly lower values compared to baseline (Figure 5B). As a majority of patients in the calcium + vitamin D group had lower BMSi values at 7 weeks compared to baseline, they were switched to active treatment and excluded from follow-up. At 20 weeks, calcium and vitamin D patients were unchanged from baseline (a result grossly skewed by the exclusion of the patients with reduced BMSi at 7 weeks), while all three other groups had significantly higher values compared to baseline. For teriparatide, risedronate, and denosumab the data suggest improvements in BMSi over the twenty-week study – very interesting findings although a bit difficult to interpret based on the biological mechanisms of these drugs which are quite different. Teriparatide stimulates remodeling (with a positive BMU balance) and periosteal apposition<sup>(35)</sup>, effectively producing enhanced levels of new bone formation. Conversely, denosumab and risedronate suppresses remodeling, with effects on periosteal bone being unclear<sup>(36–38)</sup>. Attempts were made to use multivariate statistics to control for the inhomogenous patient characteristics in each treatment group, although sample size considerations make this a

tenuous approach. Nevertheless, it would have been valuable to see individual responses to help understand if all patients increased BMSi over time or if there was heterogeneity in the response (some increase, some decrease). This also would help to overcome the inhomogeneous patient characteristics in each treatment group.

In a cohort of women from Spain and Norway, OsteoProbe was used in an attempt to explain the known fracture risk difference between these two populations<sup>(39)</sup>. All women had normal BMD and were free of any osteoporosis medications. Eight tests were conducted on the midshaft of the tibia, with no mention of test flagging. The variability within subjects was reported as 9.1%, substantially higher than the 1.65% reported by Farr et al<sup>(29)</sup>. Median BMSi values were 77 and 81 for the Norwegian and Spanish women, respectively, with a range of values from about 60–95 (Figure 5C). These are within the range of the two previous reports<sup>(29,34)</sup>. There was no correlation between BMSi and age or BMSi and BMD (hip) for either cohort (Figure 6A). The authors conclude that differences in BMSi may help explain the higher fracture risk of Norwegian populations compared to those from Spain.

The most thorough report to-date, from the perspective of describing details of OsteoProbe testing, was published by Malgo et al<sup>(40)</sup>. This study aimed to assess how BMSi differed in cohorts with low bone mass, with and without fragility fractures. A total of 63 fracture patients (males and females) were compared to 27 non-fracture patients; both groups had similar BMDs. During testing, each OsteoProbe indent was defined as ‘poor, adequate, or well performed’ – the criteria for these classifications were not stated. Indentations continued until 5 ‘adequate’ measures were obtained; this took 25 tests in some cases. Unfortunately, the number of patients that necessitated this many tests, how many bad tests existed in the various patients, or what data was included/excluded, was not presented. There was no significant difference in BMSi between men ( $81.6 \pm 0.8$ ) and women ( $80.0 \pm 0.7$ ) and no relationship with BMSi to BMD, but a significant negative relationship between BMSi and age across the whole cohort (Figure 6B). The main difference among groups was that patients with osteopenia who had fractured had a lower BMSi compared to those with osteopenia who had not fractured (Figure 5D). This was not the case in patients with osteoporosis, where those who had fractured had similar BMSi to those who had not fractured. BMSi also was not different in patients with osteopenia and osteoporosis. This suggests that patients who fracture, or who have osteoporosis but haven’t fractured, all have lower BMSi compared to patients with osteopenia who haven’t fractured. The implied value here is that a low BMSi value suggests a higher risk of fracture, although a much larger trial would be necessary to show this to be true.

Collectively, these early OsteoProbe reports show promise. The device appears to be able to differentiate disease (diabetes), treatment, and fracture populations from control populations. As outlined above, the reporting of more details regarding measurement flagging is essential in future clinical studies. It will be important to present raw data in addition to any adjusted data, and equally important to report the appropriate statistics to justify use of covariate analysis given that most reports are using these (e.g., is the bivariate relationship between variables the same in all groups examined?). Most importantly with OsteoProbe, there is an urgent need to understand if there are meaningful relationships with traditional mechanical testing-derived properties.

## Pre-clinical studies using *in vivo* RPI

Our laboratory has conducted *in vivo* BioDent assessment of dogs<sup>(19)</sup>, rats<sup>(22)</sup>, and mice<sup>(41)</sup>. The goal of these studies was to assess the utility of the BioDent to detect treatment differences<sup>(19)</sup> or to assess the reproducibility in anticipation of future longitudinal rodent studies<sup>(22)</sup> (41). The dog study protocol was designed to mimic the human studies, using a BP1 probe with periosteal scraping, preconditioning cycles, and a 10 cycle 10N force protocol. Three to five tests were performed on the mid-tibia diaphysis of each animal, matching the location assessed clinically. If a test was found to be unusable during the live animal testing, based on the shape of the test curve, a replacement test was run. Untreated animals (n=6) had IDI values that averaged ~14 $\mu$ m (range 12–16) and TID values that averaged ~170  $\mu$ m (range 100–270). Interestingly, while IDI in these dogs was in the range of the human data<sup>(8)</sup>, TID was considerably higher (over 3-fold). Within animal coefficients of variation were 19% for IDI and 26% for TID. Beyond reporting these characteristics in untreated animals, this paper also documented significantly lower TID and IDI in animals that had been treated with raloxifene<sup>(22)</sup>. This was consistent with previous data using raloxifene in dogs, showing that it enhanced material-level mechanical properties<sup>(42)</sup>.

The popularity of rodent models in skeletal research prompted us to inquire whether the BioDent device could be useful for making *in vivo* measurements in rats and mice. Our assessment of rats (six month old male Sprague Dawley, n=72)<sup>(22)</sup> utilized a BP1 probe. Tests on the tibia diaphysis were run to a 10 N maximum force for 10 cycles and included preconditioning, similar to the human and dog studies. The one difference was that the periosteum was not scraped, due to challenges with the small bone size. Three to seven tests were taken per animal, with tests thrown out that had negative IDI or negative loading slopes during testing (~15% of indents fit these criteria). Within animal (n=72) coefficients of variation were 21% for IDI and 17% for TID. Across animal variation was 25% for IDI and 21% for TID. Mean TID was 121  $\mu$ m while mean IDI was ~11  $\mu$ m; these TID values were over 2-fold higher than in humans<sup>(8)</sup>. A subset of animals was tested again, 28 days later, and showed a 13% lower IDI and a 7% lower TID versus the baseline measure (Figure 7A).

