



Accuracy and Reproducibility of Linear and Angular Measurements in Virtual Reality: a Validation Study

Asif Ahmed Anik¹ · Brian A. Xavier² · Jan Hansmann² · Emmanuel Ansong³ · Jinsong Chen⁴ · Linping Zhao⁵ · Edward Michals²

Published online: 8 August 2019

© Society for Imaging Informatics in Medicine 2019

Abstract

The purpose of this experimental study is to validate linear and angular measurements acquired in a virtual reality (VR) environment via a comparison with the physical measurements. The hypotheses tested are as follows: VR linear and angular measurements (1) are equivalent to the corresponding physical measurements and (2) achieve a high degree of reproducibility. Both virtual and physical measurements were performed by two raters in four different sessions. A total of 40 linear and 15 angular measurements were acquired from three physical objects (an L-block, a hand model, and a dry skull) via the use of fiducial markers on selected locations. After both intra- and inter-rater reliability were evaluated using inter-class coefficient (ICC), equivalence between virtual and physical measurements was analyzed via paired *t* test and Bland-Altman plots. The accuracy of the virtual measurements was further estimated using two one-sided tests (TOST) procedure. The reproducibility of virtual measurements was evaluated via ICC as well as the repeatability coefficient. Virtual reality measurements were equivalent to physical measurements as evidenced by a paired *t* test with *p* values of 0.413 for linear and 0.533 for angular measurements and Bland-Altman plots in all three objects. The accuracy of virtual measurements was estimated to be 0.5 mm for linear and 0.7° for angular measurements, respectively. Reproducibility in VR measurements was high as evidenced by ICC of 1.00 for linear and 0.99 for angular measurements, respectively. Both linear and angular measurements in the VR environment are equivalent to the physical measurements with high accuracy and reproducibility.

Highlights

- First validation study on the linear and angular measurements in a state-of-the-art VR system
- Both linear and angular measurements in the VR environment studied are statistically equivalent to physical measurements
- Reproducibility of both linear and angular measurements in the studied VR environment is high
- The accuracy of both linear and angular measurements in the studied VR environment is estimated.

✉ Linping Zhao
lpzhao99@uic.edu; lping.zhao@gmail.com

¹ Department of Mechanical Industrial Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607, USA

² Department of Radiology, University of Illinois College of Medicine, 1740 W. Taylor St., Chicago, IL 60612, USA

³ University of Illinois College of Medicine, 1853 West Polk Street, Chicago, IL 60612, USA

⁴ College of Applied Health Sciences, University of Illinois at Chicago, 1919 W. Taylor St., Chicago, IL 60612, USA

⁵ The Craniofacial Center, University of Illinois at Chicago, 811 S. Paulina St., Chicago IL60612, USA

Keywords Validation · Virtual reality · Measurements · Accuracy · Reproducibility · Equivalence

Introduction

Recent breakthroughs in head-mounted display (HMD), positional tracking system, and software platform including game engines, etc., have been shifting the paradigm of virtual reality (VR) technology [1]. The new era, characterized with much simplified hardware with higher capacity and considerable reduced cost and with a software platform where integrating the multi-media system and developing user-specific applications are easy, provides a much wider accessibility to the virtual reality technology [2]. Its impact has been spreading beyond the game and movie industry by bringing a new wave of applications in medicine, particularly, in radiology [3–5].

The latest virtual reality (VR) environments, equipped with a head-mounted display and a set of controllers, can provide a fully immersive interactive view of three-dimensional computed tomography (CT) and computed tomography angiography (CTA) scans which allow clinicians to explore 3D patient-

specific anatomy with higher degree of versatility and flexibility [6, 7]. The advanced visualization in the immersive view creates a high-fidelity virtual experience in which the user can interact with multi-dimensional objects and perceive relationships at their true scale in the same space. The VR nature helps users to identify critical features by applying scientific judgment, and instant manual measurement can be carried out on the identified area of interest [8, 9]. The physical extent, alignment, separation, and other dimensional properties become much more perceptible as the user maneuvers through the virtual space in the immersive environment [6, 10].

One such development is the ImmersiveTouch platform and application software (ImmersiveTouch® Inc., Chicago, IL) towards various clinical applications in medical and dental sectors. Among them, ImmersiveView™ (ImmersiveTouch® Inc., Chicago, IL), as shown in Fig. 1a, is a platform based on the virtual reality environment developed as a logical next step in the visualization, interpretation, and measurement of 3D medical images including CT, CTA, and CBCT. Within this VR environment, 3D representation of clinical imaging data can be quickly loaded and easily manipulated in real time, facilitating easier visualization of each patient's unique anatomy. The user can view the anatomical image from any given position, change the orientation and scale easily, and manipulate the elements of the virtual with various interaction techniques. The potential benefit of this is myriad in terms of diagnosis, surgical and treatment planning, and outcome evaluation. From a clinical application standpoint, however, it is essential and critical to validate the ability of such specific designed VR system to accurately represent and measure the patient-specific imaging data.

