Value of 3D Preoperative Planning for Primary Total Hip Arthroplasty Based on Artificial Intelligence Technology

Jiabang Huo  
The Third Affiliated Hospital of Southern Medical University

Guangxin Huang  
The Third Affiliated Hospital of Southern Medical University

Dong Han  
The Third Affiliated Hospital of Southern Medical University

Xinjie Wang  
The Third Affiliated Hospital of Southern Medical University

Yufan Bu  
The Third Affiliated Hospital of Southern Medical University

Ya Chen  
Southern Medical University

DaoZhang Cai (✉ cdz@smu.edu.cn)  
Southern medical university https://orcid.org/0000-0003-2294-5411

Chang Zhao  
The Third Affiliated Hospital of Southern Medical University

Research article

Keywords: AI, Artificial Intelligence, Preoperative Planning, total hip arthroplasty, THA

DOI: https://doi.org/10.21203/rs.3.rs-120578/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: Accurate preoperative planning is an important step for accurate reconstruction in total hip arthroplasty (THA). Presently, preoperative planning is completed using either a two-dimensional (2D) template or three-dimensional (3D) mimics software. With the development of artificial intelligence (AI) technology, AI HIP, a planning software based on AI technology can quickly and automatically identify acetabular and femur morphology, and automatically match the optimal prosthesis size. However, the accuracy and feasibility of its clinical application still needs to be further verified. The purposes of this study were to investigate the accuracy and time efficiency of AI HIP in preoperative planning for primary THA, compared with 3D mimics software and 2D digital template; and further analyze the factors that influence the accuracy of AI HIP.

Methods: A prospective study was conducted on 53 consecutive patients (59 hips) undergoing primary THA with cementless prostheses in our department. All preoperative planning was completed using AI HIP as well as 3D mimics and 2D digital template. The predicted component size and the actual implantation results were compared to determine the accuracy. The templating time was compared to determine the efficiency. Furthermore, the potential factors influencing the accuracy of AI HIP were analyzed including sex, body mass index (BMI), and hip dysplasia.

Results: The accuracy in predicting the acetabular cup and femoral stem was 74.58% and 71.19%, respectively, for AI HIP; 71.19% (P = 0.743) and 76.27% (P = 0.468), respectively, for 3D mimics; 40.68% (P < 0.001) and 49.15% (P = 0.021), respectively, for 2D digital templating. The templating time using AI HIP was 3.91±0.64 min, which was equivalent to 2D digital templates (2.96±0.48 min, P < 0.001), but shorter than 3D mimics (32.07±2.41 min, P < 0.001). Acetabular dysplasia(P = 0.021), rather than sex and BMI, was an influential factor in the accuracy of AI HIP templating. Compared to patients with developmental dysplasia of the hip (DDH), the accuracy of acetabular cup in the non-DDH group was better (P = 0.021), but the difference in the accuracy of the femoral stem between the two groups was statistically insignificant (P = 0.062).

Conclusion: AI HIP showed excellent reliability for component size in THA. Acetabular dysplasia may affect the accuracy of AI HIP templating.

Background

Total hip arthroplasty (THA) is a mature orthopedic surgery that can reduce pain and improve joint function. However, when the prosthesis is improperly selected, it will lead to a series of intraoperative and postoperative complications, such as: leg length discrepancy [1], joint dislocation [2], and periprosthetic fracture [3]. This will lead to an increase in the revision rate, resulting in patient dissatisfaction and lower clinical score.

Therefore, a successful THA requires the selection of the most suitable prosthesis size and restoration of joint biomechanics. Accurate preoperative planning can help obtain information about reconstruction
results in advance. Although they have accurate intraoperative planning for navigation and robots, they are time-consuming and costly, with high error rates of the manual system, and long learning curve involved [13]. Therefore, they cannot be an ideal conventional application planning technology at present. At present, 2D templates are mostly used in clinical practice, but their accuracy is low due to the influence of X-ray magnification [5] and projection position [6]. For example, Petretta et al. [7] compared the preoperative planning of 260 patients with the film template method and digital template; the complete accuracy rate of both methods was only 28%. Holzer et al. [8] conducted a retrospective study on 632 patients using a digital template, and the complete accuracy rate of predicting acetabular cup and femoral stem was 37% and 42%, respectively, while that of Efe [9] for 169 patients using a digital template was 33.7% and 36%, respectively. A few hospitals use 3D planning software for preoperative planning of THA. Although, the accuracy is relatively high, the acetabular rotation center, femoral osteotomy position, and femoral stem implantation depth, which require constant manual adjustment and cumbersome operation, could not be located. Zeng [10] et al. used 3D mirmics to make preoperative planning for 20 patients with Crowe IV hip dysplasia, and the complete accuracy of acetabular cup was 70% and 100% within one size. Sariali [11] et al. [6] used 3D software for preoperative planning of 30 patients receiving THA, and the accuracy of predicting one size of the femoral stem and acetabular cup was 100% and 96%, respectively. Inoue et al. [12] used 3D planning software Zed hip to plan 65 hips of 57 patients. The complete accuracies of acetabular cup and femoral stem were 92% and 65%, respectively, and the accuracy rate of one acetabular cup and femoral stem was 100% and 98%, respectively.