A similar experimental design was used to assess tibial properties in mice (16 week old female C57BL/6), although tests were performed at lower maximum force<sup>(41)</sup>. Twelve animals were tested *in vivo* using BP3 probes, 10 cycles, 2N maximum force with preconditioning but without periosteal scraping. The same indentation evaluation criteria as in rats were used to determine bad tests (30% of indentation tests fit these exclusion criteria). Within animal (n=12) coefficients of variation were 22% for IDI and 15% for TID. Across animal variation was 25% for IDI and 12% for TID. Mean TID was 32  $\mu$ m while mean IDI was ~7  $\mu$ m; these were considerably lower than in the rats/dogs due to the lower test force (2N) and surprisingly similar to humans at least for TID<sup>(8)</sup>. A subset of animals was tested again, 28 days later, and showed a 13% lower IDI and an 11% lower TID versus baseline (Figure 7B).

As a collection, these three papers show the potential utility of the BioDent as a tool for assessing properties *in vivo* across a range of animal models. They provide a basis for determining group sizes based on variation within and between animals, as well as the

change over time in untreated animals. It remains unclear why in both rodent experiments all properties were consistently lower after 28 days, but this finding emphasizes the need for controls in any longitudinal experiment. Examining these data with reference to the human work raises several questions. For both the dog and rat studies, where a similar experimental set-up was used, the TID values far exceeded those of the humans by 2–3-fold. One possible explanation, and something that has received little attention for both BioDent and OsteoProbe, is the potential role of new periosteal bone on these indentation measures. The existence of new periosteal bone formation, which is slow but still taking place in rats & dogs, would produce bone that is less mineralized and thus easier to penetrate (effectively increasing the TID). In the broader sense, because these machines are using distance-based measures, it is possible that similar changes in IDI, TID, or BMSi could occur in a very brittle bone and a very ductile bone, albeit for distinctly different reasons. It will be important for future studies to try and understand interactions between periosteal bone properties and indentation properties.

### What are these devices really measuring?

The introduction of this paper emphasized how two overarching questions surrounding these RPI devices were what they actually measure and how that relates to traditional mechanical properties. A number of papers have examined the relationships of BioDent properties to traditional mechanical properties. The early clinical work included some basic relationships between IDI and fracture toughness (Figure 4C), although the limited sample size and highly influential data point made these relationships quite preliminary<sup>(8)</sup>. Using a more robust sample size, two studies found no significant relationship between BioDent variables and fracture toughness in cadaveric<sup>(43)</sup> or mouse bone<sup>(44)</sup> while other cadaveric bone studies have shown negative relationships with TID and IDI to fracture toughness<sup>(43,45)</sup>. Significant correlations have also been noted between IDI and crack growth toughness<sup>(45)</sup>. Conversely, studies in excised rat, dog, and human bone have shown significant negative relationships between IDI and modulus of toughness derived from bending tests<sup>(24,46)</sup>. Fracture toughness represents how easily a tissue is able to resist crack initiation and propagation – true properties of the material<sup>(47)</sup>. Modulus of toughness, on the other hand, is more of a general parameter of tissue ductility and can be influenced by non-material properties. These data suggest the BioDent is actually measuring more general ductility properties, rather than microcracking.

Several papers have attempted to correlate BioDent properties to structural/compositional properties using *ex vivo* tissue samples. First cycle ID and TID have both been shown to be inversely correlated to mean calcium content and mineral matrix ratio<sup>(48)</sup>. Mixed results have been shown for relationships between BioDent and tissue material density, porosity, and collagen crosslinking suggesting at best these parameters have a minor influence on what this machine is measuring<sup>(11) (26,43)</sup>.

There are no OsteoProbe studies in which BMSi is rigorously compared to traditional mechanical properties. Using a sample size too small to make clear conclusions (n=2), BMSi TID and Vickers hardness differed in the expected directions in the sense that one specimen was inferior to another across all three measures<sup>(10)</sup>. Using *ex vivo* human samples, there

was no significant relationship between BMSi and cortical bone density, porosity, or thickness<sup>(11)</sup>. Given the rising use of OsteoProbe, it is essential to see data on the relationship between BMSi and traditional mechanical test properties (whole bone bending and fracture toughness tests) as have been reported with the BioDent.

## Summary

From the outset we posed the question of whether reference point indentation was ‘for real’ in the sense that it can provide unique insight into bone mechanical properties. The answer, unfortunately, remains to be seen. The BioDent machine has been shown, in two clinical reports, to be able to differentiate fracture versus non-fracture although we certainly don’t need a device to help us make that differentiation. Yet in the preclinical realm, the device has been shown to detect treatment effects that are consistent with traditional mechanical property changes. Furthermore, *ex vivo* tests have shown relationships between BioDent properties and post-yield properties from three-point bending tests. Taken together, this machine seems to hold promise although additional work, as outlined in the text above, would be beneficial. On the other hand, the OsteoProbe, which has overtaken the BioDent in the clinical research arena, has more questions than answers surrounding its utility. It seems able to separate populations, yet the lack of clarity regarding details of testing, compounded with the absence of any work comparing BMSi to traditional mechanics, suggests a significant amount of more rigorous work needs to be undertaken with this machine. One thing that is clear regarding RPI is that it has garnered great interest from the field, likely because of its theoretical potential – the ability to assess material-level mechanical properties *in vivo*. Whether or not this potential can be realized remains to be seen.

## Acknowledgments

All authors contributed equally to the intellectual framework of this paper and approved the final version. This work was funded in part through funding by the NIH (AR60202). We thank Drew Brown for his expertise running/interpreting the BioDent machine over the past several years. The authors have received research support, in the form of machine rental and test probes, from ActiveLife scientific (the maker of both BioDent and OsteoProbe) to perform studies over the years.

