Investigating the Reliability and Validity of the Portable Osteometric Device

Eric Anderson, Sierra W. Malis, Anna J. Osterholtz, and Molly K. Zuckerman

ABSTRACT

Metric analysis of skeletal material is integral to the analysis and identification of human remains, though one commonly used measuring device, the osteometric board, has lagged in recent advancement. Traditional boards are bulky and require manual measurement recording, potentially generating intra- and interobserver error. To address these limitations, we tested the reliability, validity, and error rates of a novel device, the Portable Osteometric Device Version 1 (PODv1), which measures distance using laser sensors with time-of-flight technology. Forty-five volunteers measured four skeletal elements with the PODv1 and a PaleoTech osteometric board in three rounds. Comparison of tibia, humerus, and femur measurements with both devices showed no significant differences, although the maximum length of the ulna did differ, potentially because of observer confusion regarding the PODv1's user instructions for this element. Our results suggest that the PODv1 is a reliable, valid measurement device compared to traditional osteometric boards. Although both device types can produce calibration, transcription, and observer errors, the time-of-flight technology and the absence of manual recording built into the PODv1 may limit those errors. These advancements and their potential positive impacts on the accuracy of osteometric data collection may have far-reaching benefits for osteological, bioarchaeological, paleopathological, and forensic anthropological data collection.

Keywords: osteometrics, anthropometrics, osteometric board, validity test

El análisis métrico del material esquelético es integral para el análisis e identificación de restos humanos, aunque uno de los dispositivos de medición más comúnmente utilizados, la tabla osteométrica, ha quedado rezagada en los avances recientes. Las tablas tradicionales son voluminosas y requieren la medición manual, lo que puede generar errores intra e inter-observador. Para abordar estas limitaciones, probamos la confiabilidad, validez y tasas de error de un nuevo dispositivo, el Dispositivo Osteométrico Portátil Versión 1 (PODv1), que mide la distancia utilizando sensores láser con tecnología de tiempo de vuelo. Cuarenta y cinco voluntarios midieron cuatro elementos esqueléticos con el PODv1 y una tabla osteométrica PaleoTech en tres rondas. La comparación de las medidas de la tibia, el húmero y el fémur con ambos dispositivos no mostró diferencias significativas, aunque la longitud máxima de la ulna difirió entre ellos, posiblemente debido a la confusión del observador en torno a las instrucciones de uso del PODv1 para este elemento. Los resultados sugieren que el PODv1 es un dispositivo de medición confiable y válido en comparación con las tablas osteométricas tradicionales. Aunque ambos tipos de dispositivos pueden implicar errores de calibración, transcripción y observación, la tecnología de tiempo de vuelo y la ausencia de necesidad de registro manual incorporadas en el PODv1 pueden limitar estos problemas. Estos avances y sus posibles impactos positivos en la precisión de la recopilación de datos osteométricos pueden tener beneficios de largo alcance para la recopilación de datos osteológicos, bioarqueológicos, paleopatológicos y antropológicos forenses.

Palabras clave: osteometría, antropometría, tablero osteométrico, prueba de validez

Since the late nineteenth century, measurements of skeletal material have become an essential part of osteometric research in osteology and other related fields and subfields, including bioarchaeology, paleopathology, and forensic anthropology (Jamison and Zegura 1974; Marks 2012, 2017). Skeletal metrics, including maximum length measurements of elements within the appendicular skeleton such as long bone lengths, are vital components of various methods, including stature estimation, and of research aims, such as the construction of biological profiles and estimations of frailty, that are commonly used in these domains (e.g., Albanese et al. 2016; Marklein et al. 2016; Spradley 2016). Various measurement devices were developed over the ensuing decades to enable the standard collection of many skeletal metrics, with osteometric boards becoming established tools for the collection of maximum length for elements in the appendicular skeleton since their invention in the late nineteenth century (e.g., femoral maximum length; Hepburn 1899; Schiller 1992; White et al. 2011). However, osteometric boards have undergone very little modification or changes to their design since their invention (Hepburn 1899; Schiller 1992). Traditional osteometric

This article has earned badges for transparent research practices: Open Materials. For details see the Data Availability Statement.

Advances in Archaeological Practice 12(2), 2024, pp. 133–142

DOI:10.1017/aap.2023.33

boards tend to be bulky, making them often inconvenient to transport for field-based data collection. In addition, they are not digital, requiring observers to manually read and record the measurements. This can lead to intra- and interobserver error, depending on how accurately observers read the measurement generated.

In this article, we present the results of our testing a new device, the Portable Osteometric Device version 1 (PODv1), for skeletal data collection, specifically the collection of maximum length for elements in the appendicular skeleton. The PODv1 was designed to eliminate the manual data recording required by traditional osteometric boards, as well as their bulkiness, and the consequent potential intra- and interobserver errors in the generated measurements. It measures distance using laser sensors with time-of-flight (ToF) technology and allows real-time digital collection of measurements as they are generated. In its current version, the PODv1 housing is made through 3D printing.

To assess the reliability, validity, and error rates of the PODv1, we compared it to the Paleo-Tech Lightweight Field Osteometric Board (PaleoTech), an industry-leading osteometric board, and assessed intra-observer and interobserver error during skeletal data collection. Forty-five volunteers collected maximum length measurements for four skeletal elements commonly used in osteometric analyses—the tibia, humerus, femur, and ulna—with the PODv1 and the PaleoTech osteometric board in three rounds of data collection. The purpose of this project was twofold: to test the reliability and validity of the PODv1 against the PaleoTech and to determine the reliability and validity of the PODv1 as a device for collecting osteometric data, specifically long bone length. As we demonstrate here, although both device types can generate calibration, transcription, and observer errors, the PODvi's time-of-flight technology and absence of a need for manual recording may limit these issues, especially transcription errors and inter- and intra-observer error. These advancements and their potential positive impacts on the accuracy, standardization, and reliability of osteometric data collection with the novel PODv1 may have far-reaching benefits for osteological, bioarchaeological, paleopathological, and forensic anthropological methods and analyses that involve maximum length measurement of skeletal elements in the appendicular skeleton.