The above 2D and 3D planning methods are subjective or time-consuming. Therefore, how can we achieve rapid planning and avoid planning deviation caused by personal experience? With the development of artificial intelligence (AI) technology, AI HIP, a planning software based on AI technology can quickly and automatically identify acetabular and femur morphology, and automatically match the optimal prosthesis size. However, the accuracy and feasibility of its clinical application still needs to be further verified. As far as we know, the accuracy of AI HIP has not been studied.

We therefore, attempted to evaluate this through a prospective study with objectives to evaluate: (1) The accuracy and time efficiency of AI HIP and compare with the 3D mimics and the 2D digital template. (2) The factors that influence the accuracy of AI HIP: Sex, BMI, and hip dysplasia.

**Materials And Methods**

This was a prospective clinical study to assess the accuracy of AI HIP in primary THA and to further analyze the factors influencing its accuracy. The research protocol was approved by the institutional review board and all participants gave their written informed consent.

From October 2019 to July 2020, preoperative planning was performed for every patient who underwent primary THA with cementless prostheses. All included patients underwent preoperative CT scans and routine X-ray examination. A CT scan from the upper pelvic edge to more than 10cm below the lesser trochanter was performed with a Toshiba Aquilion 64 CT scanner (Voltage: 120 kVp, electric current:
Intelligent mA control, according to the patient's fat and thin draw mA curve, CT matrix: 512 × 512 matrix, slice thickness: 1 mm), and were stored in DICOM format to the mobile disc.

On the eve of the surgery, each patient was planned using AI HIP, 3D mimics, and 2D digital template, which were divided into three groups: AI, 3D, and 2D groups. The results of each group were recorded, including: acetabular cup size, femoral stem size, and templating time. The templating time is the period from the beginning of importing data to the end of planning. All THAs were performed by the same senior arthroplasty surgeon (C.D.Z.), using the same posterolateral approach. The actual size of the prosthesis implanted during the operation was obtained from surgical records. The preoperative planning results of each group were compared with the actual intraoperative implant sizes to determine their accuracy. In this study, all preoperative planning was performed by the author (H.J.B.). To identify planning reliability, all plannings were checked by two experienced surgeons (Z.C. and H.G.X.). If there was a disagreement, the planning was repeated with all the three surveyors (H.J.B., Z.C. and H.G.X.) agreed upon.

To ensure that the surgeon was not bound by the planning results during the operation and to maintain the consistency between the planning and the operation, we adopted double-blind surgeons. The surveyors, the surgical procedure and preoperative planning were independent from each other.

preoperative planning

1. AI group preoperative planning

1.1 The working principle of AI HIP:

The raw CT data were input into the segmentation model and 3D recognition model, both of which are deep learning models based on neural networks. The segmentation module separates the different parts of the bone and uses a multiple 3D segmentation neural network (using Unet as baseline) to ensure segmentation accuracy, dividing the bone into different areas such as the pelvic and femur areas. The anatomical point recognition module uses the point recognition neural network of face recognition and pattern recognition technology to accurately calculate the coordinates of corresponding points, locate the key points on the skeleton, and output the coordinates of the points, such as: the lesser trochanter, greater trochanter, tear drop, and anterior superior iliac spine. Then, an automatic search engine based on big data and reinforcement learning is used to match the optimal prosthesis, and finally, the perfect intelligent results are planned (Fig. 1).