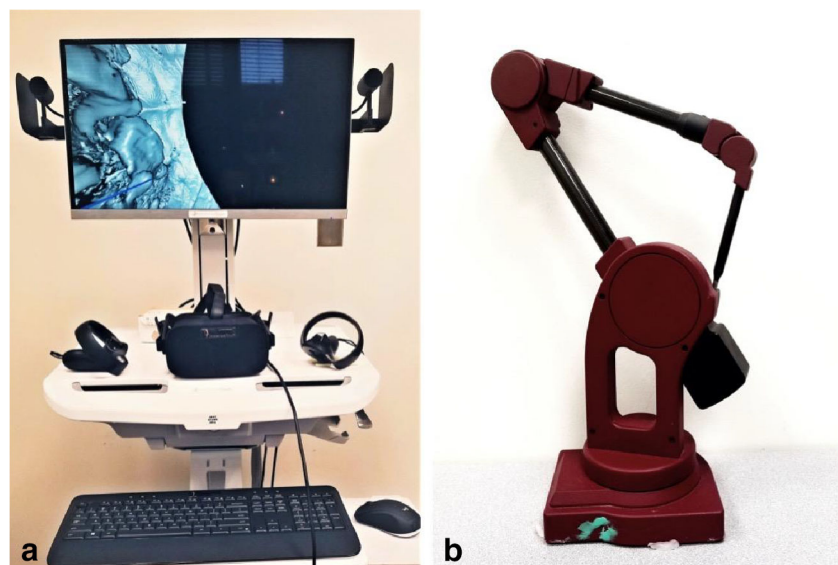
Accurate and precise anatomical measurements are necessary in multiple fields of medicine whether for diagnosis, evaluation of growth, or subsequent intervention [11]. The

platform, ImmersiveView™, has a built-in linear and angular measurement tool for quantified analysis, which is essential for diagnosis and further treatment planning. However, the accuracy and reproducibility of such virtual measurements have not been validated so far to the authors' knowledge. The purpose of this experimental study is to validate linear and angular measurements made in the VR environment using the ImmersiveView™ software in comparison with the physical measurements in a laboratory setting. This validation, focusing on the basic concepts and considering reasonable clinical relevant factors, intends to provide a benchmark reference for further clinical validation in terms of equivalence, accuracy, and reproducibility. The hypotheses to be tested are that (1) the virtual linear and angular measurements are equivalent to the physical measurements in terms of accuracy and (2) the reproducibility of virtual measurement is statistically high.

Methodology

In order to compare the virtual measurements with the physical measurements and evaluate the accuracy of the virtual measurements, the experiment is designed into two separate parts. In the first step, the accuracy of the physical measurements was established via a systematic calibration of the physical measurement instrument utilized in this study to determine experimental precision, resolution, and workspace. In the next step, both virtual and physical measurements on the same parameters were acquired and compared statistically in terms of reproducibility, equivalence, and accuracy. In addition, both inter- and intra-rater reliability were evaluated. The details of these steps are discussed in the following sections.

Fig. 1 **a** ImmersiveView™ instrument is used to acquire VR measurements for linear and angular measurements from three different objects. **b** MicroScribe CMM is used to acquire physical measurements of the same models



Calibration of the MicroScribe CMM

The MicroScribe 3Dx Digitizer (Immersion Corp., CA, USA), a coordinate measuring machine (CMM), as shown in Fig. 1b, was utilized to carry out direct physical measurements. This instrument was calibrated and evaluated in two aspects: the resolution and the consistency in the measurement space.

Firstly, the digitizer was calibrated using a set of grade 1, workshop standard gauge blocks, which has a tolerance of $+0.10$ to -0.05 μm (ASME B89.1.9M). Five-gauge blocks of 10 mm, 30 mm, 50 mm, 80 mm, and 100 mm in length were selected and measured using the MicroScribe 3Dx Digitizer from vertex to vertex along the longitudinal direction. Their spatial coordinates were recorded and used to calculate distances, which were then compared with the nominal lengths. Measurements were repeated five times, and the mean differences and their standard errors were 0.04 ± 0.03 mm, 0.05 ± 0.04 mm, -0.03 ± 0.03 mm, -0.03 ± 0.04 mm, and -0.05 ± 0.04 mm, respectively. The resolution of the digitizer was defined as the maximum absolute value of these mean differences and standard errors. Therefore, 0.09 mm was used as the resolution of the linear measurements from the MicroScribe CMM.

Secondly, the consistency of the linear measurements in the designated workspace, corresponding to the size of the human skull, was evaluated. A calibration cylinder was introduced, which was originally designated for the ICAT FLX v17 Cone Beam CT scanner (Imaging Sciences International, Hartfield, PA) to provide imaging calibration with a spatial resolution of 0.125 mm [12]. This cylinder measuring 116 mm in diameter and 180 mm in height was embedded with a set of radiopaque metal beads of 0.6 mm in diameter on its cylindrical surface. Eight columns of beads were evenly distributed circumferentially with four beads in each column longitudinally 40 mm apart with a tolerance of ± 0.25 mm. Four pairs of beads at representative locations, upper, lower, left, and right, were selected to represent the entire workspace. The spatial coordinates of each selected bead were measured, and the corresponding distances were calculated. After repeating it five times, the mean and standard error were computed and then compared with the range of $40 \text{ mm} \pm 0.25$ mm. It yielded that all mean and standard errors from four locations were within this range. This suggested that measurements at different locations in the workspace were consistent and within the tolerance as defined by this calibration cylinder.

Materials and Measurements

Three objects concurrent to the typical scenario in clinical applications were chosen and prepared for this investigation: an L-shaped block, a dry human hand model, and a dry human skull model. The L-shaped block consisted of six small rectangular blocks on the top surface. Fiducial markers of 2-mm

diameter were placed on top of each block, resulting in a total of 6 points used in the measurements. Ten linear distances between these fiducial markers were predefined to measure. Similarly, fiducial markers were placed at 6 different locations on the human hand model, and 10 distances and 7 angles were predefined. For the dry skull, the third object, a set of anatomical landmarks were selected that were easily identified externally while allowing for a wide range of angles and linear distances. Fiducial markers were placed at 10 of these landmarks, and 20 linear distances and 8 angles were predefined. All the objects involved and their corresponding landmarks are shown in Fig. 2. This is comprised of a set of measurements with a total of 40 linear distances and 15 angular measurements.