BACKGROUND

Anthropometry, Osteometrics, and Osteometric Boards

Anthropometry, and osteometrics within this domain, centers on methods for measuring the human body, including the skeleton, and has been used since the early eighteenth century for various research purposes in biology, biological anthropology, and related fields (Jamison and Zegura 1974; Marks 2012, 2017). Eighteenth- to mid-twentieth-century anthropometric research in biology and biological anthropology was primarily typological and focused on racial classification and the estimation of intelligence levels and criminal tendencies (Marks 2012, 2017). Although these domains have primarily shifted from classificatory to evolutionary approaches to human phenotypic variation, including metric variation, many biological anthropologists continue to collect and analyze anthropometric data, including osteometrics. Among other applications, skeletal metric data—especially maximum length measurements of elements within the appendicular skeleton (i.e., long bone lengths)—are commonly used to reconstruct maximum living stature, as well as growth and development patterns, and to estimate sex and age-at-death (e.g., Adams and Byrd 2002; DiGangi and Moore 2012; Moore 2012; Moore and Ross 2012). Therefore, osteometric boards are one of the most-used data collection devices in biological anthropology, including osteology, bioarchaeology, paleopathology, and forensic anthropology.

Since their development in the late nineteenth century, traditional osteometric boards have not changed substantially in design and construction (Hepburn 1899; Schiller 1992). Paul Broca (1824–1880), a French physician, anatomist, and anthropologist, developed the first osteometric board in 1888 (Schiller 1992). Hepburn (1899:111) described Broca's original design as "a flat graduated board or plank, at one end of which a flat vertical upright is fixed. Against this upright part the bone to be measured is placed, while to the opposite end of the bone a right-angled triangle of wood is applied, and the length of the bone is read off up on the graduated plank."

Adjustments to this original design have been minimal; traditional osteometric boards consist of a single board base with a fixed panel and detachable sliding panel. An object is placed between these to determine its length. Hepburn (1899) added stabilizing elements, which increased the ease of measurement observations; this adjustment also reduced inter- and intraobserver errors generated using Broca's design (Hepburn 1899).

In the twentieth century, adjustments to the traditional osteometric board were intended to make it more user-friendly, portable, and affordable. For example, some commercially available versions—Carolina Supply Company, Paleo-Tech (now produced by France Casting)—are collapsible, making them more suitable for transport and field-based data collection. Noncommercial, craft-produced versions of traditional osteometric boards also feature limited adjustments. For instance, the Abawerk osteometric board has two upright panels along the length and width of the board with metric-ruled paper placed on the base; the paper is used to generate length measurements for skeletal material placed on the board (Geise 1986). This board design was intended to make measurement generation easier for observers while increasing its accuracy: the ruled paper can be measured from multiple angles, although it is less durable than standard osteometric board materials, such as wood, metal, and plastic (Geise 1986). Indeed, many researchers forego the expense of commercially manufactured osteometric boards—which range from $75 to $1,100, with the majority retailing for more than $200 USD—by making craft versions of osteometric boards, some of which have more durable designs than others (Figure 1). For example, Naples and colleagues (2010) created an inexpensive osteometric board for classroom use, with the goal of facilitating student usage and offering more opportunities for learning forensic methods. Made of rulers, tape, glue, and cardboard, this version of a traditional osteometric board is very affordable but not particularly durable, and the fragility of its components and construction may generate intra- and interobserver error rates that are unsuitable for research use.

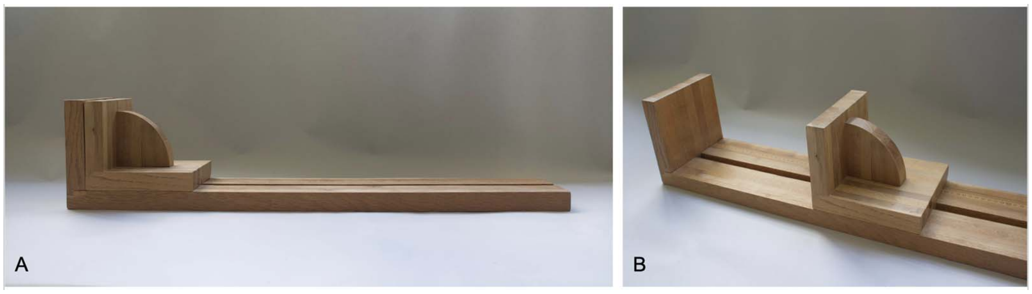

FIGURE 1. Traditional, museum-grade wooden osteometric board (A) with a sliding panel (B).

Metric Considerations

Reliability, validity, and considerations of inter- and intraobserver error are vital to anthropometry and osteometrics. Reliability refers to the measure of inter- and intraobserver error, in which an object is measured on two different occasions by at least two different observers, with little random error (Nance 1987). Validity is the degree to which a measurement consistently accomplishes its intended purpose. In other words, it is a question of how accurately the measurement generated using an osteometric board reflects the true length of the skeletal element, when measuring from a baseline that is accepted as accurate (Nance 1987). Accuracy refers to how close the measurements are to a true value (i.e., the baseline value). This type of information must be collected methodically because of human error—both inter- and intraobserver—instrumentation issues, and the observer's experience level.

Inter- and intraobserver variation is especially critical to consideration of anthropometric, including osteometric, data. Interobserver error results from inconsistent measurements by different observers, whereas intraobserver error results from inconsistent measurements taken by a single observer (Adams and Byrd 2002). Generally, interobserver error is greater than intraobserver error (Langley et al. 2018). Fewer errors also typically occur in the generation of measurements of maximum length and breadth measurements of skeletal material than in other standard skeletal measurements (Langley et al. 2018). Notably, acceptable error rates for anthropometry are <1.5% for intraobserver error and <2% for interobserver error (Perini et al. 2005). Postcranial measurements commonly considered as being "difficult" for most observers to generate, such as those of the tibia (e.g., maximum length vs. medial and lateral condylar length), should have less than a 3% error rate (Adams and Byrd 2002).