1.2 Planning process:

The DICOM format of the hip CT data was converted to " .cmg " format, the " .cmg " format image data of the patient was imported into AI HIP software (Changmugu), which automatically segmented the bones
through AI technology and established 3D models of the pelvis and femur. Through amplification or rotation, the 3D images could be visualized and the severity of the lesions could be grasped (Fig. 2). Through neural network automatic correction of the pelvis and simulating pelvic anteroposterior X-ray, leg length discrepancy, and femoral offset were calculated automatically. According to the modeling and calculation before, according to the automatic recognition of acetabular anteroposterior diameter and upper and lower diameter, the optimal acetabular cup size was selected from the prosthesis database and automatically matched to the acetabular socket model. The acetabular cup is automatically placed with an anteversion angle of 20° and an abduction angle of 40°, so that the acetabular coverage rate can be calculated automatically. The coverage rate can change with real-time changes in the abduction angle and anteversion angle. The appearance and size of the prosthesis in the database are exactly the same as the prosthesis used in the operation, and there is no prosthesis deviation. If you are not satisfied with the automatic placement of AI, you can fine-tune it. Based on the automatic recognition of bone markers such as the epiphysis of the femoral shaft and the inner and outer cortex of the femoral shaft, the software selected the optimal femoral stem size from the prosthesis database, matched it automatically to the 3D model, and the femoral head size was selected according to the size of the acetabular cup (Fig 3). After the prosthesis was placed, the leg length difference and offset under ideal conditions could be calculated automatically. At this time, femoral neck osteotomy could be performed. Finally, after the acetabular cup, femoral stem, and femoral head were placed, the software automatically calculated the length difference of the lower limbs and the height of the osteotomy (distance from the osteotomy plane to the tip of the lesser trochanter), and simulated the effect of pelvic anteroposterior radiography after operation (Fig 4, Fig 5).

2.3D group preoperative planning

2.1 Fabrication of 3D prosthesis:

In order to make the shape and size of the 3D prosthesis consistent with the prosthesis prototype used in surgery, we used a 3D scanner (Epson E-018) to scan the PINACLE acetabular cup, SUMMIT, and CORAIL stem of various sizes. The original size model was established by reverse modeling with Geomagic Studio 2012 (Raindrop) and saved in STL format.

2.2 Planning process:

The patient CT data in DICOM format of mobile disc were imported into mimics19 software (materialize). The original mask of the bone was selected by a threshold segmentation function. The pelvis and femur were segmented by region growing, editing mask function, and Boolean operation. The 3D model was then reconstructed. First simulate the acetabular cup implantation, measure the anterior and posterior diameter length of the acetabulum, select the PINNACLE 3D cup with similar anterior and posterior diameter lengths to import into mimics in STL format, make the 3D cup adhere to the anterior and
posterior wall of the acetabulum on the transverse section, the coronal, sagittal, and select the optimal size of the acetabular cup after repeated adjustments on the coronal, sagittal, and 3D pelvic models [14] [15]. The femoral stem was then planned, and the STL format of the SUMMIT or CORAIL femoral stem was imported, and the best prosthesis fitted closely with the epiphyseal cortex of the femoral shaft [12] [16]. The appropriate size of the femoral head was selected, based on the size of the acetabular cup (Fig. 6).

3. 2D group preoperative planning

2D digital templating was based on the AP pelvic X-ray, which was performed in a standard manner with a 38mm magnification marker ball in the perineum area at the level of the greater trochanter. The patient's standard pelvic anteroposterior X-ray was imported into the 2D digital template (Smart joint, Depuy Synthes), and the diameter of the marker ball was set to 38mm. The digital template prosthesis was superimposed on the AP pelvic radiograph, and the optimal size of the prosthesis was selected according to the matching situation between the: acetabular cup, the acetabulum, femoral stem, and femoral medullary cavity [17](Fig. 7).

Finally, the accuracy and time efficiency of the AI and 3D groups, AI group, and 2D group were compared, and the factors influencing the accuracy of AI HIP were further analyzed, mainly from three aspects: 1) sex; 2) BMI (according to the CHINA standard: Normal: 18.5 ~ 23.9; Pre-overweight/obese: 24–27.9; Obesity: ≥28), hip dysplasia (developmental dysplasia of the hip (DDH) group and non-developmental dysplasia of the hip (non-DDH) group).

