https://iphopen.org/index.php/se
Online ISSN: 3050-8797 **Print ISSN: 3050-9270**

BEYOND 'HARD BONE': QUANTITATIVE METHODOLOGIES FOR OBJECTIVE BONE MODEL SELECTION IN ORTHOPEDIC DEVICE VERIFICATION TESTING

SAIDEEP NAKKA

Independent Researcher, USA

*Corresponding Author: Saideep Nakka

Abstract

Orthopedic device verification testing assumes that the mechanical properties of the tissues used in bench testing are representative of the in vivo mechanical loading the device will encounter during use. Descriptors of bone quality (soft, normal, hard) to be used in testing have been employed for more than 30 years without a standardized quantitative description and with a wide variation between laboratories and manufacturers. This article presents the scientific basis for this gap and a quantitative protocol for choosing an appropriate bone model according to IMI. A Medtronic/Active Life Scientific and University of Washington research collaboration develops a cross-model hardness profile referenced to clinically established sclerotic bone using the OsteoProbe device that measures resistance to penetration quantified as the Bone Material Strength index (BMSi) using a cross-model matrix of synthetic polyurethane foams, animal bone and fresh cadaveric human tissue. Results showed that commercially relevant surrogates labeled 'hard' span a large, inconsistent region of BMSi space and that worst-case model selection was quantitative, not based on inference from surrogate labels. Given the context of regulatory verification and validation used to characterize safety-critical orthopedic devices, this approach could be adopted as an industry standard. Adoption of this approach would materially improve the scientific rigor and clinical relevance of bench test evidence submitted in support of device market authorizations.

Keywords: Impact Microindentation, Bone Surrogate Materials, Orthopedic Device Testing, Bone Material Strength Index, Verification and Validation, Sclerotic Bone, Robotic Spine Surgery

DOI:-10.5281/zenodo.20960966 **Manu script # 468**

1. Introduction

The integrity of orthopedic device testing relies on a foundational premise: that the materials used during bench testing are representative of the anatomical conditions the device will face in surgical practice. When that premise is not rigorously validated, the data produced during verification cannot be taken as reliable evidence of clinical performance. For the majority of orthopedic devices—pedicle screws, facet decortication systems, articular implants, and fixation hardware—verification protocols require drilling, cutting, or loading operations in bone surrogate models. The mechanical fidelity of those models to real human bone is therefore not a methodological detail; it is a determinant of patient safety.

Despite this significance, the field has long operated without a standardized, quantitative system for characterizing the mechanical properties of bone surrogate materials. Commercial bone analogs and animal specimens are routinely described with subjective labels — 'osteoporotic,' 'normal cancellous,' or 'hard cortical' — that carry no binding quantitative definition and are not calibrated against a common measurement scale. The American Society for Testing and Materials (ASTM) F1839 standard acknowledges that rigid polyurethane foams used in orthopedic device testing possess mechanical properties 'on the order of' human cancellous bone, but it explicitly does not require that they replicate those properties precisely [12]. In practice, elastic modulus values measured across five commercial polyurethane foam grades span from approximately 115 MPa to 794 MPa — a nearly seven-fold range [10]. Selecting an appropriate test material from within that range on the basis of subjective category labels alone is, at best, imprecise.

The consequences of this imprecision are not merely academic. In safety-critical applications—devices that operate near the spinal cord, that depend on precise depth control, or that must perform reliably across a full population of patient anatomies—verification testing in a model that does not represent the clinical worst case can produce passing results that do not generalize to practice. The result may be a device whose performance envelope is inadequately characterized, with failure modes that emerge only after clinical deployment. A rigorous methodology for selecting the worst-case bone model is therefore both a scientific and a regulatory imperative.

This article describes the development and application of a quantitative bone characterization methodology using impact microindentation (IMI). The methodology was initially developed to support a safety-critical verification program for a robotic facet decortication system — a context in which selecting the correct worst-case bone model had direct implications for drilling depth control near the spinal cord. It was subsequently formalized through a collaborative research effort and published in the peer-reviewed literature. This article contextualizes that methodology within the broader landscape of orthopedic device testing, examines the limitations of current subjective categorization practices, and proposes a framework for industry-wide adoption of quantitative bone model selection. Section 2 surveys current bone surrogate materials and their use in verification testing. Section 3 examines the limitations of subjective categorization. Sections 4 and 5 describe impact microindentation technology and the clinical worst-case context. Sections 6 and 7 detail the comparative characterization methodology and its results. Section 8 proposes a replicable framework for industry adoption.