Importantly, error rates within skeletal measurements recorded using osteometric boards, whether traditional or novel (e.g., PODv1), can be reduced through observer awareness of external factors such as humidity that can affect measurement precision and accuracy, and thus the reliability and validity of the measurements, along with consequent adjustments in practice: the observer's lack of awareness of these factors leads to measurement errors attributable to the device, observer, or both (Langley et al. 2018). Some factors are external to and independent of the observer, some are observer- and device-specific, and some represent a combination (Albrecht 1983; Geise 1986; Langley et al. 2018). For example, humidity can substantially affect measurements generated using the Abawerk osteometric board. Giese (1986) found that elevated relative humidity of 60% and higher in the local environment in which data recording occurred made the grid paper expand and changed a given measurement's outcome between 0.5 and 1.0 mm. Overall, as ambient humidity increased, so did the length of the measurements (Geise 1986).

Observer-specific factors include the observer's level of experience in collecting skeletal metric data using an osteometric board (traditional or novel), which can influence reliability. Adams and Byrd (2002) assessed reliability relative to observer experience level in postcranial metric data collection. In their study, participants in the 52nd annual American Academy of Forensic Sciences meeting with varying levels of experience (0–1 year, 1.1–5 years, 5.1–10 years, and 10+ years) in recording postcranial skeletal metrics used both digital calipers and an osteometric board to record 22 postcranial measurements. They found that less-experienced observers—with experience of five years or less—generated measurements with higher interobserver error. At the same time, they found that complacency and overconfidence among observers from every experience level also generated imprecision. Commonly observed measurement and data recording errors included transposing numbers, decimal place errors, failure to "zero out" the digital calipers prior to data collection, transcription errors, and confusion surrounding established procedures for generating standard skeletal metrics (e.g., Adams and Byrd 2002; Buikstra and Ubelaker 1994; White et al. 2011).

Reliability and replicability are fundamental to osteometrics (Adams and Byrd 2002; Langley et al. 2018). One way to enhance these two factors is to restrict analyses to osteometric measures that do not require extensive observer experience, such as maximum length measurements (vs. pubis length). These "easier" measurement types have lower error possibilities, helping decrease inter- and intraobserver errors in their collection (Adams and Byrd 2002). However, restricting methods in this fashion can also greatly limit the scope of analyses featuring osteometric data. Another way involves a variety of improvements to the devices used for collecting osteometric data so that they produce more consistent measurements, which can include using computer-assisted methods (Adams and Byrd 2002; Harris and Smith 2009). Our work represents the latter approach.

METHODS AND MATERIALS

Osteometric Laser: Preliminary Study

The PODv1 used in this study was developed from preliminary testing of an initial model that generated laser measurements using a direct ToF method (Anderson and Osterholtz 2021). ToF is a method for measuring distance by determining the time it takes for photons to travel between a sensor and an object (Koerner 2021). Anderson and Osterholtz (2021) assessed whether maximum length measurements of two cast replicas of long bones—femur, radius; Bone Clones, Human Male Skeleton, Disarticulated (SCM-192-D)—generated using a Bosch Blaze GLM 50 C laser measure (Bosch laser) were comparable to those generated with a portable PaleoTech osteometric board. Twenty volunteer observers from the Department of Anthropology and Middle Eastern Cultures at Mississippi State University (MSU), with varying levels of educational attainment in anthropology—for example, a BA degree in progress, MA degree in progress, PhD attained—and levels of experience in collecting osteometric data, collected maximum-length measurements from each element in one round.

A Pearson R test revealed a significant positive correlation between both device measurements of the radius ($r = 0.597$, $n = 20$, $p = 0.005$) and the femur ($r = 0.988$, $n = 20$, $p = 0.04$), indicating that the measurements from the two devices were in general agreement. Even though these results indicate a significant correlation, there was a difference between the radius (moderate, $r = 0.597$) and femur (strong, $r = 0.98$) correlation strength taken from the Bosch laser and the PaleoTech osteometric board. This was likely due to interobserver error associated with different levels of observer experience, which was potentially exacerbated by the absence of any instruction to them on established practices for collecting the measurements of these elements. Additionally, the initial model was not anchored to a table or a board, meaning that it and the Bosch laser had some instability, much like Broca's original device. This likely caused some variation within the measurements compared to the PaleoTech board. Although Anderson and Osterholtz (2021) showed that further refinements were needed, the results did indicate that the initial model and the direct ToF method had potential for generating some types of osteometric data comparable to a traditional osteometric board. Here, we build on this preliminary research. Specifically, testing the PODv1 addresses the technical issues identified in the preliminary research through several design modifications, using volunteer observers with consistent experience levels, and consistently providing standardized, detailed instructions on the correct use of both the traditional osteometric board (i.e., the PaleoTech) and the PODv1 in generating the maximum length measurements for each skeletal element included in the study.

Portable Osteometric Device

The PODv1 was developed to solve some of the reliability and validity issues that researchers observed with osteometric data collection using traditional osteometric boards. The PODv1 is operated by a rechargeable battery, and its results are generated through 3D printing.[1] To measure the length of skeletal elements, it uses a laser-distance–measuring sensor module with direct ToF technology and with a transmitter and a receiver. Whereas traditional osteometric boards use a board to enable measurement of an object's length, in the PODv1 the board is entirely absent. Instead, laser sensors with ToF technology are used to measure distance; that is, the length of an object. The transmitter fires a laser pulse that is reflected off an object (e.g., the sliding panel) and then captured by the receiver to measure photon travel time (Koerner 2021). The laser-distance–measuring sensor used in the PODv1 has an accuracy of ±1 mm, depending on the lighting conditions and distance being measured (AliExpress 2022). Importantly, to achieve an error rate of less than 2%, the laser pulse in ToF distance must reflect from an object with greater than 73% reflectivity (AdaFruit 2016; Jans et al. 2020). Therefore, for increased reflectivity and thus increased reliability, the PODv1 features a white filament target area, giving an 88% reflectivity (AdaFruit 2016).

Compared to traditional osteometric boards (e.g., PaleoTech; see Figure 1), the PODv1 is smaller in size and has a more open design, without an anchoring base board. The open design enables the PODv1 to generate measurements of materials ranging in length from 0.03 m to 2.00 m. This means that it can be used to generate maximum length measurements of skeletal elements spanning the full spectrum of human variation, as well as of other materials such as artifacts that are potentially recovered from field sites where the PODv1 is in use. Additionally, the PODv1's dimensions are approximately 115 × 100 × 58 mm, which makes it more portable and easier to store than traditional osteometric boards, even collapsible versions.