Statistical analysis

Quantitative data are described as mean ± standard deviation, while categorical variables are described as counts (percentage). The prediction was considered accurate if it was consistent with the size used intraoperatively. Chi-square test or Yates corrected chi-square test (when appropriate) was used to compare the accuracy of the three methods. The multivariate model was fitted using a generalized estimation equation (GEE) to estimate the accuracy of the three methods for cup and stem, respectively, adjusting for: sex, surgical side, hip dysplasia, age, and BMI. The distribution of dependent variables was binomial with log as the link function. The operation time comparison among the three methods was performed using the GEE, and the dependent variable distribution was a Gaussian distribution. The significant level alpha was set to 0.05, and all statistical analyses were performed using R 3.6.3 software.

Results

From October 2019 to July 2020, a total of 53 patients (29 men and 24 women) were included in the study. The mean age of the patients was 57.4 ± 1.7 years (27–79 years). There were 25 left and 34 right hips (59 hips). According to CHINA criteria: 6 (10%) hips were underweight, 33 (56%) were in the normal
range, 13 (22%) were pre-overweight/obese, and 7 (12%) were obese. The primary diagnosis was: DDH in 16 (27.12%) hips, osteoarthritis (OA) in 16 (27.12%) hips, osteonecrosis of the femoral head (ON) in 16 (27.12%) hips, ankylosing spondylitis (AS) in 9 (15.25%) hips, and rheumatoid arthritis (RA) in 2 (3.39%) hips. OA, ON, AS, and RA were classified as non-DDH groups, 43 (72.88%) hips. The PINACLE cup (DePuy, Warsaw, IN, USA) was used in all hips, SUMMIT (DePuy, Warsaw, IN, USA) stem was used in 43 hips, and a CORAIL (DePuy, Warsaw, IN, USA) stem was used in 16 hips.

Accuracy of prosthesis

The accuracy in predicting the acetabular cup and femoral stem was 74.58% (44 of 59) and 71.19% (42 of 59), respectively, for AI HIP group; 71.19% (42 of 59) and 76.27% (45 of 59), respectively, for 3D mimics group; 40.68% (24 of 59) and 49.15% (29 of 59), respectively, for 2D digital group. The accuracy of AI HIP was significantly higher than that of the 2D digital template, with a statistical difference between the two methods (OR: 0.196 (0.087-0.441), P < 0.001 for cup; 0.377 (0.164-0.864), P =0.021 for stem). Compared with the 3D mimics, the accuracy of AI HIP was comparable to that of the 3D mimics, with no significant difference between the two methods (cup: P =0.743, stem: P =0.468).
Table 1
Generalized estimation equation model for predicting cup consistency rate of AI, 3D and 2D methods

<table>
<thead>
<tr>
<th>Factors</th>
<th>Coefficient</th>
<th>SE</th>
<th>Wald</th>
<th>P value</th>
<th>OR(95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.838</td>
<td>1.162</td>
<td>0.520</td>
<td>0.471</td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>-0.144</td>
<td>0.440</td>
<td>0.108</td>
<td>0.743</td>
<td>0.865 (0.365-2.051)</td>
</tr>
<tr>
<td>2D</td>
<td>-1.63</td>
<td>0.414</td>
<td>15.504</td>
<td>&lt; 0.001</td>
<td>0.196 (0.087-0.441)</td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-1.275</td>
<td>0.456</td>
<td>7.831</td>
<td>0.005</td>
<td>0.279 (0.114-0.682)</td>
</tr>
<tr>
<td>Left</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>0.166</td>
<td>0.392</td>
<td>0.179</td>
<td>0.672</td>
<td>1.181 (0.547-2.547)</td>
</tr>
<tr>
<td>DDH</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-DDH</td>
<td>1.550</td>
<td>0.499</td>
<td>9.637</td>
<td>0.002</td>
<td>4.71 (1.771-12.529)</td>
</tr>
<tr>
<td>Age(yrs)</td>
<td>-0.003</td>
<td>0.013</td>
<td>0.047</td>
<td>0.828</td>
<td>0.997 (0.973-1.023)</td>
</tr>
<tr>
<td>BMI(&lt;18.5)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI[18.5-23.9]</td>
<td>0.260</td>
<td>0.683</td>
<td>0.145</td>
<td>0.704</td>
<td>1.297 (0.34-4.95)</td>
</tr>
<tr>
<td>BMI[18.5-23.9]</td>
<td>-0.222</td>
<td>0.737</td>
<td>0.090</td>
<td>0.764</td>
<td>0.801 (0.189-3.4)</td>
</tr>
<tr>
<td>BMI(≥28)</td>
<td>-0.653</td>
<td>0.805</td>
<td>0.657</td>
<td>0.418</td>
<td>0.521 (0.107-2.523)</td>
</tr>
</tbody>
</table>
## Table 2

