**Appendix 1.** Validation of trueness and precision of a novel augmented reality platform for measuring acetabular cup implant orientation in a training simulator

**Measurement**

A dedicated motion-capture laboratory was used to validate the accuracy of the enhanced augmented reality (AR) headset and phantom hip platform. Eighteen Prime 13W (OptiTrack, Corvallis, OR, USA) infrared cameras mounted on a rig offer precision tracking, validated to an accuracy of 0.3 mm and 240 frames per second. These were therefore used as the accepted reference value (ARV). The OptiTrack system allows for three-dimensional (3-D) reconstruction of marker point movements and calculates angular values using an in-house program. A set of retroreflective markers was placed on the AR system headset, the acetabular cup introducer, and the benchtop simulator to log and track the movements of each. The resulting measurements from the OptiTrack system enabled the determination of the cup orientation (Fig. 1).

Measurements of cup orientation were also made using the AR platform. The camera used black and white tracking fiducials to determine cup orientation. These were fixed onto the base and frame of the benchtop simulator, and the sides of the acetabular cup introducer, with a further marker placed on a platform fixed to the lateral aspect of the hemipelvis, superior to the acetabulum. Bespoke software, which calculated operative inclination and anteversion in relation to the anterior pelvic plane, was uploaded to the AR platform that allowed the participants to visualize their acetabular positioning relative to the target orientation. This consisted of a heads-up display (HUD) “crosshair,” within which a red dot would move with the introducer movement. When the orientation is within 1° of error from the target orientation, the dot flashes green to inform participants they have reached the desired orientation.

The cup orientation angles determined from the relative positioning of the cup introducer and the benchtop trainer were then compared with those recorded from the OptiTrack ARV to assess the level of agreement.

**Validation Testing**

The motion-capture axes were defined using a protractor fitted with three retroreflective markers, enabling post hoc 3-D reconstruction. The OptiTrack cameras were calibrated using a testing wand consisting of a further three markers, which was moved around the camera field to register the coordinate values. After validation, retroreflective markers were fixed to each of the points of interest (AR platform, introducer, and benchtop trainer). Each object was placed in the OptiTrack system alone so that the markers could be registered and grouped to form a single object on the motion-tracking software.

Each test consisted of a before-and-after phase in which the cup introducer was synchronized with the OptiTrack system. This consisted of two steps: (1) the cup introducer was moved around the acetabulum in a circular motion so that the HUD HoloLens crosshair dot was moving around the extremes of the crosshair. This was done for a total of four rotations; and (2) the cup introducer was moved back and forth along the transverse plane (from posterior to anterior) for a total of four complete motions. After synchronization, participants (three hip surgeons, including author KL) were made to orient the introducer to each of the six clinical target angles in turn (Supplemental Table 1) using the HUD crosshair as guidance. The target angle at which the HUD dot flashes green was remotely determined by a computer with which each angle was selected. After each target was achieved, the synchronization steps were repeated to conclude the test. Each participant repeated the test twice.

**Analysis of Measurements**

As a result of common confusion over the exact definitions of terms such as “accuracy” and “precision,” the latest definitions from the American Society for Testing and Materials (ASTM) outlined by Langlois and Hamadouche were used [[2](#_ENREF_2)]. Accuracy is defined as the degree of total error of the system considering both the degree of systematic error (trueness) and the degree of random error (precision). To validate the accuracy of the HoloLens, both trueness and precision were reported. Trueness is defined as the degree of closeness of values between the measured system and the ARV, also referred to as the bias. Precision is defined as the amount of agreement between reported measurements obtained through repetition. It has no bearing on the closeness of the true ARV.

Analysis was performed using Microsoft Excel (Microsoft Inc, Redmond, WA, USA) and SPSS software (SPSS Inc, Chicago, IL, USA). Data were tested for normality and the Wilcoxon signed-rank test to compare values with a significance value of 0.05. Scatter graphs for each test were constructed using raw data from the AR platform and OptiTrack system. The values for the periods at which target angles were achieved were collected, and the means of these were used for further analysis. Trueness was computed as the mean difference between the mean inclination and anteversion values determined by the AR platform and OptiTrack system. Precision was computed as the SD of the differences between the means of inclination and anteversion values.

Scatterplots and Pearson’s coefficients were constructed to show the correlation between the mean inclination and anteversion measurements (Fig. 2), and Bland-Altman plots used to compare agreement between AR platform and OptiTrack system measurements (Fig. 3). Linear regression was also computed against the means of the AR and OptiTrack to check for proportional bias.

Repeatability and reproducibility were computed to evaluate intrarater and interrater reliability, respectively, using two-way mixed effects, absolute consistency intraclass correlation coefficients (ICC). Repeatability was calculated by performing the ICC index across tests performed by the same operator with reproducibility performed across tests with differing operators.

**Accuracy of the AR Platform**

The trueness/bias was 0° (95% confidence interval limit 1°) for inclination and -1° (95% CI limit 1.8°) for anteversion. The distribution of the difference between methods was shown to be nonparametric for both inclination (p < 0.01) and anteversion (p < 0.001). The Wilcoxon signed-rank test showed a statistically significant difference between the bias and zero for both inclination (Z = 276, p < 0.01) and anteversion (Z = 37, p < 0.001). Precision was 0° for inclination and 1° for anteversion. Mean AR platform values for inclination were 42° ± 5° and 20° ± 4° for anteversion. Mean OptiTrack system angles 41° ± 5° and 21° ± 4° for inclination and anteversion, respectively. Repeatability for the AR platform was 0.995 (ICC) for inclination and 0.989 for anteversion. Reproducibility for the AR platform was 0.999 for inclination and 0.995 for anteversion (Supplemental Table 2).