The PODv1 has two primary components and a straightforward setup. The first, a laser panel, is an upright panel containing the laser, the control unit, and the clamp stabilizer. The second component, the sliding panel, has a foldaway sliding stabilizer and stores the parts for the clamp stabilizer (Figure 2). The PODv1 is set up by assembling the clamp stabilizer and attaching it to the laser panel to create a 90-degree angle, which is then placed against a table (or similar flat surface) with a corresponding 90-degree angle. Next, the clamp stabilizer secures the PODv1 to the table with a clamp. After the PODv1 is secured to a table, the sliding panel is assembled and placed against the table. Then, the skeletal material (or other object for measurement) is placed against the laser panel. the sliding panel then secures the material, and the measurement generated is displayed on an LCD screen (Figure 3).

Portable Osteometric Device Features

The PODv1 has four different operating functions: Live Mode, Precision Mode, Calibration, and History. Only the Live and Precision Modes are used to measure the length of an object. The Live measurement is automatically engaged after turning the PODv1 on and displays a continuous measurement. This mode displays an active 25%, midpoint, and 75% of the measured object (see Figure 3A). It also displays the last Precision Mode measurement taken under the Live measurement reading (Figure 3A). The Precision Mode takes 20 measurements and displays their average (Figure 3B). The Calibration function recalibrates the device, returning it to zero. For recalibration to occur, however, a 100 mm block, which is incorporated into the design, must be placed between the two panels during calibration. The History function shows the last five Precision Mode measurements (Figure 3C).

Volunteer Observers

Volunteer observers ($N = 45$) were recruited from MSU ($n = 23$) and from participants in the 2022 American Association of Biological Anthropologist (AABA; $n = 22$) conference.[2] A power analysis, consisting of a paired t-test with an $alpha = 0.05$ and $power = 0.80$, determined the minimum sample size to be 34 observers. This test was performed using GPower (3.1). IRB approval was obtained

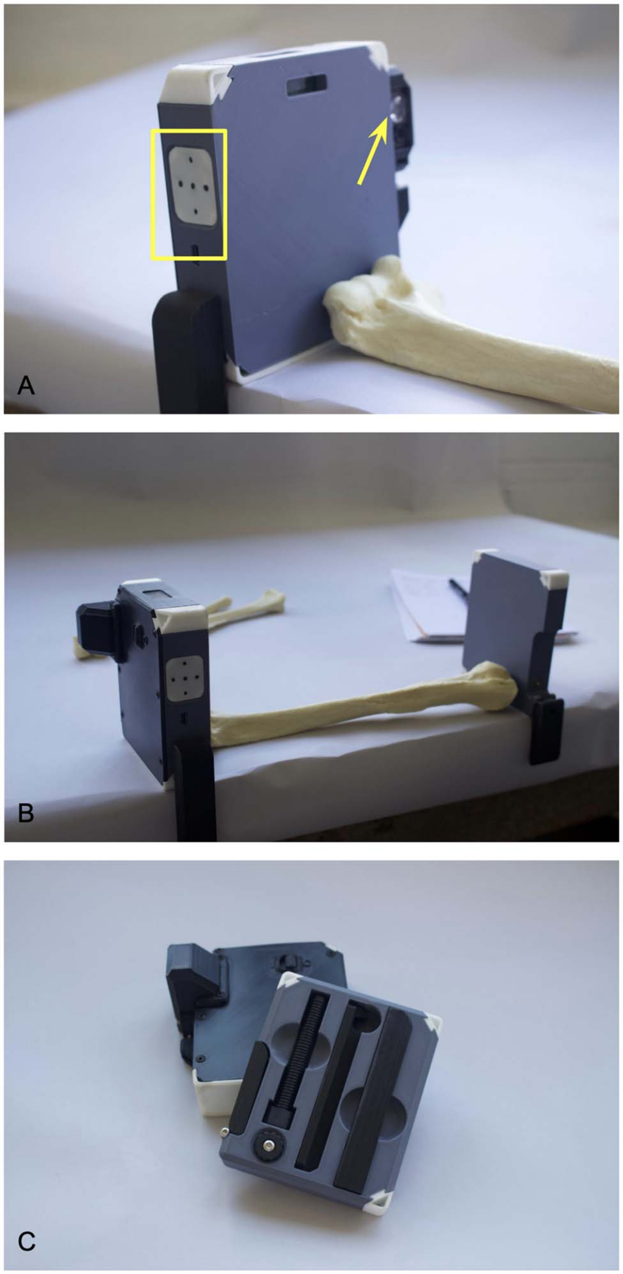

FIGURE 2. PODv1, including (A) the front portion of the Laser Panel with the laser on the right side, the control unit on left side of the Laser Panel (yellow box), the laser (yellow arrow), (B) the Sliding Panel with foldaway Sliding Stabilizer on right side, and (C) the Clamp Stabilizer storage on the back of Sliding Panel.

prior to data collection (MSU IRB-21-457). Volunteers at both MSU and the AABA were required to have completed an undergraduate or graduate-level human osteology course, whether at MSU or a previous degree-granting institution, or possess equivalent experience, such as lab work or taking part in a bioarchaeological field school, prior to data collection. Everyone was given written instructions and watched a six-minute video (made by the first author), both of which provided the same set of detailed instructions on how to use both the PODv1 and the Paleo-Tech Lightweight Field Osteometric Board. Additionally, each volunteer was given written and pictorial instructions for collecting measurements of the four bone elements. These instructions directly followed the procedures described in "Data Collection Procedures for Forensic Skeletal Material 2.0" (Langley et al. 2016). Langley and coworkers' guide was used because it provides clear, detailed, and user-friendly instructions on how to collect the four measurements used in this study.