The CT images were acquired using the GE Revolution Evo (GE Healthcare, Milwaukee, WI, USA) following an established clinical protocol as follows: 120 kVp, 250 mA, and a rotation time of 0.5 s with an in-plane pixel size of 0.325×0.325 mm and slice thickness of 0.625 mm, and exported as DICOM (Digital Imaging and Communication in Medicine) files. The DICOM files were processed by the ImmersiveView™ (ImmersiveTouch Inc., Chicago, IL, USA) software, which converted the scan data into surface-rendered 3D models. The models were visualized and interacted within the ImmersiveView™ VR environment with the use of an Oculus Rift HMD (Oculus VR, Menlo Park, CA, USA) and a pair of controllers. The virtual measurements were acquired on the 3D models using ImmersiveView™ built-in measurement tools.

Measurement Protocol

The raters were instructed to measure from the same location on each fiducial marker, using either the CMM stylus or the ImmersiveView™ measurement tools. Practice and instruction sessions were carried out on these three objects prior to the measurement sessions.

Both raters first performed measurements with the calibrated MicroScribe CMM for direct physical measurements followed by measurements using the ImmersiveView™ measurement tool in the VR environment. With a short break between sessions, both physical and VR measurements were then repeated four more times. The data from the first session was considered a learning step and was discarded, while data collected from the other four sessions were used for further analysis.

In each physical measurement session, the coordinates of each fiducial marker were acquired and recorded following a standardized, predetermined sequence. The linear distances and angles between fiducial markers were calculated afterwards.

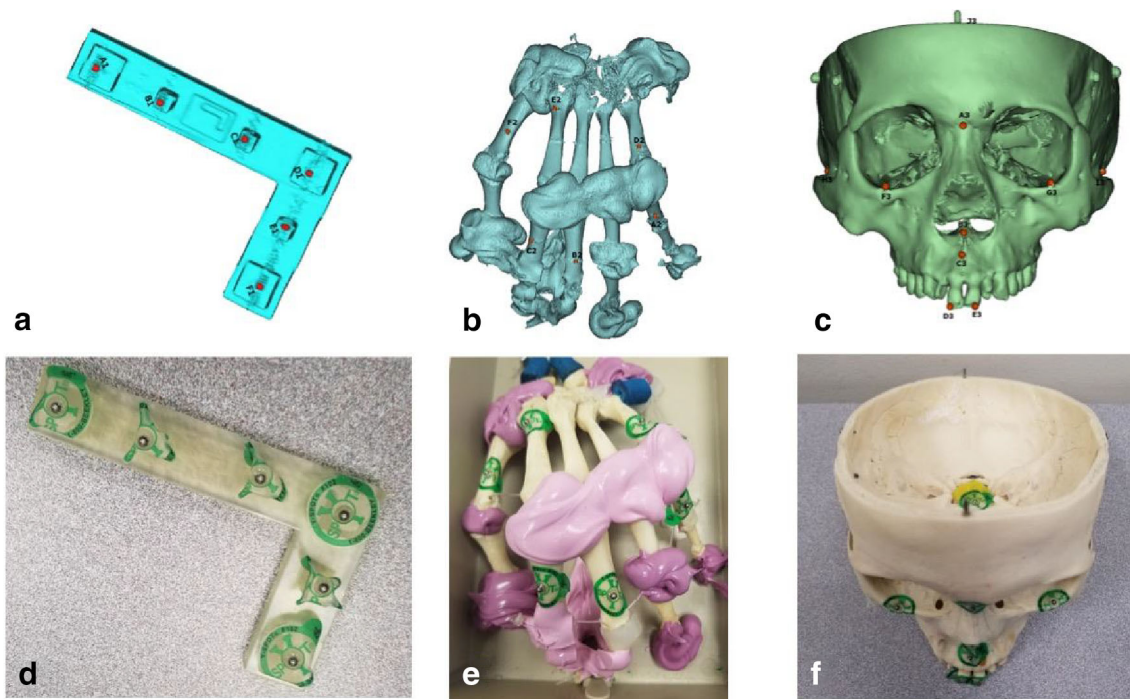
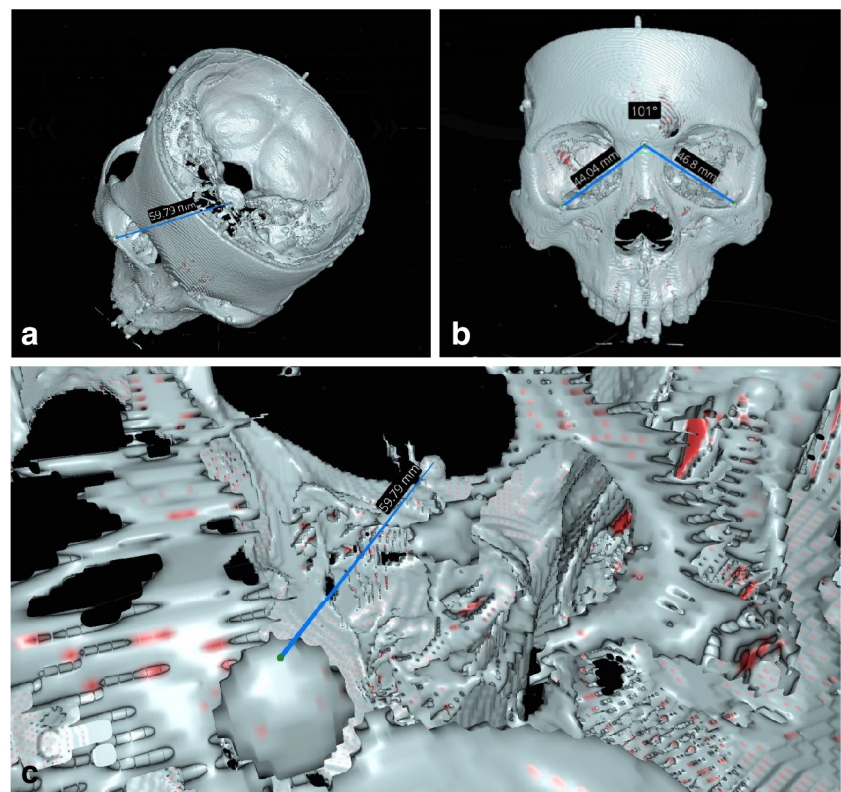