<table>
<thead>
<tr>
<th>Factors</th>
<th>Coefficient</th>
<th>SE</th>
<th>Wald</th>
<th>P value</th>
<th>OR(95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.709</td>
<td>1.141</td>
<td>0.385</td>
<td>0.535</td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>0.331</td>
<td>0.456</td>
<td>0.526</td>
<td>0.468</td>
<td>1.392(0.570-3.401)</td>
</tr>
<tr>
<td>2D</td>
<td>-0.976</td>
<td>0.423</td>
<td>5.311</td>
<td>0.021</td>
<td>0.377(0.164-0.864)</td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.759</td>
<td>0.397</td>
<td>3.643</td>
<td>0.056</td>
<td>0.468(0.215-1.021)</td>
</tr>
<tr>
<td>Left</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>0.361</td>
<td>0.372</td>
<td>0.944</td>
<td>0.331</td>
<td>1.435(0.693-2.971)</td>
</tr>
<tr>
<td>DDH</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-DDH</td>
<td>0.599</td>
<td>0.461</td>
<td>1.692</td>
<td>0.193</td>
<td>1.821(0.738-4.492)</td>
</tr>
<tr>
<td>Age(yrs)</td>
<td>0.032</td>
<td>0.014</td>
<td>5.436</td>
<td>0.020</td>
<td>1.033(1.005-1.061)</td>
</tr>
<tr>
<td>BMI(&lt;18.5)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI[18.5-23.9]</td>
<td>-0.312</td>
<td>0.650</td>
<td>0.230</td>
<td>0.632</td>
<td>0.732(0.205-2.620)</td>
</tr>
<tr>
<td>BMI[18.5-23.9]</td>
<td>-0.898</td>
<td>0.678</td>
<td>1.756</td>
<td>0.185</td>
<td>0.407(0.108-1.538)</td>
</tr>
<tr>
<td>BMI(≥28)</td>
<td>0.565</td>
<td>0.850</td>
<td>0.442</td>
<td>0.506</td>
<td>1.76(0.333-9.315)</td>
</tr>
</tbody>
</table>

### Time efficiency

The preoperative planned operation time of the AI group was 3.91±0.64 minutes, 32.07±2.41 minutes for the 3D group, and 2.96±0.48 minutes for the 2D group. The operation time of AI HIP was 28.105 min shorter than that of the 3D mimics (P <0.001), and 0.891 min longer than that of the 2D digital template (P <0.001) [Table 3].
Table 3
Generalized estimation equation model for Time efficiency of AI, 3D and 2D methods

<table>
<thead>
<tr>
<th>Factors</th>
<th>Coefficient</th>
<th>SE</th>
<th>Wald</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.604</td>
<td>0.881</td>
<td>16.73</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>AI</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D</td>
<td>28.105</td>
<td>0.279</td>
<td>10158.119</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>2D</td>
<td>-0.891</td>
<td>0.113</td>
<td>62.435</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Female</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.470</td>
<td>0.197</td>
<td>5.712</td>
<td>0.017</td>
</tr>
<tr>
<td>Left</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>0.619</td>
<td>0.780</td>
<td>0.629</td>
<td>0.428</td>
</tr>
<tr>
<td>DDH</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-DDH</td>
<td>0.020</td>
<td>0.324</td>
<td>0.004</td>
<td>0.950</td>
</tr>
<tr>
<td>Age(yrs)</td>
<td>-0.007</td>
<td>0.008</td>
<td>0.913</td>
<td>0.339</td>
</tr>
<tr>
<td>BMI(&lt;18.5)</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI[18.5-23.9]</td>
<td>-0.044</td>
<td>0.394</td>
<td>0.013</td>
<td>0.910</td>
</tr>
<tr>
<td>BMI[24-27.9]</td>
<td>0.014</td>
<td>0.395</td>
<td>0.001</td>
<td>0.971</td>
</tr>
<tr>
<td>BMI(≥28)</td>
<td>0.168</td>
<td>0.531</td>
<td>0.101</td>
<td>0.751</td>
</tr>
</tbody>
</table>

Analysis of factors influencing the accuracy of the AI HIP software

1. Sex: the accuracy of acetabular cup in female patients was 76.00% (19 of 25) and male patients was 73.53% (25 of 34), there was no statistically significant difference between the two groups (P = 0.829). The accuracy of the femur stem in female patients was 68.00% (17 of 25) and male patients was 73.53% (25 of 34), and there was no statistically significant difference between the two groups (P = 0.643). It can be seen that sex has no effect on the accuracy of planning acetabular cup and femoral stem by AI HIP.