The first author proctored the volunteer observers at MSU, and the third and fourth authors proctored those at the AABA; the only assistance proctors provided was initially calibrating the PODv1 and assisting any volunteer who had difficulty operating either or both devices. After watching the video and reading the instructions, each volunteer observer independently assembled and operated each device. Each used both devices to measure the four skeletal elements in each of three consecutive rounds. In sum, each observer collected measurements on each element six times for a total of 24 measurements per observer. Volunteer observers entered each measurement into a Microsoft Excel spreadsheet. To limit bias and maintain anonymity, each volunteer recorded their measurements, number of years of experience using traditional osteometric boards, and their primary field of study (e.g., bioarchaeology, forensics). Experience level was categorized into three groups: novice, intermediate, and expert. The novice group incorporated those who had less than 3 years of experience with traditional osteometric boards, members of the intermediate group were those with 3–10 years of experience, and the expert group was made up of those with 10 or more years of experience. No other personal information was recorded (Supplemental Table 1).

Baseline Measurements of Skeletal Elements

We selected the four skeletal elements used in this study—femur, tibia, humerus, and ulna—to represent the spectrum of difficulty involved in collecting osteometric data for each, including maximum length. The tibia is considered to be difficult to measure because of its complex epiphyseal morphology (e.g., intercondylar tubercles; Adams and Byrd 2002), whereas the femur and humerus are not hard to measure, given their simpler epiphyseal morphology. The ulna lies between them; although it is not considered to be difficult to measure (Adams and Byrd 2002), its distinct epiphyseal morphology—olecranon and styloid processes—can create challenges for observers seeking to measure its maximum length.

The first, third, and fourth authors created the baseline skeleton measurement for each of the four bone elements—tibia, femur, humerus, ulna—drawing from two sources. At MSU, the authors generated baseline measurements for, and the volunteer observers collected measurements from, four complete dry bone elements representing four adult individuals from within the Đurđevac-Sošice Commingled Collection,[3] which is curated at MSU. This collection was chosen due to its accessibility, the permission granted by the descendant community for inclusion of the remains within research contexts, and the completeness of the selected skeletal elements. For the AABA volunteers, the authors generated baseline measurements for, and the volunteer observers collected measurements from, four cast skeletal elements drawn from two adult individuals (male, female) in the MSU Biological Anthropology teaching

FIGURE 3. LCD screen on PODv1 including (A) Live Measurement display screen, (B) Precision Mode, and (C) Calibrate and History.

collection: Bone Clones, Human Female Skeleton, Disarticulated (SCM-191-D) and Human Male Skeleton, Disarticulated (SCM-192-D). Both baselines were generated using the same Paleo-Tech Lightweight Field Osteometric Board. Each author collected the measurements from each element in three rounds. All the measurements for each element were then averaged to create a baseline, true measurement for each element, both dry bone and cast. The baseline measurement for each element was then compared to the measurements generated by the volunteer observers at MSU and the AABAs using the PODv1 and the PaleoTech osteometric board, respectively, to test interobserver error and validity. Statistical analyses were completed using RStudio Statistical software.

RESULTS

PODv1 vs. PaleoTech

As mentioned, volunteer observers collected three separate measurements of each of the four skeletal elements using both the PODv1 and PaleoTech devices, which totaled 1,077 measurements: 270 femoral measurements, 269 humeral measurements, 268 tibial measurements, and 270 ulnar measurements. Because the variances of the two independent samples are not equal and they do not have a normal distribution, a Wilcoxon rank sum test with continuity correction was used to compare the maximum length measurements of the four skeletal elements collected by the volunteers using the PODv1 and the PaleoTech board (Wilcoxon 1945). The length measurements of the humerus and tibia generated with both devices showed no statistically significant difference within the subsample of MSU observers, and the femur, humerus, and ulna measurements generated with both devices show no statistically significant differences within the AABA volunteer subsample (Table 1). Measurements of the femur and ulna were significantly different between devices in the MSU subsample, however, as were those of the tibia between the two devices in the AABA subsample. Volunteer observers also varied in their experience level between the two subsamples, with higher experience levels found at the AABA than at MSU (Table 2).

Means for each of the four skeletal elements collected using each device in both subsamples did not reveal either a larger or smaller mean length measurement trend (Table 3). However, standard deviation rates for the length measurements *did* show a trend; within the MSU and the AABA subsamples, measurements generated with the PODv1 produced a larger value than those generated with the PaleoTech board (Table 3).

TABLE 1. Wilcoxon Test Results Comparing Wilcoxon Score (*W*) Derived from the Length Measurements between the PODv1 and the PaleoTech for Each Skeletal Element for Each Subsample.

Element	Subsample	W	p-Value
Femur	MSU	1616.5	0.001*
	AABA	2415.0	0.266
Humerus	MSU	2629.5	0.282
	AABA	2134.0	0.960
Tibia	MSU	2367.0	0.808
	AABA	1660.0	0.016*
Ulna	MSU	3550.5	<0.001*
	AABA	2263.0	0.690

* Statistical significance is $p = 0.05$.

TABLE 2. Experience Levels of Volunteer Observers from Each Sample (*n*).

Experience Level	Subsample	n
Novice	MSU	21
	AABA	4
Intermediate	MSU	6
	AABA	8
Expert	MSU	3
	AABA	10

TABLE 3. Mean and Standard Divisions (*sd*) for Each Element between the Devices.

		PODv1		PaleoTech	
Element	Subsample	Mean (mm)	sd	Mean (mm)	sd
Femur	MSU	412.550	3.5833	411.457	1.2359
	AABA	285.954	2.9866	286.432	1.9412
Humerus	MSU	336.230	3.1580	337.090	1.7380
	AABA	306.762	2.4180	306.926	0.7519
Tibia	MSU	349.540	3.4700	350.216	1.8954
	AABA	345.788	3.1404	344.623	2.4239
Ulna	MSU	260.681	2.4222	262.348	0.9483
	AABA	250.738	2.7742	250.885	1.9279

https://doi.org/10.1017/aap.2023.33 Published online by Cambridge University Press

TABLE 4. Interobserver Percent Error Rate for the MSU Subsample.

Device	Femur	Ulna	Tibia	Humerus
PODv1	0.6300	0.82	0.38	0.33
PaleoTech	0.0067	1.03[a]	0.11[a]	0.07

[a] One outlier was removed from the ulna and two from the tibia maximum length measurements. The original tibia percent error rate was 10.421%; the original ulna percent error rate was 1.0239% before removing outliers.