Fig. 2 Representation of the 3D model with fiducial marker (red dot) landmarks attached to the **a** L-block, **b** hand, and **c** skull and the corresponding actual physical model with a fiducial marker (silver dot) attached to the **d** L-block, **e** hand, and **f** skull

Since the virtual models could be rotated and magnified, raters were encouraged to manipulate the 3D model and adjust the measurement marker placements during the recordings. Once the rater was satisfied, the

readings of a distance or an angle were then recorded. An example of high magnification utilized in the virtual environment along with linear and angular measurements is shown in Fig. 3.

Fig. 3 **a** Linear measurement in VR. **b** Angular measurement in VR. **c** High-magnification state in the VR environment



Statistical Analysis

Intra- and Inter-rater Reliability

Two raters performed all direct physical measurements and VR measurements. This was repeated four times (or four trials). The intra-rater reliability in both direct physical and VR measurements was evaluated using the inter-class coefficient (ICC) for both raters, individually. The inter-rater reliability between the two raters in both physical and VR measurements was evaluated using ICC based upon the mean of all the trials. All ICC analyses were performed with the two-way random effects model, a single rater, and absolute agreement.

Analysis of Reproducibility

Reproducibility of VR measurements is the degree of agreement among all four trials for all measurements of the two raters. This was analyzed using ICC with the two-way random effects model, a single rater, and absolute agreement. Furthermore, reproducibility was estimated using the repeatability coefficient, following a procedure utilized by Bland and Altman [13]. This procedure involves calculating the within-subject standard deviation (S_w) from the square root of the residual mean square in one-way ANOVA, taking measurements as the factor. The repeatability coefficient is defined as the $1.96\sqrt{2} S_w$, or $2.77 S_w$ for 95% measurements [5]. As a comparison, the repeatability coefficient of physical measurements was also calculated.

Analysis of Equivalence Between VR and Physical Measurements

The two-tailed paired t test was performed for all VR and physical measurements to test the null hypothesis, which can be interpreted as a mean difference of both types of measurements equal to zero [14]. If the null hypothesis is rejected ($p < 0.05$), then the equivalence between MicroScribe CMM and VR measurements will not hold. If the p value is sustained ($p > 0.05$), then the equivalence of the two methods is likely.

The equivalence between these two measurement methods was evaluated using Bland-Altman plots, which test the level of agreement between direct physical and VR measurements by calculating the mean difference (bias) and the 95% limit of agreement [13].

To further evaluate the equivalence, we attempted to identify whether the VR measurements, compared with the corresponding physical measurements, were within a strict predefined equivalence margin $[-\theta, \theta]$. The two one-sided tests (TOST) procedure was used for all linear and angular VR measurements with $\alpha = 0.05$, a $(1 - 2\alpha)$ confidence interval, and $\theta = 0.5$ mm for linear measurements and $\theta = 0.7^\circ$ for

angular measurements [15]. The predefined level of acceptance was chosen based on the previously found resolution of the MicroScribe CMM in the calibration stage. In the TOST procedure, the null hypotheses stated that the true value and measured value were not equivalent. And the alternative hypothesis is an effect that falls within the equivalence bounds or the absence of an effect that is worthwhile to examine. Therefore, if the confidence interval for the mean difference between the two methods is completely contained within the equivalence margin, the null hypothesis is rejected, and these two methods are considered equivalent ($p < 0.05$) within the equivalence margin. Such equivalence margin, in turn, provides an estimate to the accuracy of the VR measurements, given that the accuracy of the physical measurements is known.

All tests were performed with Statistica software 6.0 (Statistica for Windows; Statsoft, Tulsa, OK, USA) and IBM SPSS Statistics 22 (IBM SPSS, SPSS Inc., Chicago, IL, USA).

Result

Intra- and Inter-rater Reliability of VR and Physical Measurements

Both raters yielded high degree of intra-rater reliability in both VR measurements and physical measurements as evidenced by ICC from 0.99 to 1.00, as shown in Table 1. Inter-rater reliability was also high in both VR and physical measurements as evidenced by the ICC of 1.00 for linear measurements and 0.99 for angular measurements. Such consistency among raters provided a necessary basis for further analysis in reproducibility, equivalence, and accuracy.

Reproducibility of VR Measurements

The ICCs among four trials of VR measurements from the two raters together were 1.00 for linear and 0.99 for angular measurements, indicating a high degree of reproducibility. Such high degree of reproducibility is comparable with that of the physical measurements using CMM: the ICCs among four trials of physical measurements were 1.00 for linear and 1.00 for angular measurements.