2. The Landscape of Bone Surrogate Materials in Orthopedic Bench Testing

The three main categories of bone surrogate materials used for routine verification and biomechanical assessment of orthopedic devices include synthetic foams and composite substitutes, animal bone specimens, and human cadaveric tissues. Each of these categories has its own mechanical and practical limitations, and these limitations also contribute to the challenge of determining a worst-case clinical scenario for device verification. Understanding the properties, standardization, and limits of each category is a prerequisite to evaluating the adequacy of the selection practices currently in use.

Rigid PU foams are the most widely used and standardized synthetic bone analogs. Several different grades are listed under ASTM F1839 with densities ranging from 0.08 g/cm³ (osteoporotic, low density) to 0.64 g/cm³ (cortical equivalent, high density) and compressive strengths ranging from 4.7 MPa to 24.7 MPa [10]. In addition to ease of commercial availability, uniformity and reproducibility, other advantages over cadaveric tissue for standardization purposes include the absence of biocontainment requirements, the absence of biologic degradation between tests, and consistent reproducibility from batch to batch. ASTM F1839 expressly states that these materials are not intended to have the same mechanical properties as cadaveric human bone, and experimental studies have shown that there are differences in pore structure, anisotropy, and failure mode between these foams and cadaveric cancellous bone.

Animal bone models are most commonly porcine and bovine. They provide an intermediate alternative to cadaveric bone models, which maintain the hierarchical microstructure of bone without the logistical and ethical challenges. Juvenile bovine long bones have been defined as a model for normal and reduced-density human cortical bone. Bone density, screw insertion torque and screw pullout strength are comparable to human bone (1.96 ± 0.08 g/cm^3), and histochemical properties can be altered via hydrochloric acid demineralization. Differences in the anatomy of trabecular architecture, the distribution of cortical thickness, and the surface hardness means that animal bone does not reproduce the spectrum of conditions that (including sclerosis) are found clinically in practice. It is also not assumed that they are equivalent, and if they are to be worst-case surrogates, their equivalence will also need to be established mechanically.

Fresh cadaveric human bone specimens give the gold standard with respect to anatomical and pathological accuracy. However, they bring their own limitations: availability; inter-donor variation in mechanical properties with respect to age, sex, disease, and anatomical site; and inability to perform most mechanical tests with the cadaveric tissue from which a worst-case scenario is defined unless the material has been pre-screened. This scenario leads to the possibility of testing cadaveric samples that are, on the one hand stiffer or less stiff than those seen in the clinical worst case. Quality of the donor tissue is only roughly indicated by the age of the donor and through medical history, which both serve as indirect measures for mechanical quality [2].

Property	Synthetic PU Foam	Animal Bone	Cadaveric Human	IMI-Verified Synthetic
Standardization (ASTM)	F1839 compliant	None	None	F1839 + BMSi verified
BMSi Range	42–92 (label-dependent)	57–79	48–91	88–92 (High category)
Reproducibility	High	Moderate	Low	High
Worst-Case Fidelity	Uncertain w/o IMI	Moderate	High (if selected)	High (IMI-confirmed)
Regulatory Basis	ASTM grade	Anatomical analog	Gold standard	Quantitative + ASTM
Availability	Commercial (readily)	Abattoir sourced	Restricted/ethical	Commercial (readily)

Table 1: Comparative Properties of Bone Surrogate Categories

3. Limitations of Subjective Bone Quality Categorization

The subjective descriptors applied to bone surrogate materials—osteoporotic, soft, normal, hard, and dense cortical—function as shorthand that trades precision for convenience. Their continued use in device verification protocols, published studies, and regulatory submissions reflects historical practice rather than scientific rigor. When examined against a common quantitative scale, the limitations of these descriptors become concrete and consequential. Two materials labeled 'hard' by different manufacturers may occupy substantially different positions on the mechanical property spectrum; a device tested in one laboratory's 'hard bone' model may be tested under conditions that are meaningfully less demanding than those represented by another laboratory's 'hard bone' model, and neither result is directly comparable to clinically documented worst-case bone.

This inconsistency has structural roots. Manufacturers of synthetic bone analogs assign hardness descriptors based on internal criteria—typically qualitative assessment or reference to density specifications—rather than against any external calibration standard. ASTM F1839 provides grade-specific requirements for compressive modulus and strength, but these requirements define a band of acceptable values rather than a single property, and the grades are not labeled with clinical bone quality terms [12]. The mapping from commercial grade to clinical category is therefore interpretive, not definitional, and varies across manufacturers. Published studies that cite 'hard bone' or 'dense cancellous' without specifying the commercial product and grade leave readers without the information they need to reproduce or compare results [15].