TABLE 5. Interobserver Percent Error Rate for the AABA Subsample.

Device	Femur	Ulna	Tibia	Humerus
PODv1	0.20	0.81	0.33	0.71
PaleoTech	0.05	0.16	0.00	0.24

Note: One volunteer observer's measurement of the humerus using the PODv1 is missing.

Error rates between the two devices were also similar. Interobserver percent error rates were calculated using the baseline values and the total average of the volunteer observer's measurements for each element collected using the two devices within both subsamples, MSU and AABA (Tables 4 and 5). Results show that both devices generated less than a 2% error rate with each skeletal element within both subsamples. The PaleoTech board generated an overall lower percent error rate than did the PODv1 in each subsample.

Intraobserver error among the measurements generated by the volunteer observers also showed little differences between the devices. Intraobserver error, calculated by averaging the standard deviation rate for the three measurements for each element with both devices across the subsamples, yielded low to no variation across the observer's three rounds of measurements between the devices. However, the PaleoTech board did yield length measurements with lower standard deviation rates.

DISCUSSION

This study tested the reliability and validity of the PODv1 against an industry-leading osteometric board, the Paleo-Tech Lightweight Field Osteometric Board, by providing measurements of interobserver and intraobserver error during the collection of osteometric data by volunteer observers. Overall, results show that the PODv1 is a reliable and valid device for collecting osteometric data, specifically the maximum length of appendicular skeletal elements.

The length measurement results show that the PODv1 and PaleoTech boards generate different levels of accuracy for different elements, however. Although length measurements collected using the PODv1 for the humerus and tibia did not vary substantially from those collected with the PaleoTech board, those of the femur and ulna did. These results are somewhat confusing given Adams and Byrd's (2002) designations of "difficult" for the tibia, "easy" for the humerus and femur, and a mid-range designation for the ulna. However, these results can likely be explained by the volunteer observers' lack of familiarity with the PODv1 compared to a traditional osteometric board, such as the PaleoTech. They may also be attributable to misinterpretation and confusion surrounding the directions for collecting measurements using the PODv1, which is discussed later. Additionally, interobserver error for the length measurements generated using both devices showed that the PODv1 produced a slightly larger standard deviation rate than the PaleoTech board within both the MSU and AABA subsamples; as discussed later, this result suggests the need for modifications to the PODv1's design. Notably, however, use of both devices produced very little interobserver error, with a percent error rate of less than 2% for the elements measured. This indicates that measurements produced using the PODv1, at least in this initial test, fall within the accepted anthropometric standards for measurement errors (Adams and Byrd 2002; Perini et al. 2005). In sum, these results suggest that the PODv1 is a reliable and valid measuring device for collection of some types of osteometric data, with variance in the accuracy of the measurements likely attributable to observer- and device-specific factors, operating both solo and in combination; this study did not test the role of external factors in the PODv1's reliability.

Factors affecting the performance of the PODv1 and, in some cases, the PaleoTech board, included calibration issues, transcription errors, experience level, and observer errors. As observed by proctors, volunteers encountered calibration issues with the PODv1. Specifically, a software glitch caused the PODv1 intermittently to incorrectly calibrate or fall out of calibration during measurement rounds. The incorrect calibration was correctable by conducting a calibration for the device (see the section "Materials and Methods"). In contrast, calibration issues arose with the PaleoTech board when the sliding panel would loosen from the base board, causing the panel to lose its parallel orientation. As a result, different length measurements could be generated from different sides of the sliding panel for a given element; adjusting the bolt that secures the sliding panel corrected this calibration issue. Transcription errors also contributed to some variation in measurements recorded between the two devices. Some errors exceeding 50 mm from the mean were found within data from the MSU and AABA subsamples for both the PaleoTech and the PODv1. These are likely attributable to volunteers either misreading the measurement, whether manually on the board or digitally on the PODv1, or incorrectly entering the reading into Excel. However, these errors were rare; four occurred within the measurements produced with the PaleoTech board and three for the PODv1. The slight difference may be attributed to the PODv1's ability to display a digital length measurement. Additionally, differences in observer experience levels may have also produced variation within their length measurements. Both devices showed a general trend of less deviation as experience level increased. However, measurements generated with the PaleoTech board showed less deviation among each group than those from the PODv1. This effect was likely exacerbated by observer familiarity with traditional osteometric boards and the novelty of and consequent low familiarity with the PODv1.

This lower familiarity is also closely tied to user and volunteer observer error and its role in creating variation within length measurements produced using the two devices. Misunderstanding and confusion as to established practices for collecting maximum

https://doi.org/10.1017/aap.2023.33 Published online by Cambridge University Press

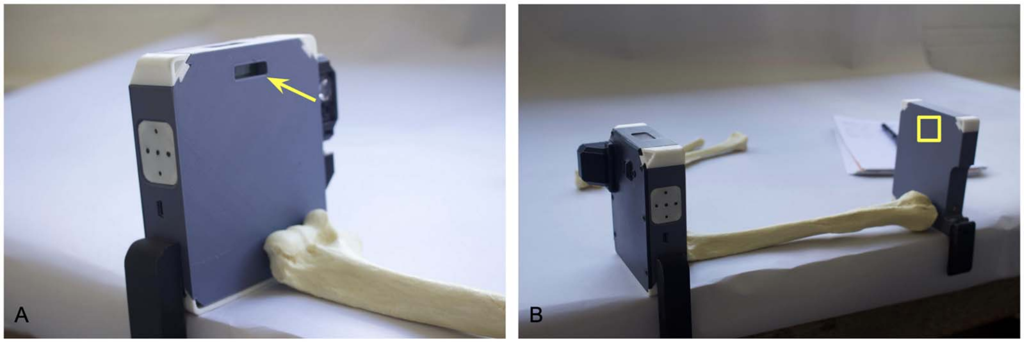

FIGURE 4. (A) The future location of the laser (yellow arrow) and (B) target area (yellow square).

length measurements for the femur, humerus, ulna, and tibia were common among volunteer observers. Proctors observed that volunteers often did not follow the suggested protocol of reading the written instructions and watching the instructional video but instead relied on only one instructional source, combined partial components of both resources, or did not review either one. This may have been due to confusion among the volunteers, misreading or misunderstanding of the instructions, or complacency and overconfidence. User error was notably higher in the MSU subsample, however, suggesting a close linkage with experience. For example, the percent error rate in the MSU subsample for tibia length measurements—the most "difficult" element to measure in this study—with the PaleoTech (1.0259%) and PODv1 (0.8213%) was higher than the other element measurements. In contrast, among the AABA subsample, percent error rates for the tibia were not higher than for the other three elements. Notably, the AABA sample included more expert observers than did the MSU sample, suggesting that error rates and experience are tightly linked, particularly for data collection from more "difficult" skeletal elements.