The repeatability coefficients of VR measurements were 1.74mm for linear distance measurements and 2.86° for angular measurements, compare to that of physical measurements of 1.11mm and 1.65° , respectively (Table 2). Considering the size of fiducial markers to be 2 mm in diameter, these repeatability coefficients also indicated a high degree of repeatability and reproducibility. Such high repeatability provides a

Table 1 Inter-rater and intra-rater reliability for 40 distance and 15 angular measurements from two raters. A value of above 0.9 indicates strong reliability between the measurements

Method	Types	Intra-rater reliability		Inter-rater reliability
		Rater AA (4 trials)	Rater BX (4 trials)	Rater AA vs rater BX (mean of 4 trials)
ImmersiveView™	Linear	1.00	1.00	1.00
	Angular	0.99	0.99	0.99
MicroScribe CMM	Linear	1.00	1.00	1.00
	Angular	1.00	1.00	0.99

solid foundation for further evaluation of the agreement between measurements using the MicroScribe CMM and the ImmersiveView™ using the Bland-Altman plot (Fig. 4).

Equivalence Between VR and Physical Measurements

The two-tailed paired *t* test yielded a *p* value of 0.513 for linear measurements and a *p* value of 0.433 for angular measurements (Table 3), which is larger than the predefined confidence level of 0.05. Thus, the null hypothesis cannot be rejected or the mean difference between the two methods equal to zero cannot be rule out. This suggests that the equivalence between VR measurements and physical measurements is likely.

The Bland-Altman plots in Fig. 4 revealed good agreement and no systematic bias between VR and physical measurements. For linear measurements as shown in Fig. 4a, the 95% limit of agreement (−1.21, 1.27) contains 93.8% (75/80) scores, while the mean difference between VR and physical measurements is 0.03 mm with a 95% confidence interval (−0.17, 0.11). For angular measurements, as shown in Fig. 4b, the 95% limit of agreement (−2.47, 2.47) contains 93.3% (28/30) scores, while the mean difference between VR and physical measurements is 0.001° with a 95% confidence interval (−0.47, 0.47).

Further, the TOST procedure yielded equivalence between VR and physical measurements in all measurements (all *p* < 0.001) from three objects (L-block, hand, skull models) at the predefined equivalence

Table 2 Reproducibility quantified by the inter-class coefficient (ICC) and the repeatability coefficient [2] of VR and physical measurements of 40 linear measurements and 15 angular measurements

Rater	Types	ICC		Repeatability coefficient	
		VR	CMM	VR	CMM
2 raters (AA and BX)	Linear	1.00	1.00	1.74 mm	1.11 mm
	Angular	0.99	1.00	2.86°	1.65°

VR VR measurements via ImmersiveView™, CMM physical measurements via the MicroScribe CMM

margins of ±0.5 mm for distance measurements and ±0.7° for angular measurements, respectively, as shown in Table 4. Considering the accuracy of the physical measurements via MicroScribe CMM to be 0.09 mm, which is much smaller, one can use these equivalence margins to estimate the accuracy of the VR measurements using ImmersiveView™ software with confidence.

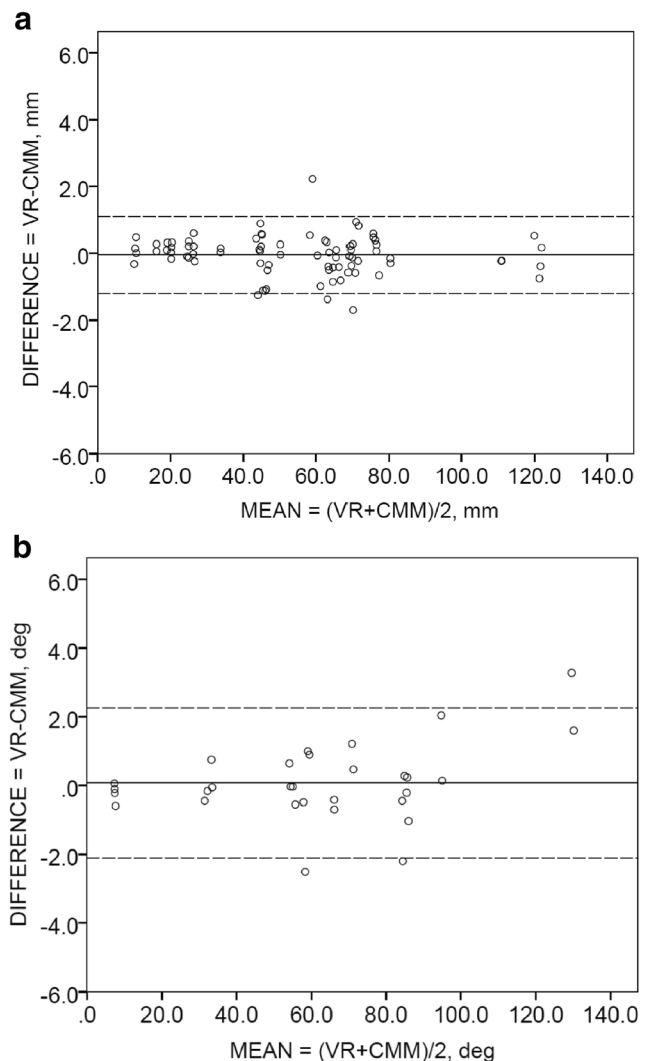


Fig. 4 Bland-Altman plots for the comparison of **a** linear measurements and **b** angular measurements from both raters (AA and BX)

Table 3 Paired *t* test result for 40 linear distance and 15 angular measurements for all three objects obtained in both VR environment and direct physical measurements. The null hypothesis of the paired *t* test states that the mean differences between both the measurements should be equal to zero

Types	Mean difference	Standard deviation	Standard error	95% confidence interval for mean difference		<i>t</i> statistics	<i>p</i> value
				Lower	Upper		
Linear	0.063	0.615	0.072	−0.047	0.080	0.505	0.513
Angular	0.085	1.264	0.108	−0.129	0.300	0.785	0.433

Discussion

The primary purpose of this study was to validate the VR measurements in a specific VR environment before moving it into a real-world clinical application. As the initial step, this study is designed to identify the difference, if any, among VR measurements and physical measurements under a controlled condition with selected objects, fiducial landmarks, and experienced raters (AA and BX). With a broad range of measurements (linear and angular) and stringent experimentation, we hope this primary study will act as a benchmark reference for further experimental and clinical validations.