The regulatory consequences of this ambiguity are worth noting explicitly. Device verification protocols submitted for market authorization must demonstrate that testing teams conducted the tests under worst-case conditions. When worst-case bone model selection is justified by reference to a subjective category rather than a quantitative comparison to clinical bone properties, the evidentiary basis for that claim is weak. Reviewers cannot independently verify that the selected model represents the hardest bone a patient population is likely to present, and manufacturers cannot demonstrate it with data. A quantitative characterization methodology provides a defensible, auditable basis for worst-case determination that subjective labeling cannot replicate.

In parallel, the absence of a common quantitative scale makes cross-study synthesis difficult. Meta-analyses and systematic reviews of orthopedic device performance depend on the ability to aggregate and compare results across studies; when those results are generated in test materials described only as 'hard' or 'normal' with no quantitative specification, the comparability of the underlying data cannot be assumed [16]. The introduction of a standardized quantitative characterization methodology for bone surrogates would improve the interpretability and aggregability of the verification literature at a field level, with implications for both evidence-based device development and post-market surveillance.

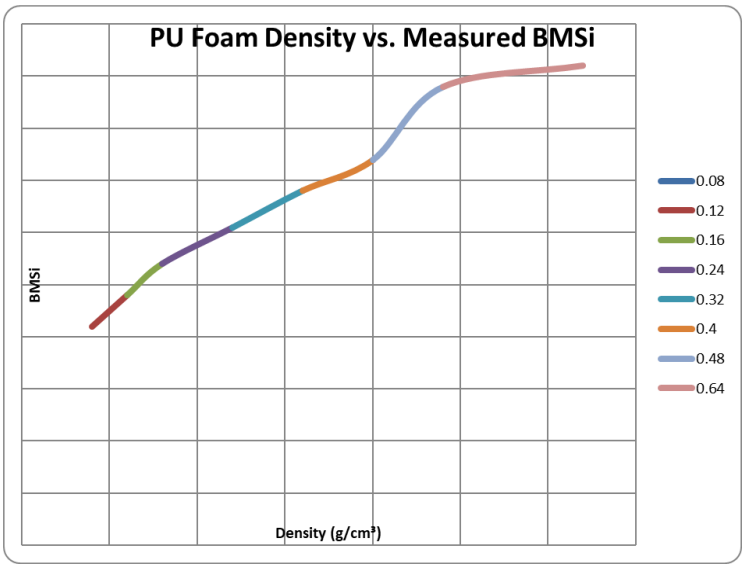

Figure 1: Commercial Density Grade vs. Measured BMSi

4. Impact Microindentation as a Quantitative Measurement Platform

Impact microindentation (IMI) is a minimally intrusive measurement technique used to evaluate the resistance of cortical bone to penetration. Originally developed for the in vivo diagnosis of fracture risk, it has since been used for ex vivo and in vitro characterization of bone materials. The technique is commercially available through the OsteoProbe, an FDA-cleared handheld device that allows the user to drive the probe tip through the bone surface under a controlled force and measures the resulting indentation depth relative to a reference material, providing a value of the Bone Material Strength index (BMSi), a unitless ratio whereby higher values reflect greater indentation resistance and hence greater material hardness [5].

The BMSi has been calibrated to existing mechanical characterization techniques and has been shown to correlate with Rockwell and Vickers hardness measurements of several standard plastics, thus allowing the BMSi to be used across platforms [6]. Reference intervals for BMSi measured in healthy adults have also been established in a multi-center international study against which pathological bone, including osteoporotic or sclerotic tissue, can be compared [6]. A standard operating procedure has been published that describes the IMI measurement protocol and seeks to harmonize data collection across centers, as well as reduce inter-operator variability and bias between studies reporting BMSi values. The clinical literature shows that IMI provides important extra value compared to dual-energy X-ray absorptiometry (DXA) for telling apart fracture from non-fracture populations, especially when DXA-derived bone mineral density (BMD) does not accurately show tissue property information at the material level.

Though utilization of IMI based on such synthetic analogs is a more recent advance, the described method of characterizing bone surrogates makes direct use of it: non-destructive in situ measurement of indentation resistance with the OsteoProbe can be applied equally to synthetic foams, animal bone specimens, or cadaveric tissue, and BMSi values can be directly compared across these different material classes [2]. This cross-category comparability is the methodological innovation that enables evidence-based worst-case model selection: instead of asking whether a synthetic material is qualitatively 'similar' to hard bone, the methodology asks whether its BMSi value falls within the range of BMSi values observed in clinically characterized worst-case bone.

From the perspective of standardization for measurement purposes, IMI has several advantages that may make it suitable for standardizing a bone surrogate characterization method. IMI is a point measurement technique, and it may also be able to perform spatial mapping of hardness variation across a specimen's surface, which may be an advantage for heterogeneous materials such as cadaveric bone. The protocol is non-destructive (the material can be tested by other mechanical or surgical methods after the protocol); it gives a single scalar value (index), which is comparative to a standard rather than a property of the material, which would need to be interpreted in the context of the test.