The PODv1's use of direct ToF methods to collect osteometric data may make it a promising candidate device for anthropometry despite its limitations. Its reliance on battery power, even though the battery is rechargeable, may limit its use in some field data collection settings. Calibration also represents another limitation. If a researcher fails to calibrate the PODv1 before use, inaccurate measurements could result; this limitation does not apply to traditional osteometric boards. Yet, just as with the impact on humidity on some traditional osteometric boards or the loose sliding panel on the Paleo-Tech Lightweight Field Osteometric Board used in this study, observer awareness of external and device-specific factors—and accommodations for them—can reduce consequent error.

CONCLUSION

As we demonstrate here, the PODv1 represents a novel, reliable, and valid alternative to traditional osteometric boards for the collection of osteometric data, especially length measurements of human appendicular skeletal elements. The PODv1 introduced and tested here is in a Beta phase of development, and future versions would benefit from minor adjustments to increase its reliability. These include corrections to the software so that the PODv1 saves and maintains calibration, adjustments of the laser's position to the top of the device, refinements in its alignment, and resolution of the calibration issue that volunteer observers experienced. Proctors and volunteer observers noted that misalignment of the PODv1 was a common issue, causing the long axes of the panels to no longer run parallel to each other. This created variation in the measurement lengths produced for a given element, so that the sides of the sliding panel then produced unreliable measurements. This issue is compounded by the laser's current placement on the side of the PODv1. Necessary changes to a future version of the PODv1 include moving the laser to the top and center of the laser panel and moving the target region to the center of the sliding panel, which will minimize error (Figure 4). It would also be beneficial to have a wider base for both panels, a heavier sliding panel, and visible indicators on a base mat that would make visual alignment easier.

The ToF method has multiple advantages over traditional osteometric boards. In addition to the PODv1's portability, the ToF technology and digital display of data may reduce data collection time while decreasing errors in measurement reading, recording, and transcription. Currently, the POD takes an average of 20 measurements in Precision Mode within a matter of seconds. This feature decreases the time needed to take multiple measurements of a single element, which are required to determine reliability and validity; traditional osteometric boards cannot surpass this capability. For example, if the PODv1 (vs. the PaleoTech board) had been used to create the baseline measurements used in this study, it could have produced three different Precision Mode measurements, equal to 60 different measurements, in one round.

Future iterations of the device will incorporate Bluetooth technology, allowing collected data to automatically transfer to an app called ARC Data.[4] The ARC Data app stores all the collected data for a measurement session; the data are then stored on a device or can upload to a database or secure data storage location later. Although connectivity issues may limit the use of this feature, this adjustment would minimize transcription errors and decrease data entry time.

Overall, the current capabilities of the PODv1 and any additional adjustments made in future versions of the device may have far-reaching benefits for osteometric data collection in human osteology, bioarchaeology, paleopathology, and forensic anthropology.

Acknowledgments

The authors would like to thank the volunteers from MSU and the AABA 2022 conference for their contributions to this research, Shane Miller for statistical guidance, and Arlin Castellanos for help in translating the abstract. No permits were required for this study. IRB approval was obtained prior to data collection (MSU IRB-21-457).

Funding Statement

Funding support was provided by the Cobb Institute of Archaeology and the Mississippi State University Center for Entrepreneurship and Outreach.

Data Availability Statement

All data presented in this study are included in this article and the Supplemental Material.

Competing Interests

Eric Anderson and Sierra Malis are shareholders in Advanced Research Collection Technologies LLC.

Supplemental Material

To view supplemental material for this article, please visit https://doi.org/10.1017/aap.2023.33.

Supplemental Table 1. POD Data.

NOTES

1. 3D printing technology and supporting hardware were used to create the housing unit for the laser because of its low expense and efficiency. The PODv1 housing was designed in Autodesk Fusion 360 CAD software. This CAD software is user-friendly and open source. After the 3D model was rendered in CAD, the PODv1 was placed into the PrusaSlicer slicing software program to be transitioned into a series of layers and a format for printing. This slicing software was used because it was compatible with the Creality Ender-3 V2 FDM 3D printer, a printer with ABS (acrylonitrile butadiene styrene) filament, accessible to the first author. ABS filament was chosen because it is a low-cost, durable material for creating the PODv1 housing.
2. At MSU, volunteers were recruited via an email sent to all BA and MA students who had completed AN 4313/6313 Human Osteology at MSU within the past four years (2017–2021) or an equivalent course at their previous BA degree-granting institution. At the AABA meetings, volunteers were verbally recruited by the third and fourth authors in and around poster and podium sessions focused on human osteology and related topics.
3. The Đurđevac-Sošice Commingled Collection results are from excavations at a medieval Catholic Church site (the Church of St. George) conducted by Osterholtz in collaboration with the Koprivnica Town Museum. The commingled collection consists of the remains of individuals buried within the church that were disturbed during later burial activities. The church was used as a cemetery space between the eleventh and eighteenth centuries. These remains are currently curated at Mississippi State and are used with the permission of the descendant community for both research and teaching purposes.
4. As of December 2022, the ARC Data app was completed with the help of the National Strategic Planning & Analysis Research Center at MSU and is currently being tested.

REFERENCES CITED

AdaFruit. 2016. Adafruit VL53L0X Time of Flight Micro-LIDAR Distance Sensor Breakout. Adafruit Learning System. https://learn.adafruit.com/adafruit-vl53l0x-micro-lidar-distance-sensor-breakout/overview, accessed March 2, 2021.