Our results revealed high accuracy and reproducibility of VR measurements that is equivalent to the corresponding physical measurements, which validates that accurate and precise geometric measurements are feasible in a virtually simulated environment. There was a high level of agreement for repeated measurements between both raters. Independent of the objects used, all linear and angular measurements showed high concordance with the physical measurements. This suggests that these measurements carried out in a VR environment can be used to analyze and accurately measure complex anatomical structures and with agreeable accuracy and precision.

Numerous methods have been applied previously to measure physical objects in the virtual environment [16–18]. Brady et al. developed a volume rendering technique in a virtual platform called “Crumbs” for the visualization of volumetric data and fiber tracking [17]. Bethel et al. developed a virtual measurement method to obtain point measurement in 3D structures from stereo image pairs generated by a scanning

electrode microscope (SEM) [16]. In their model, the user can perform distance and angular measurements using virtual sensors such as a virtual protector and caliper between features present in the stereo image pairs. Hagedorn et al. created an immersive visualization environment by using three-screen visual display to measure quantitative analysis of the tissue engineering aspects [19]. Preim et al. described a 3D interactive quantitative analysis method in a virtual environment to measure spatial resolution [20]. Similarly, Reitingner et al. presented a set of virtually enhanced tools to measure distance, angles, and volumes for different anatomical structures in a conference [21]. They argued that their virtual method can augment more natural interactions, and the user can perform measurement more effectively compared with the non-interactive two-dimensional system. However, these aforementioned virtual reality-based measurement tools have not been properly validated with rigorous evaluation.

To the author’s knowledge, this is the first study to validate the VR measurement tool via comparing virtually simulated linear and angular data with physical measurements corroborated with detailed statistical analysis. This study suggests the reproducibility of the measurements made in the VR environment with ImmersiveView™ is high as evidenced by the ICC (0.99). Such high reproducibility is comparable with that of the direct physical measurements made with the MicroScribe CMM. In addition, the repeatability coefficient was also included in this study. With the same units of the measures, it can provide an estimate to the repeatability in a way that is easy to associate with the practice and applications. The slightly higher repeatability coefficients of VR measurements (1.74 mm for linear distance and 2.86° for angular

Table 4 Equivalence between physical measurements using the MicroScribe CMM and VR measurements using ImmersiveView™ established via two one-sided tests (TOST) with equivalence margins of ± 0.5 mm for linear measurements and ± 0.7° for angular measurements in three objects from both raters combined. *p* values < 0.05 indicate the null hypotheses were rejected and the equivalence was established

Object	Type	Mean difference (CMM – VR)	95% confidence interval for mean difference		<i>p</i> value
			Lower	Upper	
L-block	Linear	−0.02	−0.16	0.11	<0.001
Hand	Linear	−0.03	−0.30	0.24	0.038
Hand	Angular	−0.06	−0.62	0.49	0.031
Skull	Linear	−0.09	−0.25	0.08	<0.001
Skull	Angular	0.01	−0.43	0.44	0.007

measurements) than those of physical measurements (1.11 mm and 1.65°, respectively) (Table 2) indicated that physical measurements have a higher degree of reproducibility than VR measurements. Such difference may be related to the voxel size of the CT scan images, which has a slice thickness of 0.625 mm, compared with 0.63 mm (1.74 mm, –1.11 mm). It suggests that the voxel size of the CT scans is still a major factor that affects the accuracy and reproducibility of the VR measurements. Even so, the low variation among repeated measures reflects high internal consistency and can be considered clinically acceptable [22].

In this study, we choose to compare VR measurements with the physical measurements. This approach was chosen to answer the following questions: (1) what is the accuracy of the physical measurements? (2) Are VR measurements equivalent to physical measurements? To answer the first question, considerable efforts have been made to assess the resolution of the MicroScribe CMM and its consistency in the workspace. The spatial resolution of 0.09 mm and workspace consistency measured within 0.25 mm were eventually established using this MicroScribe CMM system.

Utilizing high-contrast metallic fiducial markers for both physical measurements and CT scan images provided a simplified condition and was able to specify the exact locations on different objects for further linear and angular measurement calculation in this study. This is different from the real-world clinical setting but provided a benchmark reference for the further clinical validation and application. A variety of factors may come into play that can potentially affect the accuracy, even the feasibility of the measurements, if the clinical task should involve measurement of the patient's anatomy including bony components with adjacent soft tissue and potentially contrast-enhanced vessels with adjacent soft tissue. For example, the measurements between bony components and blood vessel (with a contrast agent) are feasible with ImmersiveView™, but its accuracy of such process needs to be further verified with proper statistical evaluation before introducing to a clinical platform. However, the measurement of bony components and soft tissue combination are not feasible with this software because hard tissue does not have the same Hounsfield unit as that of soft tissue and thus requires different threshold values for segmentation in CT images. Multiple-value thresholding, however, was not available in this software platform at the moment this study was conducted. The solution may lay either on the multiple organ segmentation and surface modeling, which obviously introduces additional segmentation errors compared with the voxel-based approach, or on a “free oblique plane” approach in which the gray scale image is available with an oblique plane in any orientation in conjunction with the three dimensional rendering in the VR environment. The accuracy and reproducibility of measurements in these approaches warrant validation in the further study.