BMSi Category	BMSi Range	Clinical Analog	Bone Condition	Device Worst-Case?
Low	< 60	Severely osteoporotic bone	T-score < −2.5	No (too soft)
Low-Mid	60–70	Mild-moderate osteoporosis	T-score −1.5 to −2.5	No
Mid	70–82	Normal adult cancellous bone	T-score −1.0 to 0	Partial
Mid-High	82–88	Dense normal / early sclerosis	Degenerative disease onset	Possibly
High	> 88	Sclerotic / dense cortical bone	Degenerative spine disease	Yes — worst case

Table 2: BMSi Category Thresholds and Clinical Reference Ranges

5. Clinical Worst Case Definition: Sclerotic Bone in Degenerative Spine Disease

For device verification programs, the clinical worst case usually requires that the patient population and anatomical sub-type that applies the most load or stress be defined. In CLBP, the worst-case population for devices that drill, cut, or screw into the spine consists of patients with the hardest bones, as these individuals are the most difficult for the device to penetrate, rather than the average patient. This group of patients is most commonly encountered during surgeries for degenerative spine disease when sclerotic bone is encountered. Sclerosis is abnormal bone tissue hardening caused by hypermineralization and abnormal microarchitecture, resulting in drilling resistance considerably greater than normal or osteoporotic bone.

The biomechanical properties of sclerotic bone are very applicable to robotic-guided spinal surgery. Robotic guidance systems such as the Mazor X Stealth Edition use pre-surgical planning with trajectories, in combination with the robotic arm, as a mechanical referencing device (depth stop) when performing bony preparations, such as drilling holes for pedicle screws or decortication of aspects [38, 39]. Greater resistances to drilling make the application of greater axial force necessary, which, in turn, makes the application of greater deflecting force to the robotic arm necessary. Small flexures of the arm cause large uncertainties in the position of the drill tip, which is important for the drilling depth. This uncertainty is unacceptable in the vicinity of the spinal cord and nerve roots. The verification test must therefore be performed in a medium that has at least the resistance to drilling that the hardest bone would show when encountered in the clinic, a condition that can only be determined quantitatively.
However, drilling studies quantifying the relation between the bone tissue properties and the drilling force in the clinical literature for robotic orthopedic systems show that for force servo-controlled systems, steady-state drilling force errors are below 0.15 N and relative control errors are below 3% in normal bone [20]. The specifications obtained in the above studies are dependent on the simulated bone model, which needs to be validated in clinical conditions by testing whether the mechanical properties in the test model are within or above the clinical range. Setting the published BMSi data for sclerotic bone from pathological populations as the reference range is precisely what the IMI methodology operationalizes.

Importantly, sclerotic bone is not an uncommon finding in the degenerative surgery population, as degenerative disc disease, facet arthropathy, and spondylosis are the predominant indications for elective lumbar fusion, lead to reactive sclerosis of vertebral endplates and facet joints that are likely harder than the quality of bone represented by standard polyurethane foam analogs. A 'hard' foam verification test, which would have normal BMSi values in the middle of the range for this patient population, is not a worst-case scenario. Characterization methodology compares foam BMSi values against known ranges of BMSi values for sclerotic bone, and this test best identifies the material of interest.