## References

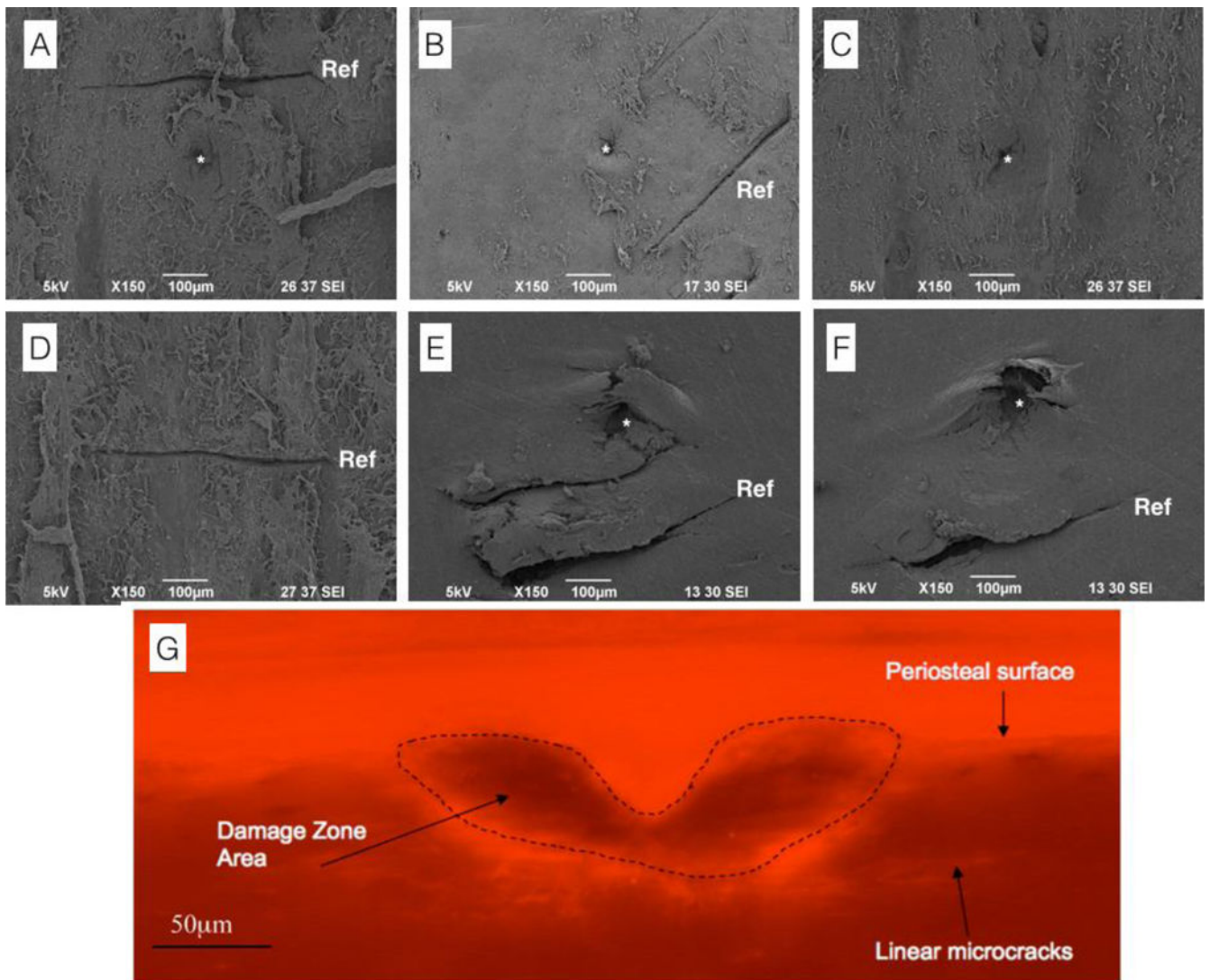
1. Khosla, S. *J Bone Miner Res.* Vol. 18. John Wiley and Sons and The American Society for Bone and Mineral Research (ASBMR); 2003 Jun. Surrogates for fracture endpoints in clinical trials; p. 1146-9.
2. Hansma PK, Turner PJ, Fantner GE. Bone diagnostic instrument. *Review of Scientific ...* 2006; 77:075105.
3. Thurner PJ. Atomic force microscopy and indentation force measurement of bone. *WIREs Nanomed Nanobiotechnol.* 2009 Nov; 1(6):624–49.
4. Zysset PK. Indentation of bone tissue: a short review. *Osteoporos Int.* 2009 Jun; 20(6):1049–55. [PubMed: 19340511]
5. Oliver, WC.; Pharr, GM. *Journal of Materials Research.* Vol. 7. Cambridge University Press; 1992. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments; p. 1564-83.
6. Li X, Bhushan B. A review of nanoindentation continuous stiffness measurement technique and its applications. *Materials Characterization.* 2002 Feb; 48(1):11–36.