Adams, Bradley, and John Byrd. 2002. Interobserver Variation of Selected Postcranial Skeletal Measurements. *Journal of Forensic Sciences* 47(6):1193–1202.

Albanese, John., Andrew Tuck, José Gomes, and Hugo Cardoso. 2016. An Alternative Approach for Estimating Stature from Long Bones That Is Not Population-or Group-Specific. *Forensic Science International* 259:59–68.

Albrecht, Gene. 1983. Humidity as a Source of Measurement Error in Osteometrics. *American Journal of Physical Anthropology* 60(4):517–521.

AliExpress 2022. "35.4US $ 9% OFF|50m/164ft Laser Distance Measuring Sensor Range Finder Module Low Cost Diastimeter Single & Continuous Measurement| Distance Measuring Sensor |Measuring Sensormeasure Distance—AliExpress." Aliexpress.com. www.aliexpress.com/item/32823813140.html?src=ibdm_d03p0558e02r02&sk=&aff_platform=&aff_trace_key=&af=&cv=&cn=&dp, accessed January 28, 2022.

Anderson, Eric, and Anna Osterholtz. 2021. The Future Is Coming: Osteometric Laser. Poster presentation at the Mississippi State University Graduate Student Symposium, Mississippi State University, Mississippi State.

Buikstra, Jane, and Douglas Ubelaker. 1994. *Standards for Data Collection from Human Skeletal Remains*. Research Series No. 44. Arkansas Archeological Survey, Fayetteville.

DiGangi, Elizabeth, and Megan Moore. 2012. Introduction to Skeletal Biology. In *Research Methods in Human Skeletal Biology*, edited by Elizabeth DiGangi and Megan Moore, pp. 3–28. Academic Press, Boston.

Geise, Marie. 1986. Technical Report: The Effects of Humidity on the Abawerk Osteometric Board. *American Journal of Physical Anthropology* 71(4):485–486.

Harris, Edward, and Richard Smith. 2009. Accounting for Measurement Error: A Critical but Often Overlooked Process. *Archives of Oral Biology* 54(1):107–117.

Hepburn, David. 1899. A New Osteometric Board. *Journal of Anatomy and Physiology* 34(1):111–112.

Jamison, Paul, and Stephen Zegura. 1974. A Univariate and Multivariate Examination of Measurement Error in Anthropometry. *American Journal of Physical Anthropology* 40(2):197–203.

Jans, Ryan, Adam Green, and Lucas Koerner. 2020. Characterization of a Miniaturized IR Depth Sensor with a Programmable Region-of-Interest That Enables Hazard Mapping Applications. *IEEE Sensors Journal* 20(10):5213–5220.

Koerner, Lucas. 2021. Models of Direct Time-of-Flight Sensor Precision That Enable Optimal Design and Dynamic Configuration. *IEEE Transactions on Instrumentation and Measurement* 70:1–9.

Langley, Natalie, Lee Jantz, Shauna McNulty, Heli Maijanen, Stephen Ousley, and Richard Jantz. 2018. Error Quantification of Osteometric Data in Forensic Anthropology. *Forensic Science International* 287:183–189.

Langley, Natalie, Lee Jantz, Stephen Ousley, Richard Jantz, and George Milner. 2016. *Data Collection Procedures for Forensic Skeletal Material 2.0*. Forensic Anthropology Center. Department of Anthropology, University of Tennessee, Knoxville.

Marklein, Kathryn, Rachael Leahy, and Douglas Crews. 2016. In Sickness and in Death: Assessing Frailty in Human Skeletal Remains. *American Journal of Physical Anthropology* 161(2):208–225.

Marks, Jonathan. 2012. Why Be against Darwin? Creationism, Racism, and the Roots of Anthropology. *American Journal of Physical Anthropology* 149(55):95–104.

Marks, Jonathan. 2017. *Is Science Racist?* Polity Press, Malden, Massachusetts.

Moore, Megan. 2012. Sex Estimation and Assessment. In *Research Methods in Human Skeletal Biology*, ed. Elizabeth DiGangi and Megan Moore, pp. 91–116. Academic Press, Boston.

Moore, Megan, and Anna Ross. 2012. Stature Estimation. In *Research Methods in Human Skeletal Biology*, ed. Elizabeth DiGangi and Megan Moore, pp. 151–179. Academic Press, Boston,.

Nance, Jack. 1987. Reliability, Validity, and Quantitative Methods in Archaeology. In *Qauntitative Research in Archaeology: Progress and Prospects*, edited by Mark Aldenderfer, pp. 245–293. SAGE Publications, Newbury Park, California.

Naples, Virginia, David Breed, and Jon Miller. 2010. A Skeleton Tells Its Own Story: Forensic Analyses of Skeletal Elements for the Science Classroom Laboratory. *American Biology Teacher* 72(3):162–711.

Perini, Talita, Glauber Lamiera de Oliveira, Juliana dos Santos Ornellas, and Fátima Palha de Oliveira. 2005. Technical Error of Measurement in Anthropometry. *Revista Brasileira de Medicina do Esporte* 11(1):81–90.

Schiller, Francis. 1992. *Paul Broca, Founder of French Anthropology, Explorer of the Brain*. Oxford University Press, Oxford.

Spradley, Kate. 2016. Metric Methods for the Biological Profile in Forensic Anthropology: Sex, Ancestry, and Stature. *Academic Forensic Pathology* 6(3):391–399.

White, Tim, Michael Black, and Pieter Folkens. 2011. *Human Osteology*. 3rd ed. Academic Press, New York.

Wilcoxon, Frank. 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1(6):80–88.

AUTHOR INFORMATION

Eric Anderson ■ Cultural Surveys Hawaii, Kailua, HI, USA (ehanderson2034@gmail.com, corresponding author)

Sierra W. Malis ■ Department of Anthropology and Middle Eastern Cultures, Mississippi State University, Mississippi State, MS, USA; National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

Anna J. Osterholtz and **Molly K. Zuckerman** ■ Department of Anthropology and Middle Eastern Cultures, Mississippi State University, Mississippi State, MS, USA

https://doi.org/10.1017/aap.2023.33 Published online by Cambridge University Press