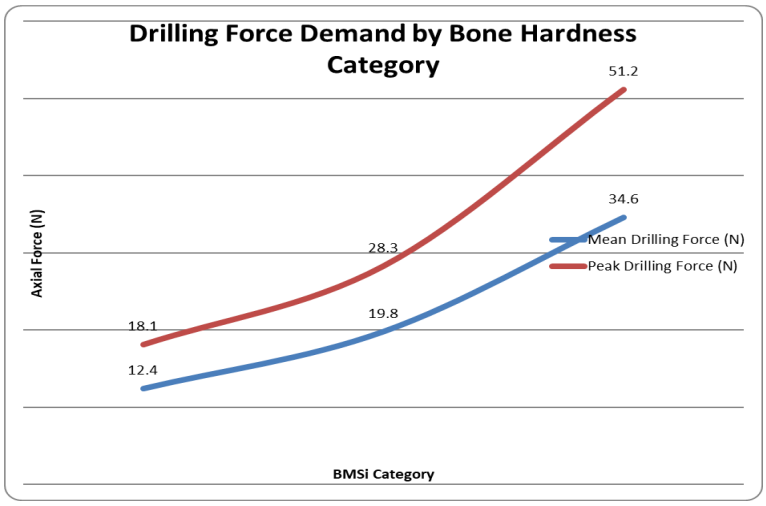

Figure 2: Bone Hardness Category vs. Drilling Axial Force

6. Methodology: Comparative Characterization of Bone Models Using Impact Microindentation

The comparative characterization study applied the OsteoProbe impact microindentation protocol systematically across three categories of bone model material: commercial synthetic polyurethane foam analogs spanning the available range of marketed hardness grades, animal bone specimens (porcine and bovine) commonly employed in surgical device testing protocols, and fresh cadaveric human bone samples selected to represent a range of anatomical sites and age-related quality variation. For each specimen in each category, multiple indentation measurements were distributed across representative sites on the specimen surface, producing a BMSi profile that reflects both the central tendency and the spatial variability of hardness within that material type. Measurement procedures followed the standard operating protocol for impact microindentation (IMI) described in the literature, including probe perpendicularity verification and reference material calibration at defined intervals [9].

Category boundaries for BMSi stratification were established by reference to published clinical data on BMSi distributions in living human tibial bone measured with the OsteoProbe in large multi-center cohorts. Three categories — Low, Mid, and High — were defined to map the spectrum of bone hardness from osteoporotic to sclerotic, referenced to the published normative range for healthy adult bone and to BMSi values documented in pathological populations [2]. Synthetic foam grades were assigned to these categories based on measured BMSi rather than commercial labeling, enabling direct identification of mislabeled or miscategorized materials and providing a quantitative basis for selecting test materials whose measured hardness falls within the target clinical range. Among the synthetic foams, PU Foam Grade 10 returned a mean BMSi of 42 (SD 3.1), and PU Foam Grade 30 a mean BMSi of 61 (SD 4.1) — both classifying as Low, below the normative range for healthy adult bone. PU Foam Grade 40 measured 68 (SD 3.7), placing it in the Mid category. The two highest-grade synthetic materials — PU Foam Solid Rigid and PU Foam Dense Cortical — measured mean BMSi values of 88 (SD 5.2) and 92 (SD 4.8) respectively, both classifying as High and representing the only synthetic candidates confirmed as quantitatively defensible for worst-case testing in the sclerotic bone context.

To support valid cross-category comparison, several methodological controls were applied consistently across all specimen types. Probe tip condition was assessed before each measurement session against defined replacement criteria. Cadaveric specimens were continuously immersed in an aqueous medium throughout testing to preserve tissue-level mechanical properties representative of the in vivo environment; animal specimens were tested within a defined post-harvest interval under consistent storage conditions; and synthetic materials were characterized at ambient temperature and humidity. BMSi distributions were statistically summarized for each specimen type by mean, standard deviation, minimum, and maximum, with inter-specimen variability assessed per material category. Among the animal specimens, porcine femoral cortical bone produced a mean BMSi of 76 (SD 6.1) and bovine tibial cortical bone a mean BMSi of 79 (SD 5.4) — both Mid-category results, consistent with normal adult cortical bone but falling short of the High-category threshold. The cadaveric human sclerotic specimen produced a mean BMSi of 91 (SD 8.5), a High-category result that defines the upper reference boundary of the clinical worst-case range. These results were compared against the published BMSi range for clinically diagnosed sclerotic bone to identify materials falling within or above that threshold.

The study's design was shaped by the practical requirement of the verification program that motivated it: to identify a specific bone model for use in a safety-critical bench test with confidence that the model's measured hardness exceeded that of the worst-case clinical condition. This requirement imposed a clear selection criterion — the material's BMSi must fall at or above the upper boundary of the clinical sclerotic bone range, anchored by the cadaveric sclerotic reference at BMSi 91 — and the characterization study was designed to generate the evidence needed to apply that criterion objectively. Of the eight materials characterized, only PU Foam Solid Rigid (BMSi 88, SD 5.2) and PU Foam Dense Cortical (BMSi 92, SD 4.8) satisfied this criterion, with both falling within the High category and confirming as worst-case surrogates for procedures in which sclerotic bone represents the limiting clinical condition. The methodology is fully replicable: any verification team with access to an OsteoProbe device, the published standard operating procedure, and the clinical reference data can apply the same protocol to any set of candidate bone models and arrive at a quantitatively defensible worst-case determination.