In order to assess the equivalence between VR measurements and physical measurements, we chose a systematic approach from the paired *t* test to Bland-Altman plots and to TOST. The paired *t* test is simple and straightforward, but a *p* value larger than 0.05 only suggests that the equivalence cannot be ruled out. Bland-Altman plots demonstrated such equivalence visually with 95% limits of agreement, while the TOST procedure provided further evidences with quantity associated with the clinical application. The predefined equivalence margins ± 0.5 mm for linear measurement and $\pm 0.7^\circ$ for angular measurement are in agreement with the known factors such as the CT voxel size of 0.325 mm in plane and 0.625 mm in slice thickness. The equivalence between the two methods was established by the measurements from all three objects in this study with 95% confidence intervals and corresponding equivalence margin. Furthermore, the equivalence margin provided an estimation of the accuracy for VR measurements. Considering the spatial resolution of physical measurements to be 0.09 mm in the MicroScribe CMM for linear distances, it is conceivable that reducing the voxel size of the medical image might improve the accuracy of the linear and angular measurements in the VR environment.

The tested version of this VR software, ImmersiveView™, is limited to CT, CBCT, and CTA. It does not support other image types such as MRI, ultrasound, or PET. In our study, we only considered CT images which are one of the conventional and widely used modalities for clinical diagnosis and surgical and treatment planning. Other approaches might have to be considered while designing validation experiments for different image types other than CT such as MRI, since fiducial markers made of metals may not be utilized during MRI scanning. In such case, we need to look for alternatives or consider identifiable markers made of materials with reasonable contrast.

Although within acceptable limitations, angular measurement showed more disparity than linear measurements. One probable explanation is that, unlike linear distance measurements where only two points need to be selected, angular measurements need at least three points to be involved. One additional point may increase the random uncertainty of the measure. This effect can potentially be reduced using smaller diameter fiducial markers, which, however, warrants a further study.

Raters' experience can add to the potential bias while taking measurements, specifically the VR measurements. To minimize bias from the recollection of the memory of previous trials/sessions, we have included a wide range of measurements in random orders at different time intervals for each set of measurements. This was also an essential consideration in the analysis of the reproducibility. In addition to the ICCs, the repeatability coefficient was also included in this study. With the same units of the measures, it may provide an estimate to

the repeatability in a way that is easy to associate with the practices and applications.

The major limitation of this study is the number of measurements (40 linear and 15 angular measurements) which may not provide strong power statistically [15]. The sample size was chosen based on the users' experience to manage an elongated session in the VR environment which may lead to dissatisfaction if extended. Too many measurements will lead to a prolonged experimental setting and thus may adversely affect the performance of the raters. Future studies will include the effect of measurements in the virtual environment with multiple raters and their variability in different situations compared with the conventional physical measurements and measurements from available 3D image analysis software.

While using virtual reality–based environment such as ImmersiveView™, it was clear that judgment of the individuals can play a critical role in the accuracy of the overall results. For example, while taking the measurements in the VR environment, it was often seen that considerable visual acuity was required to place the pointer at the exact same location and better acuteness would likely improve overall accuracy. Similarly, difference between the physical and VR measurements can be reduced by interactively enlarging the virtual scene so that small displacements from the actual position on the surface of the fiducial marker are less noticeable. The ability to provide magnification without distortion of image quality is one of the most essential aspects of the virtual environment.

Conclusion

A validation study of the basic measurement tools in the VR environment has been successfully conducted, through integrating available measurement devices and statistical analysis tools, under clinically oriented yet controlled conditions with selected objects (including dry skull), fiducial markers, medical CT scans, and selected raters. This investigation revealed that linear and angular measurements acquired within the ImmersiveView™ VR environment are equivalent to those made physically. There is a high degree of reproducibility and accuracy (through equivalence to the physical measurements) of linear and angular measurements in the VR environment, independent of the shape or complexity of the objects. The demonstrated results may provide a reference for more rigorous validation studies in a clinical setting in the future.

Clinical Relevance

Application software in a virtual immersive environment has the potential to be applied to its full extent in the medical and dental sectors in order to analyze anatomical structures with

greater efficacy [23, 24]. With the capacity to measure within clinical demanded accuracies, it can be used for critical measurements for diagnosis, treatment, and surgical planning such as craniofacial reconstructive surgery including orthognathic surgery, spine and orthopedic surgery, trauma care, neurosurgery, and interventional vascular surgery [25, 26]. In addition, virtual reality technology also likely has a role to facilitate in future interventions such as microsurgery and nanosurgery [27].

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent This article does not contain patient data.