7. Zhang Y-R, Du W, Zhou X-D, Yu H-Y. Review of research on the mechanical properties of the human tooth. *Int J Oral Sci.* 2014 Apr 18; 6(2):61–9. [PubMed: 24743065]
8. Diez-Perez, A.; Güerri, R.; Nogues, X.; Cáceres, E.; Peña, MJ.; Mellibovsky, L.; Randall, C.; Bridges, D.; Weaver, JC.; Proctor, A.; Brimer, D.; Koester, KJ.; Ritchie, RO.; Hansma, PK. *J Bone Miner Res.* Vol. 25. Wiley Subscription Services, Inc., A Wiley Company; 2010 Aug. Microindentation for in vivo measurement of bone tissue mechanical properties in humans; p. 1877-85.
9. Bridges D, Randall C, Hansma PK. A new device for performing reference point indentation without a reference probe. *Rev Sci Instrum.* 2012; 83(4):044301. [PubMed: 22559552]
10. Randall, C.; Bridges, D.; Güerri, R.; Nogues, X.; Puig, L.; Torres, E.; Mellibovsky, L.; Hoffseth, K.; Stalbaum, T.; Srikanth, A.; Weaver, JC.; Rosen, S.; Barnard, H.; Brimer, D.; Proctor, A.; Candy, J.; Saldana, C.; Chandrasekar, S.; Lescun, T.; Nielson, CM.; Orwoll, E.; Herthel, D.; Kopeikin, H.; Yang, HTY.; Farr, JN.; McCready, L.; Khosla, S.; Diez-Perez, A.; Hansma, PK. *J Med Device.* Vol. 7. American Society of Mechanical Engineers; 2013 Dec. Applications of a New Handheld Reference Point Indentation Instrument Measuring Bone Material Strength; p. 410051-6.
11. Karim, L.; Van Vliet, M.; Bouxsein, ML. *Bone.* Elsevier B.V; 2015 Apr 7. Comparison of cyclic and impact-based reference point indentation measurements in human cadaveric tibia; p. 1-6.
12. Hansma, P.; Turner, P.; Drake, B.; Yurtsev, E.; Proctor, A.; Mathews, P.; Lelujian, J.; Randall, C.; Adams, J.; Jungmann, R.; Garza-De-Leon, F.; Fantner, G.; Mkrtyan, H.; Pontin, M.; Weaver, A.; Brown, MB.; Sahar, N.; Rossello, R.; Kohn, D. *Rev Sci Instrum.* Vol. 79. American Institute of Physics; 2008. The bone diagnostic instrument II: Indentation distance increase. 064303–064303–8
13. Hansma P, Yu H, Schultz D, Rodriguez A, Yurtsev EA, Orr J, Tang S, Miller J, Wallace J, Zok F, Li C, Souza R, Proctor A, Brimer D, Nogues-Solan X, Mellibovsky L, Peña MJ, Diez-Ferrer O, Mathews P, Randall C, Kuo A, Chen C, Peters M, Kohn D, Buckley J, Li X, Pruitt L, Diez-Perez A, Alliston T, Weaver V, Lotz J. The tissue diagnostic instrument. *Rev Sci Instrum.* 2009 May. 80(5):054303. [PubMed: 19485522]
14. Randall C, Mathews P, Yurtsev E, Sahar N, Kohn D, Hansma P. The bone diagnostic instrument III: testing mouse femora. *Rev Sci Instrum.* 2009 Jun.80(6):065108. [PubMed: 19566227]
15. Thurner PJ, Erickson B, Turner P, Jungmann R, Lelujian J, Proctor A, Weaver JC, Schitter G, Morse DE, Hansma PK. The Effect of NaF In Vitro on the Mechanical and Material Properties of Trabecular and Cortical Bone. *Adv Mater.* 2009 Jan 26; 21(4):451–7.
16. Schultz D, Proctor A, Brimer D. Quantified Touch: A new sense for research and diagnostics. 2013 Aug 8.:1–7.
17. Jenkins T, Coutts LV, Dunlop DG, Oreffo ROC, Cooper C, Harvey NC, Thurner PJ, OSTEO group. Variability in reference point microindentation and recommendations for testing cortical bone: maximum load, sample orientation, mode of use, sample preparation and measurement spacing. *Journal of the Mechanical Behavior of Biomedical Materials.* 2015 Feb.42:311–24. [PubMed: 25455607]
18. Beutel, BG.; Kennedy, OD. *Bone.* Vol. 75. Elsevier Inc; 2015 Jun 1. Characterization of damage mechanisms associated with reference point indentation in human bone; p. 1-7.
19. Aref M, Gallant MA, Organ JM, Wallace JM, Newman CL, Burr DB, Brown DM, Allen MR. In vivo reference point indentation reveals positive effects of raloxifene on mechanical properties following 6 months of treatment in skeletally mature beagle dogs. *Bone.* 2013 Oct; 56(2):449–53. [PubMed: 23871851]
20. Güerri-Fernández RC, Nogués X, Quesada Gómez JM, Torres del Pliego E, Puig L, García-Giralt N, Yoskovitz G, Mellibovsky L, Hansma PK, Diez-Perez A. Microindentation for in vivo measurement of bone tissue material properties in atypical femoral fracture patients and controls. *J Bone Miner Res.* 2012 Dec 18; 28(1):162–8. [PubMed: 22887720]
21. Srisuwananukorn, A.; Allen, MR.; Brown, DM.; Wallace, JM.; Organ, JM. *BoneKEY Reports.* Vol. 4. Nature Publishing Group; 2015 Jun 17. In vivo reference point indentation measurement variability in skeletally mature inbred mice; p. 1-5.

22. Allen MR, Newman CL, Smith E, Brown DM, Organ JM. Variability of in vivo reference point indentation in skeletally mature inbred rats. *Journal of Biomechanics*. 2014 Jul; 47(10):2504–7. [PubMed: 24856912]
23. Coutts, LV.; Jenkins, T.; Li, T.; Dunlop, DG.; Oreffo, ROC.; Cooper, C.; Harvey, NC.; Thurner, PJ.; Arden, NK.; Latham, JM.; Taylor, P.; Baxter, M.; Moss, N.; Ball, C.; Chan, K.; group, TO. *Journal of the Mechanical Behavior of Biomedical Materials*. Vol. 46. Elsevier; 2015 Jun 1. Variability in reference point microindentation and recommendations for testing cortical bone\_ Location, thickness and orientation heterogeneity; p. 292-304.
24. Granke, M.; Coulmier, A.; Uppuganti, S.; Gaddy, JA.; Does, MD.; Nyman, JS. *Journal of the Mechanical Behavior of Biomedical Materials*. Vol. 37. Elsevier; 2014 Sep 1. Insights into reference point indentation involving human cortical bone\_Sensitivity to tissue anisotropy and mechanical behavior; p. 174-85.
25. Setters A, Jasiuk I. Towards a standardized reference point indentation testing procedure. *Journal of the Mechanical Behavior of Biomedical Materials*. 2014 Jun.34:57–65. [PubMed: 24556325]
26. Abraham AC, Agarwalla A, Yadavalli A, McAndrew C, Liu JY, Tang SY. Multiscale Predictors of Femoral Neck in situ Strength in Aging Women: Contributions of BMD, Cortical Porosity, Reference Point Indentation, and Nonenzymatic Glycation. *J Bone Miner Res*. 2015 Jun.:n/a–n/a.
27. Shane E, Burr D, Abrahamsen B, Adler RA, Brown TD, Cheung AM, Cosman F, Curtis JR, Dell R, Dempster DW, Ebeling PR, Einhorn TA, Genant HK, Geusens P, Klaushofer K, Lane JM, McKiernan F, McKinney R, Ng A, Nieves J, O’Keefe R, Papapoulos S, Howe TS, van der Meulen MCH, Weinstein RS, Whyte MP. Atypical subtrochanteric and diaphyseal femoral fractures: second report of a task force of the American Society for Bone and Mineral Research. *J Bone Miner Res*. 2014 Jan; 29(1):1–23. [PubMed: 23712442]
28. Allen MR, Burr DB. Bisphosphonate effects on bone turnover, microdamage, and mechanical properties: what we think we know and what we know that we don’t know. *Bone*. 2011; 49(1):56–65. [PubMed: 20955825]
29. Farr JN, Drake MT, Amin S, Melton LJ, McCready LK, Khosla S. In vivo assessment of bone quality in postmenopausal women with type 2 diabetes. *J Bone Miner Res*. 2014 Apr; 29(4):787–95. [PubMed: 24123088]
30. Reinwald S, Peterson R, Allen M, Burr D. Skeletal changes associated with the onset of type 2 diabetes in the ZDF and ZDSD rodent models. *American Journal of Physiology-Endocrinology And Metabolism*. 2009; 296(4):E765. [PubMed: 19158319]
31. Prisby RD, Swift JM, Bloomfield SA, Hogan HA, Delp MD. Altered bone mass, geometry and mechanical properties during the development and progression of type 2 diabetes in the Zucker diabetic fatty rat. *Journal of Endocrinology*. 2008 Sep 10; 199(3):379–88. [PubMed: 18755885]
32. Nyman, JS.; Even, JL.; Jo, C-H.; Herbert, EG.; Murry, MR.; Cockrell, GE.; Wahl, EC.; Bunn, RC.; Lumpkin, CK., Jr; Fowlkes, JL.; Thrailkill, KM. *Bone*. Vol. 48. Elsevier B.V; 2011 Apr 1. Increasing duration of type 1 diabetes perturbs the strength–structure relationship and increases brittleness of bone; p. 733-40.
33. Jepsen KJ, Schlecht SH. Biomechanical mechanisms: resolving the apparent conundrum of why individuals with type II diabetes show increased fracture incidence despite having normal BMD. *J Bone Miner Res*. 2014 Apr; 29(4):784–6. [PubMed: 24496824]
34. Mellibovsky L, Prieto-Alhambra D, Mellibovsky F, Güerri-Fernández R, Nogues X, Randall C, Hansma PK, Diez-Perez A. Bone Tissue Properties Measurement by Reference Point Indentation in Glucocorticoid-Induced Osteoporosis. *J Bone Miner Res*. 2015 Mar.:n/a–n/a.
35. Hirano T, Burr DB, Turner CH, SATO M, Cain RL, Hock JM. Anabolic effects of human biosynthetic parathyroid hormone fragment (1–34), LY333334, on remodeling and mechanical properties of cortical bone in rabbits. *J Bone Miner Res*. 1999 Apr; 14(4):536–45. [PubMed: 10234574]
36. Kostenuik PJ, Smith SY, Jolette J, Schroeder J, Pyrah I, Ominsky MS. Decreased bone remodeling and porosity are associated with improved bone strength in ovariectomized cynomolgus monkeys treated with denosumab, a fully human RANKL antibody. *Bone*. 2011 Aug; 49(2):151–61. [PubMed: 21457806]
37. Ominsky MS, Stouch B, Schroeder J, Pyrah I, Stolina M, Smith SY, Kostenuik PJ. Denosumab, a fully human RANKL antibody, reduced bone turnover markers and increased trabecular and