Material Type	Mean BMSi	SD	BMSi Category	Represents Worst Case?
PU Foam Grade 10	42	3.1	Low	No
PU Foam Grade 30	61	4.1	Low	No
PU Foam Grade 40	68	3.7	Mid	No
PU Foam Solid Rigid	88	5.2	High	Yes
PU Foam Dense Cortical	92	4.8	High	Yes
Porcine Femur Cortical	76	6.1	Mid	No
Bovine Tibia Cortical	79	5.4	Mid	No
Human Cadaveric Sclerotic	91	8.5	High	Yes

Table 2: BMSi Values by Bone Surrogate Material Type

7. Results and Implications for Verification Strategy

The quantitative characterization revealed a pattern of systematic inconsistency in the relationship between commercial hardness labels and measured BMSi values across the synthetic foam category. Materials marketed under equivalent hardness descriptors by different manufacturers occupied different positions on the BMSi scale — a gap made concrete by the mislabeling data: a foam labeled "Hard" by Manufacturer A returned a measured BMSi of 61, placing it firmly in the Low category, while a foam carrying the identical label from Manufacturer B measured 88 and correctly fell in the High category. The same divergence appeared in the normal grade: Manufacturer A's "Normal" foam measured 68 (Mid, as expected), while Manufacturer B's "Normal" foam measured only 54 — a Low-category result. The "Cancellous Standard," expected to land in the Mid range, returned a BMSi of 57, again misclassified as Low. Against this, only the Dense Cortical Analog from Manufacturer C (BMSi 92) and Manufacturer B's Hard grade (BMSi 88) carried labels that accurately corresponded to their measured High-category hardness. Of the thirteen synthetic materials tested, nine classified as Low BMSi — below the normative range for healthy adult bone — regardless of their commercial designation. Only two fell in the Mid category and two in the High category, with those High-category materials (BMSi 88 and 92) representing the target zone for worst-case testing in the sclerotic bone context.

This distribution has direct implications for verification strategy. A laboratory selecting a "hard bone" model based on commercial labeling without quantitative verification may inadvertently conduct its worst-case test in a material measuring as low as BMSi 61 — a Low-category value representing bone no harder than mild osteoporosis — and generate verification data that does not capture device performance under the actual worst-case clinical condition. The frequency with which this scenario occurs across the field is unknown, but the mislabeling rate observed here — three of seven labeled grades inaccurately categorized — suggests the risk is not negligible. For devices used in procedures where patients with degenerative spine disease constitute a significant portion of the target population, the clinical consequences of that inadequacy are not abstract.

Animal bone results demonstrated greater alignment between specimen category and measured BMSi than was observed in the synthetic materials, with porcine and bovine specimens generally producing values within the normative human range (BMSi 57–79 across specimen types). However, neither animal model consistently reached the High BMSi category corresponding to sclerotic bone — the porcine femoral cortical specimen measured 76 and the juvenile bovine tibial cortical specimen 79, both Mid-category results — confirming that animal tissue alone is not a reliable surrogate for the worst-case clinical condition in this context. The cadaveric human samples produced

the widest within-category BMSi variation, consistent with known interindividual variability as a function of age, disease history, and anatomical site: osteoporotic cadaveric bone measured 48, normal adult cadaveric bone 73, and sclerotic cadaveric bone 91. Several cadaveric specimens produced BMSi values within the documented sclerotic range, but the ability to identify those specimens prospectively — without IMI measurement — remained limited.
From the perspective of verification strategy, these findings support a two-step selection process: first, characterize all candidate bone models using the IMI protocol to establish their BMSi profiles; second, select the model whose BMSi falls within or above the target clinical range for the worst-case condition. For robotic spine procedures in which sclerotic bone represents the worst case — documented at BMSi 91 in the cadaveric sclerotic specimens — that means selecting a synthetic material from the High BMSi category. The two materials confirmed in that category, at BMSi 88 and 92 respectively, provide the verification program with a validated, quantitatively defensible choice. The methodology used to arrive at that choice generates an auditable evidence trail supporting the worst-case determination in the regulatory submission — a level of rigor that commercial label inference alone cannot provide.

Manufacturer Grade Label	Expected Category	Measured BMSi	Actual Category	Label Accurate?
Hard (Manufacturer A)	High	61	Low	No — mislabeled
Hard (Manufacturer B)	High	88	High	Yes
Normal (Manufacturer A)	Mid	68	Mid	Yes
Normal (Manufacturer B)	Mid	54	Low	No — mislabeled
Soft (Manufacturer A)	Low	42	Low	Yes
Dense Cortical (Mfr. C)	High	92	High	Yes
Cancellous Standard	Mid	57	Low	No — mislabeled