**Correlation Between the AR Platform and the OptiTrack System**

There was a significant correlation between AR platform and OptiTrack system measurements for both inclination (r = 0.996, Pearson’s correlation coefficient, p < 0.001) and anteversion (r = 0.974, Pearson’s correlation coefficient, p < 0.001).

**Bland-Altman Plots of Agreement**

Bland-Altman 95% level of agreement was calculated as 1° for inclination and 2° for anteversion. The 95% CI lines were drawn at 1° and -1° for inclination and 1° and -3° for anteversion. Ninety-six percent of inclination values and 94% of anteversion values were shown to be within the 95% CIs. The linear regression calculated was insignificant (p > 0.05).

**Conclusions**

Translated angular values obtained from the AR platform are very accurate with the error when compared with the ARVs are only 0° for inclination and 1° for anteversion. The translation into angular values is also highly precise with an error of 0° for inclination and 1° for anteversion.

The Pearson scatterplots show good correlation and the Bland-Altman plots showed that outliers constituted only 4% and 6% of inclination and anteversion values, respectively. Furthermore, the plot shows that the differences between measurement methods stay fairly consistent as the mean increases for both inclination and anteversion along with consistent variability; therefore, accuracy is independent of the magnitude of the angular value. The possibility of proportional bias can be excluded as a result of the insignificant linear regression analysis against the mean.

Both intrarater and interrater reliability for the AR platform were shown to be very strong through the ICC computations; thus, values are replicated to a very high degree within and between operators. The ICC tests reflect both the degree of correlation and agreement between the measurements taken by the AR platform, making them a very strong measure of reliability.

The reproducibility values are greater than those of the repeatability for both inclination and anteversion, indicating that there is greater agreement between measurements obtained across different operators than between a single operator. This is unusual; however, it is probably the result of random individual variations of measurement and thus is negligible, because the discrepancy between each pair is < 0.01%.

This AR platform was far more true and precise than any previously described conventional floor-mounted navigation systems. Kalteis et al. [[1](#_ENREF_1)] demonstrated a bias and precision of 2.9° and 2.2°, respectively, for inclination, and 4.2° and 3.3° for anteversion when using imageless navigation. Ybinger et al. [[3](#_ENREF_3)] reported a bias of 3.6° and precision of 4.5° for inclination and 6.6° and 7.4° for anteversion Although the platform has been adapted for training for arthroplasty orientation in this study, further validation within the surgical setting is required to explore the platform’s use as an intraoperative navigation tool.

**References**

1. Kalteis T, Handel M, Herold T, Perlick L, Baethis H, Grifka J. Greater accuracy in positioning of the acetabular cup by using an image-free navigation system. *Int Orthop*. 2005;29:272-276.

2. Langlois J, Hamadouche M. Current recommendations for assessing the reliability of a measurement tool: a survival guide for orthopaedic surgeons. *Bone Joint J*. 2016;98-B:166-172.

3. Ybinger T, Kumpan W, Hoffart HE, Muschalik B, Bullmann W, Zweymuller K. Accuracy of navigation-assisted acetabular component positioning studied by computed tomography measurements: methods and results. *J Arthroplasty*. 2007;22:812-817.

**Fig. 1** AR platform validation within an OptiTrack motion capture rig.

D:\Dropbox\PhD\Enhanced AR\CORR Manuscript\CORR Review\Appendix Figure 1.tif

**Fig.2A-B** Scatterplots demonstrate Pearson’s correlation coefficients between the AR platform and OptiTrack system measurements for (A) inclination and (B) anteversion D:\Dropbox\PhD\Enhanced AR\CORR Manuscript\CORR Review\Appendix Figure 2.tif

**Fig. 3A-B** Bland-Altman plots demonstrate agreement between AR platform and OptiTrack methods for (A) inclination and (B) anteversion. Middle red line = mean; outer red lines = 2 SD from mean.

D:\Dropbox\PhD\Enhanced AR\CORR Manuscript\CORR Review\Appendix Figure 3.tif

**Supplemental Table 1.** Orientation targets reflecting clinically relevant acetabular component inclination and anteversion target angles

|  |  |
| --- | --- |
| Target angle (inclination°-anteversion°) | Clinical situation |
| 40°-20° | Male hip resurfacing |
| 35°-25° | Female hip resurfacing |
| 45°-25° | Female hip arthroplasty |
| 50°-20° | Contralateral hip arthrodesis |
| 45°-15° | Male hip arthroplasty |
| 35°-15° | Male hip resurfacing (small) |

**Supplemental Table 2.** Accuracy of AR platform for acetabular cup orientation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measurement | AR inclination | AR/OptiTrack inclination | OptiTrack inclination | AR anteversion | AR/OptiTrack anteversion | OptiTrack anteversion |
| Mean ± SD | 41.6° ± 5.4° |  | 41.4° ± 5.3° | 20.0° ± 4.1° |  | 20.9° ± 4.0° |
| Trueness (bias) |  | 0.24° |  |  | -0.90° |  |
| Bland-Altman limits of agreement |  | 0.91° |  |  | 1.78° |  |
| Precision |  | 0.46° |  |  | 0.91° |  |
| Repeatability | 0.995\* |  |  | 0.989\* |  |  |
| Reproducibility | 0.999\* |  |  | 0.995\* |  |  |

\*Significant value for intraclass correlation coefficient = p < 0.05; AR = augmented reality.