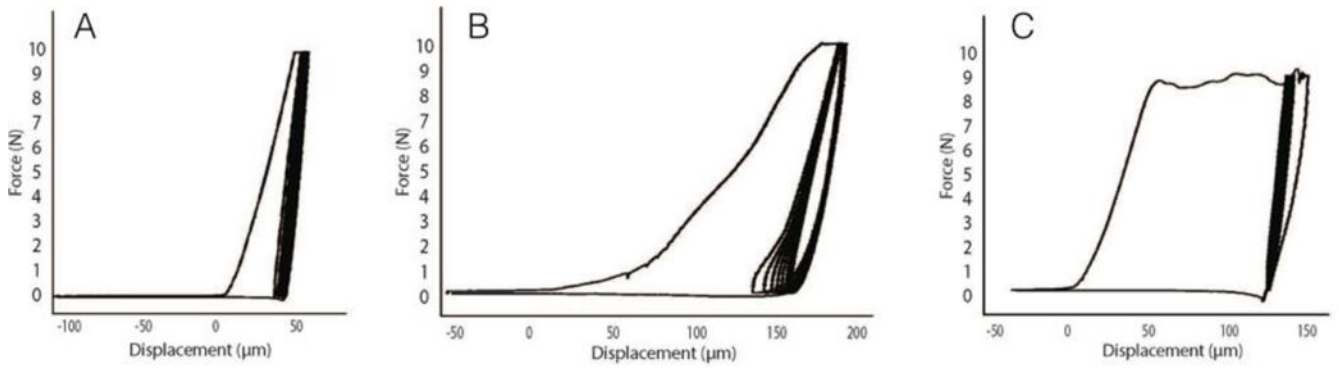
- cortical bone mass, density, and strength in ovariectomized cynomolgus monkeys. *Bone*. 2011 Aug; 49(2):162–73. [PubMed: 21497676]
38. Zebaze, RM.; Libanati, C.; Austin, M.; Ghasem-Zadeh, A.; Hanley, DA.; Zanchetta, JR.; Thomas, T.; Boutroy, S.; Bogado, CE.; Bilezikian, JP.; Seeman, E. *Bone*. Vol. 59. The Authors; 2014 Feb 1. Differing effects of denosumab and alendronate on cortical and trabecular bone; p. 173-9.
  39. Duarte Sosa D, Vilaplana L, Güerri R, Nogues X, Fagerland M, Diez-Perez A, Eriksen E. Are the High Hip Fracture Rates Among Norwegian Women Explained by Impaired Bone Material Properties? *J Bone Miner Res*. 2015 Apr.:n/a–n/a.
  40. Malgo F, Hamdy NAT, Papapoulos SE, Appelman-Dijkstra NM. Bone Material Strength as measured by microindentation in vivo is decreased in patients with fragility fractures independently of Bone Mineral Density. *Journal of Clinical Endocrinology & Metabolism*. 2015 Apr 6. jc.2014-4346-7.
  41. Srisuwananukorn A, Allen MR, Brown DM, Wallace JM, Organ JM. In vivo reference point indentation measurement variability in skeletally mature inbred mice. 2015 May 13.:1–19.
  42. Allen M, Hogan H, Hobbs W, Koivuniemi A, Koivuniemi M, Burr D. Raloxifene enhances material-level mechanical properties of femoral cortical and trabecular bone. *Endocrinology*. 2007; 148(8):3908–13. [PubMed: 17478550]
  43. Granke M, Makowski AJ, Uppuganti S, Does MD, Nyman JS. Identifying novel clinical surrogates to assess human bone fracture toughness. *J Bone Miner Res*. 2015 Jan.:n/a–n/a.
  44. Carriero A, Bruse JL, Oldknow KJ, Millán JL, Farquharson C, Shefelbine SJ. Reference point indentation is not indicative of whole mouse bone measures of stress intensity fracture toughness. *Bone*. 2014 Dec.69:174–9. [PubMed: 25280470]
  45. Katsamenis, OL.; Jenkins, T.; Thurner, PJ. *Bone*. Vol. 76. Elsevier Inc; 2015 Jul 1. Toughness and damage susceptibility in human cortical bone is proportional to mechanical inhomogeneity at the osteonal-level; p. 158-68.
  46. Gallant MA, Brown DM, Organ JM, Allen MR, Burr DB. Reference-point indentation correlates with bone toughness assessed using whole-bone traditional mechanical testing. *Bone*. 2013 Mar; 53(1):301–5. [PubMed: 23274349]
  47. Bonfield W, Datta PK. Fracture toughness of compact bone. *Journal of Biomechanics*. 1976; 9(3): 131–4. [PubMed: 1254608]
  48. Milovanovic P, Zimmermann EA, Riedel C, Scheidt AV, Herzog L, Krause M, Djonic D, Djuric M, Püschel K, Amling M, Ritchie RO, Busse B. Multi-level characterization of human femoral cortices and their underlying osteocyte network reveal trends in quality of young, aged, osteoporotic and antiresorptive-treated bone. *Biomaterials*. 2015 Mar.45:46–55. [PubMed: 25662494]