References

- de Giorgio A, Romero M, Onori M, Wang L: Human-machine collaboration in virtual reality for adaptive production engineering. *Procedia Manufacturing* 11:1279–1287, 2017. <https://doi.org/10.1016/j.promfg.2017.07.255>
- Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, Danovitch I: Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. *Innovations in clinical neuroscience*, 14(1–2):14–21, 2017. <http://www.ncbi.nlm.nih.gov/pubmed/28386517>
- Riener R., Harders M. Introduction to virtual reality in medicine. In: *Virtual Reality in Medicine*, London: Springer, 2012. https://doi.org/10.1007/978-1-4471-4011-5_9
- King F, Jayender J, Bhagavatula SK, Shyn PB, Pieper S, Kapur T, Fichtinger G: An immersive virtual reality environment for diagnostic imaging. *J Med Robot Res* 1(01):1640003, 2016. <https://doi.org/10.1142/S2424905X16400031>
- Sutherland J, Belec J, Sheikh A, Chepelev L, Althobaity W, Chow BJW, Mitsouras D, Christensen A, Rybicki FJ, La Russa DJ: Applying modern virtual and augmented reality technologies to medical images and models. *J Digit Imaging* 32(1):38–53, 2019. <https://doi.org/10.1007/s10278-018-0122-7>
- Vertemati M, Cassin S, Rizzetto F, Vanzulli A, Elli M, Sampogna G, Gallieni M: A virtual reality environment to visualize three-dimensional patient-specific models by a mobile head-mounted display. *Surg Innov* 26:359–370, 2019. <https://doi.org/10.1177/1553350618822860>
- Jerdan SW, Grindle M, van Woerden HC, Kamel Boulos MN: Head-mounted virtual reality and mental health: critical review of current research. *JMIR Serious Games* 6(3):e14, 2018. <https://doi.org/10.2196/games.9226>
- Theart RP, Loos B, Powrie YS, Niesler TR: Improved region of interest selection and colocalization analysis in three-dimensional fluorescence microscopy samples using virtual reality. *PLoS One* 13(8):e0201965, 2018. <https://doi.org/10.1371/journal.pone.0201965>
- Theart RP, Loos B, Niesler TR: Virtual reality assisted microscopy data visualization and colocalization analysis. *BMC Bioinformatics* 18(2):64, 2017. <https://doi.org/10.1186/s12859-016-1446-2>

10. Sharma A, Bajpai P, Singh S, Khatter K: Virtual reality: blessings and risk assessment. arXiv preprint, 2017. <https://arxiv.org/abs/1708.09540v1>
11. Ball J, Balogh E, Miller BT eds. Improving diagnosis in health care. National Academies Press, 2015 <https://doi.org/10.7205/MILMED-D-15-00562>
12. de Oliveira MV, Wenzel A, Campos PS, Spin-Neto R: Quality assurance phantoms for cone beam computed tomography: a systematic literature review. *Dentomaxillofac Radiol* 46(3):20160329, 2017. <https://doi.org/10.1259/dmfr.20160329>
13. Bland JM, Altman DG: Measuring agreement in method comparison studies. *Stat Methods Med Res* 8:135–160, 1999. <https://doi.org/10.1177/096228029900800204>
14. Norušis MJ: SPSS 14.0 Guide to Data Analysis. USA: Prentice Hall, 2002
15. Schuirman DJ: A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm* 15:657–680, 1987. <https://doi.org/10.1007/BF01068419>
16. Bethel EW, Bastacky SJ, Schwartz K: Interactive stereo electron microscopy enhanced with virtual reality. *Proc. SPIE* 4660, Stereoscopic Displays and Virtual Reality Systems IX, 2002. <https://doi.org/10.1117/12.468055>
17. Brady R, Pixton J, Baxter G, Moran P, Potter CS, Carragher B, Belmont A: Crumbs: a virtual environment tracking tool for biological imaging. *Biomedical Visualization*, pp. 18–25, 1995. <https://doi.org/10.1109/BIOVIS.1995.528701>
18. McCarthy CJ, Alvin YC, Do S, Dawson SL, Uppot RN: Interventional radiology training using a Dynamic Medical Immersive Training Environment (DynaMITE). *J Am Coll Radiol* 15(5):789–793, 2018. <https://doi.org/10.1016/j.jacr.2017.12.038>
19. Hagedorn JG, Dunkers JP, Satterfield SG, Peskin AP, Kelso JT, Terrill JE: Measurement tools for the immersive visualization environment: steps toward the virtual laboratory. *J Res Natl Inst Stand Technol* 112(5):257–270, 2007. <https://doi.org/10.6028/jres.112.019>
20. Preim B, Tietjen C, Spindler W, Peitgen H: Integration of measurement tools in medical 3d visualizations. *IEEE Visualization, VIS*, 21–28, 2002. <https://doi.org/10.1109/VISUAL.2002.1183752>
21. Reitinger B, Schmalstieg D, Bornik A, Beichel R: Spatial analysis tools for virtual reality-based surgical planning. In *3D User Interfaces, IEEE*, pp. 37–44, 2006. <https://doi.org/10.1109/VR.2006.121>
22. Scholtes VA, Terwee CB, Poolman RW: What makes a measurement instrument valid and reliable? *Injury* 42(3):236–240, 2011. <https://doi.org/10.1016/j.injury.2010.11.042>
23. Pantelidis, P., Chorti, A., Papagiouvanni, I., Paparoidamis, G., Drosos, C., Panagiotakopoulos, T., & Sideris, M. Virtual and augmented reality in medical education. *Medical and Surgical Education - Past, Present and Future*, 2018 <https://doi.org/10.5772/intechopen.71963>
24. Aim F, Lonjon G, Hannouche D, Nizard R: Effectiveness of virtual reality training in orthopaedic surgery. *Arthroscopy* 32(1):224–232, 2016. <https://doi.org/10.1016/j.arthro.2015.07.023>
25. Tepper OM, Sorice S, Hershman GN, Saadeh P, Levine JP, Hirsch D: Use of virtual 3-dimensional surgery in post-traumatic craniomaxillofacial reconstruction. *J Oral Maxillofac Surg* 69(3):733–741, 2011. <https://doi.org/10.1016/j.joms.2010.11.028>
26. Orentlicher G, Goldsmith D, Horowitz A: Applications of 3-dimensional virtual computerized tomography technology in oral and maxillofacial surgery: current therapy. *J Oral Maxillofac Surg* 68(8):1933–1959, 2010. <https://doi.org/10.1016/j.joms.2010.03.013>
27. McCloy R, Stone R: Science, medicine, and the future, virtual reality in surgery. *BMJ* 323(7318):912–915, 2001. <https://doi.org/10.1136/bmj.323.7318.912>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.