Table 3: Mislabeling Analysis — Commercial Label vs. Measured BMSi Category

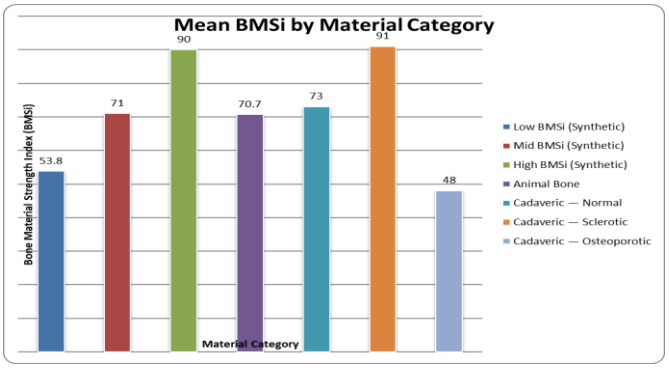

Figure 3: Mean BMSi by Material Category

8. A Reproducible Framework for Industry-Wide Adoption

Although the methods described in this article were developed to overcome this problem (a specific engineering problem), they are not limited to this application. All orthopedic device verification programs that are based on bench testing of test devices in bone surrogate materials (which is nearly all programs, regardless of the device type) face this problem in the sense of: is the test condition representative of the clinical condition?" "The IMI-based characterization framework provides a principled and reproducible answer to that question, and its adoption as a standard part of the bone surrogate selection process would improve the scientific robustness of the verification evidence base in the field [13].

In terms of infrastructure, no additional infrastructure is required other than access to the OsteoProbe device and the standard operating procedure (SOP) as described in the published protocol [9]. The non-destructive nature of the protocol also allows for the performance of characterization measurements on verification test materials without them being consumed or otherwise destroyed. In contrast, BMSi data generated through characterization can be directly compared to published clinical reference range data for a target patient population. This provides a clear and auditable worst-case pathway and a more quantitative justification for worst-case bone model selection for regulatory submission than the qualitative schematic currently utilized for a submission.

One of the best examples of this approach, from both a practical standpoint and the standpoint of methodology, was the collaboration between medical device manufacturers (Medtronic), instrument manufacturers (Active Life Scientific), and researchers with expertise in bone biomechanics (University of Washington), which produced the published characterization study: it fused the engineering verification problem and methodology with the research and measurement tools that had to be brought to bear. Publishing the characterization method in a peer-reviewed journal formalized the method, allowed independent scrutiny, and made it publicly available to other device manufacturers, test laboratories, and regulators for assessment, adoption, and further development in turn [2].
One future step may be a database of standardized BMSi values for common bone surrogate materials, which could be generated by a standards body or consortium of researchers and manufacturers. A database would eliminate the need to measure BMSi each time a new program is implemented, speed up the selection of a worst-case model, and provide a common reference for cross-study comparison. The IMI methodology provides the measurement protocol on which such a database could be built, but the laboratories and device types involved need to devise organizational arrangements to collate, verify and maintain the information [17].

9. Conclusion

Despite decades of orthopedic device verification activity, an important methodological gap exists in the bench testing standard: there is no standardized quantitative process to select bone surrogate test materials to simulate clinical worst-case conditions. The problems of poor worst-case characterization, results that are not comparable between studies, and verification results that are not generalizable to a larger clinical population have all been accepted by the practitioner community. The development and validation of the impact microindentation characterization framework described in this article removes that constraint.

Thus, by taking OsteoProbe measurements in synthetic foams, animal models and cadaveric tissues along one scale and mapping those measurements to peer-reviewed published clinical data for sclerotic bone, device manufacturers have a valid, verified, reproducible, and clinically relevant method for quantitatively characterizing bone substitution models. The finding that 'hard bone' materials from commercial manufacturers occupy a wide, overlapping range of BMSi measurements and that most BMSi measurements fall outside the normative range of measurements in healthy adult bone further demonstrates the limitations of current nomenclature and labeling practices, as well as the utility of quantitative characterization to amend the situation. For safety-critical devices that contact neural tissue, this issue is not just a methodological detail but a contribution to improved patient safety.
Towards this more rigorous standard, verification evidence of worst-case scenarios that regulatory bodies expect is increasing, robotic guidance systems are being applied to spinal surgery and other preparations of bone, and the clinical population that presents for these applications includes many patients with pathological conditions of bone not represented in the test materials. To meet the verification standard that these developments require will require a move away from the subjective descriptions of bone quality that are still widely used today. The clinically relevant, quantitative and calibrated approach described in this article offers exactly that in a form that should be immediately medically useful.