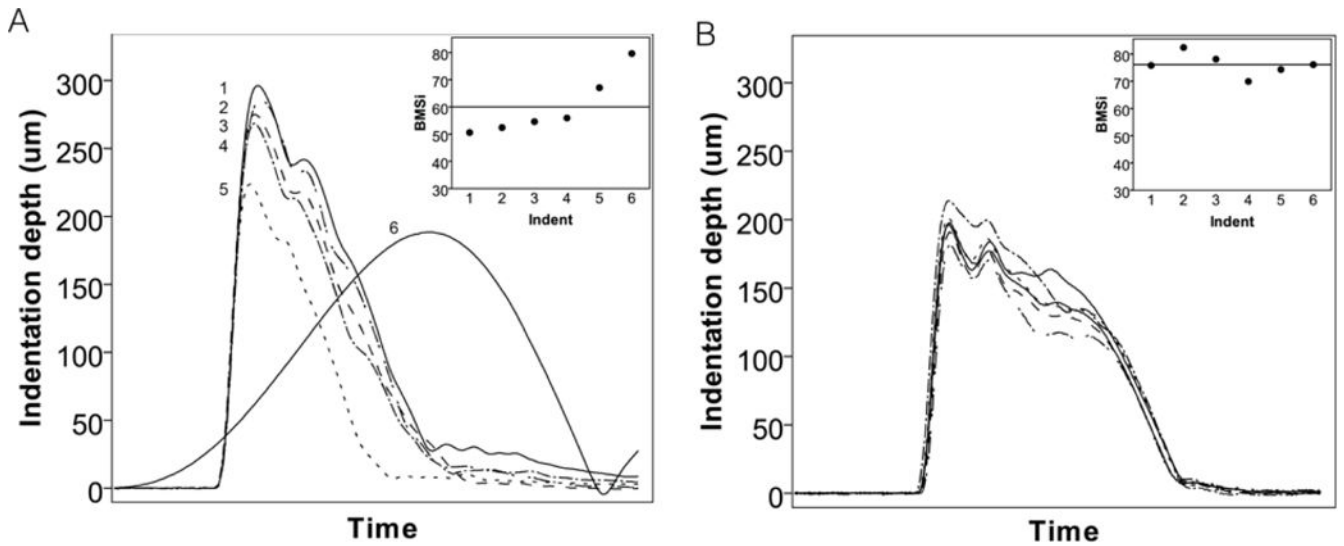
**Figure 1.**

Scanning electron microscopy and photomicrograph images of BioDent-induced damage. (A) Damage induced by an *ex vivo* test using a BP1 reference probe and a 10N 20 cycle test. (B) Damage induced by an *in vivo* test using a BP1 reference probe and a 10N 20 cycle test. (C) Damage induced by an *ex vivo* test using a BP3 reference probe and a 10N 20 cycle test – note the absence of a reference probe crack. (D) Damage induced by a BP1 reference probe following application of a reference force (~10N) without running an actual test. (E–F) Damage induced by an *ex vivo* test using a BP1 reference probe and a 10N, 20 cycle test showing apparent interaction between the reference probe damage and the test region damage. The authors' interpretation, based on this qualitative observation, is that the reference probe and test probe damage can interact. (G) Photomicrograph of the damage zone following an *ex vivo* test and prepared with basic fuchsin staining to visualize microcracks. Note the cracks extending radially outward from the test site. Section was made perpendicular to the plane of testing. Panel G reproduced with permission from<sup>(18)</sup>. In SEM images the reference probe crack is noted by 'Ref', test probe damage noted by '\*'



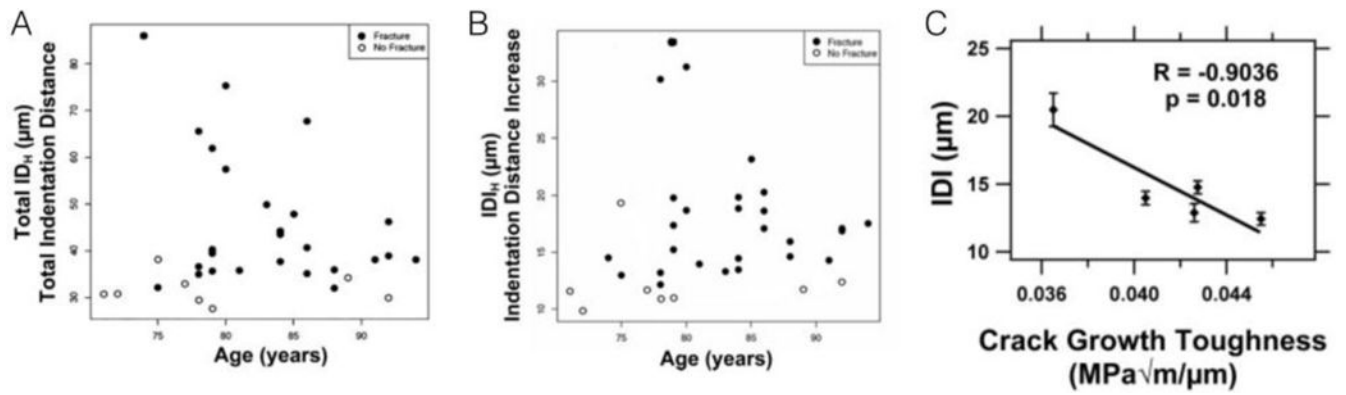
**Figure 2.**

Sample test curves acquired during *in vivo* BioDent testing of dogs. (A) A typical good test with an IDI of 12  $\mu\text{m}$  and a TID of 85  $\mu\text{m}$ . The high quality of the test is based on having data values that are within the range observed in these samples as well as the shape of the curves. (B) A test that falls into a gray area for inclusion. While the IDI is quite typical (11  $\mu\text{m}$ ), the TID is on the high side. More of a concern is the large toe region of the first cycle, as well as the toe region of the subsequent cycles. In our experience, this test would be replaced. (C) A clear bad test in which the IDI (45  $\mu\text{m}$ ) and TID (225  $\mu\text{m}$ ) are both high and the curve is clearly misshapen.



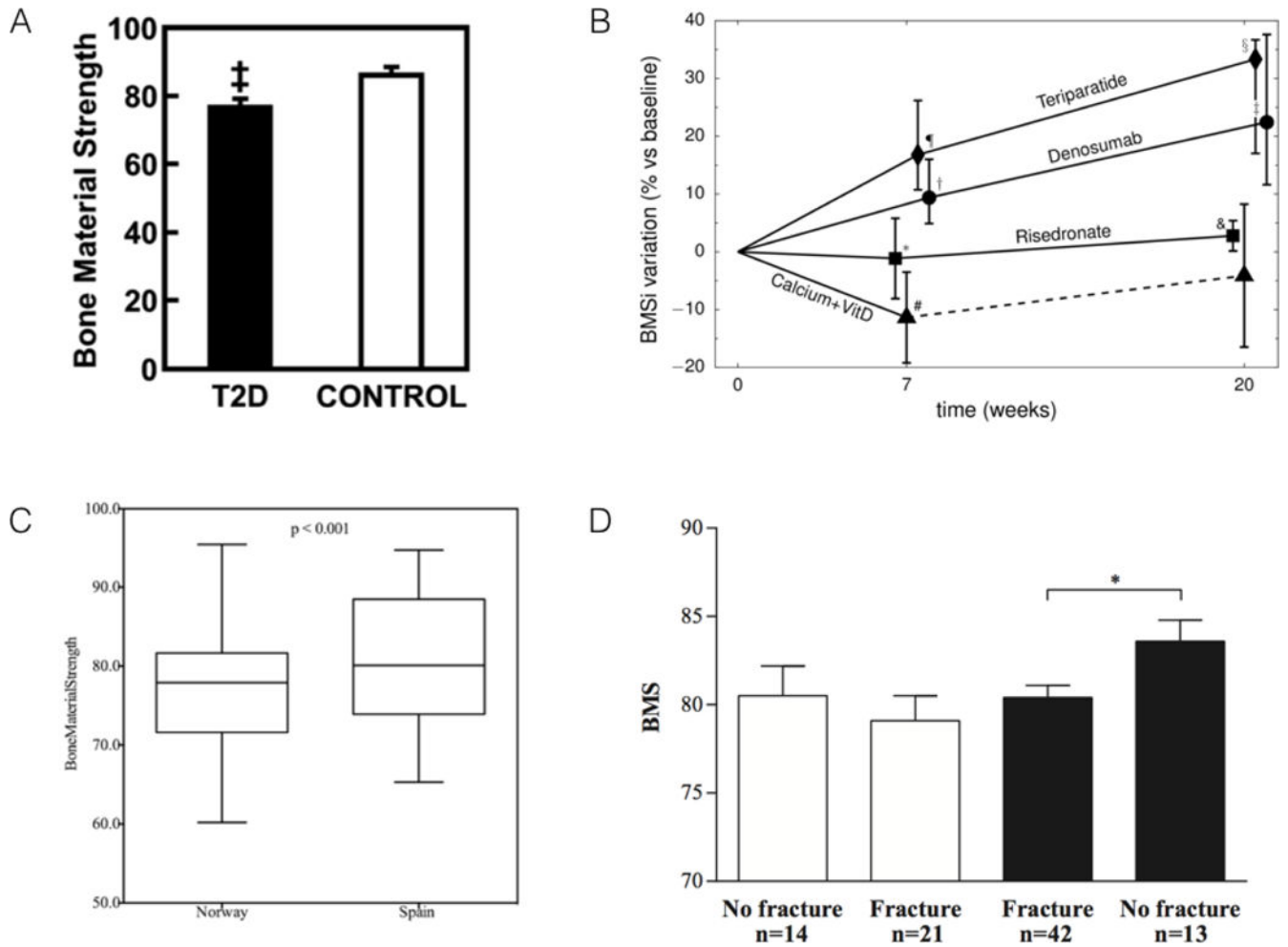
**Figure 3.**

Sample test curves from *in vivo* OsteoProbe tests in two skeletally mature beagle dogs. Curves were plotted using the stored raw data while the BMSi inset represents the raw data provided by the machine during the actual testing session. (A) This animal had four consistent BMSi values followed by two that were notably higher. Examining the curves (post-testing) revealed that while indent 5 had a normal shape, indent 6 was distinctly different and clearly not a valid test. Distinguishing whether either test 5 or 6 are ‘valid’ based on BMSi is challenging as although these values are higher than the mean for this animal, they are in line with values from a different animal (shown in B).