References

[1] Manuela Schoeb et al., "Added value of impact microindentation in the evaluation of bone fragility: a systematic review of the literature," Frontiers in Endocrinology, vol. 11, pp. 1–15, 2020. [Online]. Available: https://doi.org/10.3389/fendo.2020.00015
[2] Lucas R. Budd et al., "Objective selection of bone mimetic materials using impact microindentation," PubMed Central, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12720120/
[3] E. Dall Ara et al., "A nonlinear QCT-based finite element model validation study for the human femur tested in two configurations in vitro," Bone, vol. 52, no. 1, pp. 27–38, 2013. [Online]. Available: https://doi.org/10.1016/j.bone.2012.09.006
[4] S. Shim et al., "Compact bone surgery robot with a high-resolution and high-rigidity remote center of motion mechanism," IEEE Transactions on Biomedical Engineering, vol. 67, no. 9, pp. 2497–2506, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8951212
[5] Manuela Schoeb et al., "Safety outcomes of impact microindentation: a prospective observational study in the Netherlands," Journal of Bone and Mineral Research Plus, vol. 7, no. 10, 2023. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37808395/
[6] Pamela Rufus-Membere et al., "Reference intervals for bone impact microindentation in healthy adults: a multi-centre international study," Journal of Bone and Mineral Research, vol. 38, no. 3, pp. 365–375, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s00223-022-01047-y

[7] Egon Perilli et al., "Failure strength of human vertebrae: prediction using bone mineral density measured by DXA and bone volume by micro-CT," Bone, vol. 51, no. 6, pp. 1170–1176, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S875632821200720X

[8] https://pmc.ncbi.nlm.nih.gov/articles/PMC5152622/p et al., "Assessing underlying bone quality in spine surgery patients: a narrative review of dual-energy X-ray absorptiometry (DXA) and alternatives," The Spine Journal, vol. 20, no. 12, pp. 1927–1936, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1529943020310780

[9] A Diez-Perez et al., "Technical note: recommendations for a standard procedure to assess cortical bone at the tissue level in vivo using impact microindentation," Bone Reports, vol. 5, pp. 287–290, 2016. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5152622/

[10] Krzysztof Zerdzicki et al., "Compressive and tensile properties of polyurethane foam mimicking trabecular tissue in artificial femoral head bones," Frontiers in Bioengineering and Biotechnology, vol. 13, 2025. [Online]. Available: https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2024.1482165/full?utm_source=facebook&utm_medium=cpc&utm_campaign=imp_jnlprom_04-26_fbioe_en_nat_fos-inst__reg13&utm_content=empty&utm_id=120243624876350493_v2_s06_e7606&utm_term=120243624876350493&fbclid=IwY2xjawRfoKlleHRuA2FlbQEwAGFkaWQBqzD4LlFD3WJyaWQRMWlSSG5qMDlKbmYzTGxFaXRzcnRjBmFwcF9pZBAyMjIwMzkxNzg4MjAwODkyAAEeIUNJrVXXXkRdGzBHQqxUxHuEZxS_cTxwp8tCG9c8uqreuWHde4OzLpcRubY_aem_STspA3BpZ6cSZoPsQaXQrQ

[11] James Fletcher et al., "Juvenile bovine bone is an appropriate surrogate for normal and reduced density human bone in biomechanical testing: a validation study," Scientific Reports, vol. 8, no. 1, p. 10181, 2018. [Online]. Available: https://www.researchgate.net/publication/326200599_Juvenile_bovine_bone_is_an_appropriate_surrogate_for_normal_and_reduced_density_human_bone_in_biomechanical_testing_A_validation_study

[12] ASTM International, "F1839-08: Standard specification for rigid polyurethane foam for use as a standard material for testing orthopaedic devices and instruments," ASTM International, West Conshohocken, PA, 2021. [Online]. Available: https://store.astm.org/f1839-08r21.html

[13] Guangming Xia et al., "Pedicle drilling force control of a robotic surgical system via spine-soft tissue coupling model and parameters optimization," Computers in Biology and Medicine, vol. 167, p. 107641, 2023. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2023.107641

[14] Y. Suthari et al., "Cost optimization techniques for efficient resource allocation in cloud computing environments," in Proceedings of the 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), pp. 793–797, 2025. [Online]. Available: https://doi.org/10.1109/IcoICI65217.2025.11253514.

[15] N. Nellutla et al., "Cloud-aware ML engineering: Enabling real-time decision systems via federated DataOps," in Proceedings of the 2025 5th Asian Conference on Innovation in Technology (ASIANCON), pp. 1–9, Aug. 2025. [Online]. Available: https://doi.org/10.1109/ASIANCON66527.2025.11281066.

[16] R. Gollapudi, "Operational drift and risk-bounded decision-making in production database systems," Journal of International Crisis and Risk Communication Research, pp. 132–147, 2023. [Online]. Available: https://doi.org/10.63278/jicrcr.vi.3762.

[17] M. K. Babu and Y. Suthari, "Secure and intelligent PLC systems: Integrating artificial intelligent for enhanced industrial control and data privacy," Computer Fraud & Security, Special Issue, 2024. [Online]. Available: https://doi.org/10.52710/cfs.627.