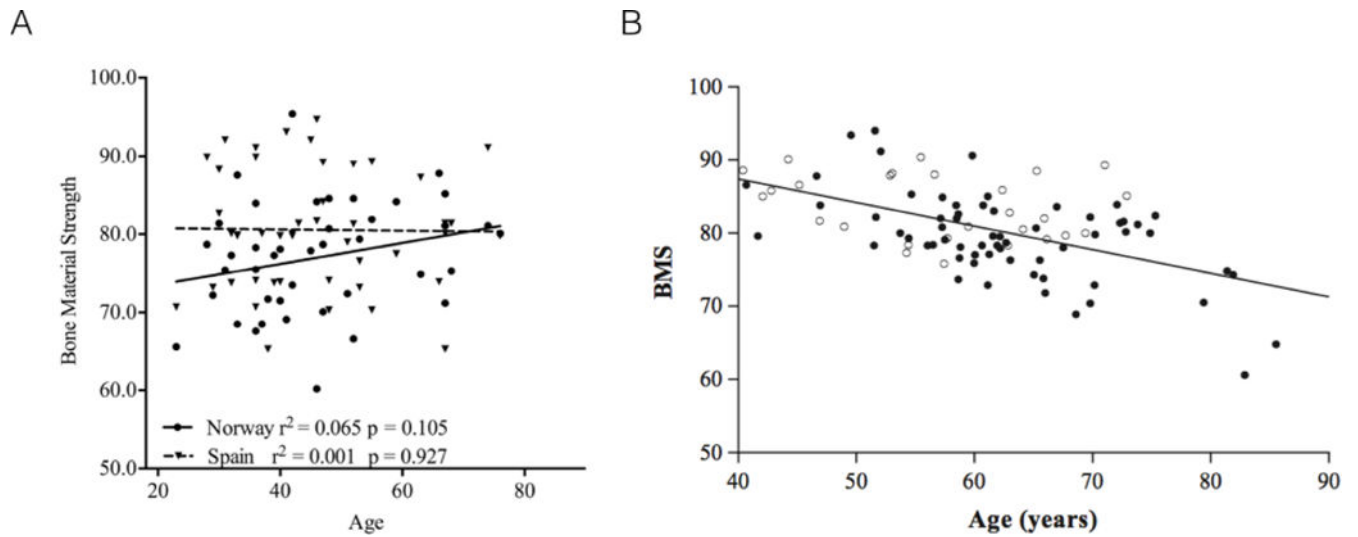
**Figure 4.**

*In vivo* BioDent measures of total indentation distance (A) and indentation distance increase (B) from individual fracture and non-fracture patients. Plots show the variability among patients within each cohort. Relationships between IDI and fracture toughness (both assessed on *ex vivo* specimens) are significant yet clearly driven by one influential point (C). Panels reproduced with permission from<sup>(8)</sup>.

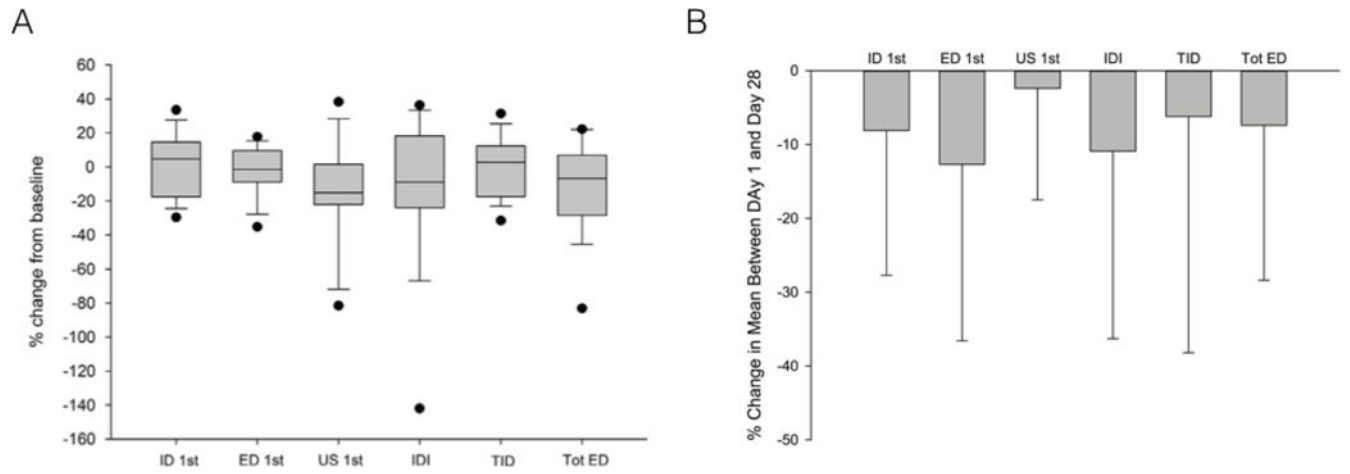


**Figure 5.**

Key figures from each of the four clinical papers using OsteoProbe. (A) BMS is significantly lower in patients with type 2 diabetes (T2D) compared to control. (B) Changes in BMSi over time in patients treated with a combination of glucocorticoids and various other interventions (see text for details). (C) Difference in BMSi between subjects from Norway and Spain – populations known to have different fracture risks. (D) BMSi values for osteoporotic (white bars) and osteopenic (dark bars) patients with and without fracture. Note, despite figure axes using the term Bone Material Strength (or BMS), this is the same as BMSi, which is the more conventional term for this variable. Figures reproduced with permission from<sup>(29,34,39,40)</sup>.



**Figure 6.** Associations between age and BMS have been shown to be non-existent (A) or negative (B). Note, despite figure axes using the term Bone Material Strength (or BMS), this is the same as BMSi, which is the more conventional term for this variable. Figures reproduced with permission from<sup>(39,40)</sup>.



**Figure 7.** Longitudinal measures using BioDent have shown non-significant decreases in most properties after 28 days in both rats (A) and mice (B). Data are plotted as percent difference between baseline and day 28. Figures reproduced with permission from<sup>(22,41)</sup>.

**Table 1**

Characteristics of the two reference point indentation devices

	<b>BioDent</b>	<b>OsteoProbe</b>
Method of use	Mounted *	Hand-held
Manufacturer suggested research target	Pre-clinical	Clinical
Reference probe	Three types	None
Test probe size	~370 microns	~ 370 microns
Probe tip radius	2.5 microns	<10 microns
Overcoming soft tissue	Scraping, preload, preconditioning	Unknown
Loading force **	User defined (2–10 N)	40 N
Loading rate	40 N/sec	120,000 N/sec
Loading cycles	User defined (up to 20)	1
Preconditioning	Optional in cycle number and force	During 10N preload
Data visualization during test	Force versus distance plots and data	Raw (unnormalized) BMSi value

\* Although the majority of studies have used the system mounted, data does exist using the BioDent as a hand-held device<sup>(17)</sup>.

\*\* Biodent uses a true load-control feedback system while Osteoprobe generates a load using a spring-based system without feedback.



**Table 2**

Output variables of the two reference point indentation devices

	<b>BioDent*</b>	<b>OsteoProbe</b>
Variables	1st ID, $\mu\text{m}$ TID, $\mu\text{m}$ IDI, $\mu\text{m}$ Creep ID, $\mu\text{m}$ US (ave of 3-last cycle), $\text{N}/\mu\text{m}$ Energy (ave of 3-last cycle), $\mu\text{J}$	BMSi (dimensionless ratio of indentation distances in sample and plastic)

\* These variables are the ones provided by the manufacturer software. Several groups have written analyses programs to assess cycle-by-cycle data which provides the opportunity to generate additional variables<sup>(19,24)</sup>. ID - indentation distance; TID - total indentation distance; IDI – indentation distance increase; US – unloading slope; BMSi – bone material strength index.