2. BMI: The accuracy of acetabular cup: normal group (18.5 ~ 23.9) was 78.79% (26 of 33); the overweight/obese group (24–27.9) was 69.63% (9 of 13); obese group (BMI≥28) was 69.63% (4 of 7), and there was no statistically significant difference among the three groups (P = 0.600).
Concerning the accuracy of the femur stem, the normal group (18.5 ~ 23.9) was 66.67% (22 of 33); the overweight/obese group (24–27.9) was 76.92% (10 of 13), and the obese group (BMI≥28) was 85.71% (6 of 7). There was no statistically significant difference between the three groups (P = 0.725). Therefore, BMI has no effect on the accuracy of planning acetabular cup and femoral stem using AI HIP.

3. Hip dysplasia: Concerning the accuracy of acetabular cup, the DDH group was 50% (8 of 16), while the non-DDH group was 83.72% (36 of 43). The difference between the two groups was statistically significant (P = 0.021). As for the accuracy of the femoral stem, the DDH group was 50% (8 of 16), while the non-DDH group was 79.07% (34 of 43) respectively. There was no statistically significant difference between the two groups (P = 0.062). Therefore, in DDH patients, the accuracy of AI HIP to predict the acetabular cup seemed low, though, it had little impact on the accuracy of the femoral stem (Table 4, 5).

### Table 4
Chi square test for the accuracy of predicting cup of AI, 3D and 2D methods

<table>
<thead>
<tr>
<th>Factors</th>
<th>subgroup</th>
<th>N</th>
<th>Inaccurate (n=15)</th>
<th>Accuracy (n=44)</th>
<th>Chi square</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis (%)</td>
<td>DDH</td>
<td>16</td>
<td>8(50.00%)</td>
<td>8(50.00%)</td>
<td>5.328</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>non-DDH</td>
<td>43</td>
<td>7(16.28%)</td>
<td>36(83.72%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (%)</td>
<td>female</td>
<td>25</td>
<td>6(24.00%)</td>
<td>19(76.00%)</td>
<td>0.046</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>34</td>
<td>9(26.47%)</td>
<td>25(73.53%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI(%)</td>
<td>BMI(&lt;18.5)</td>
<td>6</td>
<td>1(16.67%)</td>
<td>5(83.33%)</td>
<td>1.869</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>BMI[18.5-23.9]</td>
<td>33</td>
<td>7(21.21%)</td>
<td>26(78.79%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI[24-27.9]</td>
<td>13</td>
<td>4(30.77%)</td>
<td>9(69.23%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI(≥28)</td>
<td>7</td>
<td>3(42.86%)</td>
<td>4(57.14%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors</td>
<td>subgroup</td>
<td>N</td>
<td>Inaccurate (n=17)</td>
<td>Accuracy (n=42)</td>
<td>Chi square</td>
<td>P value</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>----</td>
<td>------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>Diagnosis (%)</td>
<td>DDH</td>
<td>16</td>
<td>8 (50.00%)</td>
<td>8 (50.00%)</td>
<td>3.492</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>non-DDH</td>
<td>43</td>
<td>9 (20.93%)</td>
<td>34 (79.07%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex (%)</td>
<td>female</td>
<td>25</td>
<td>8 (32.00%)</td>
<td>17 (68.00%)</td>
<td>0.215</td>
<td>0.643</td>
</tr>
<tr>
<td></td>
<td>male</td>
<td>34</td>
<td>9 (26.47%)</td>
<td>25 (73.53%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI(%)</td>
<td>BMI(&lt;18.5)</td>
<td>6</td>
<td>2 (33.33%)</td>
<td>4 (66.67%)</td>
<td>1.317</td>
<td>0.725</td>
</tr>
<tr>
<td></td>
<td>BMI[18.5-23.9]</td>
<td>33</td>
<td>11 (33.33%)</td>
<td>22 (66.67%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI[24-27.9]</td>
<td>13</td>
<td>3 (23.08%)</td>
<td>10 (76.92%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BMI(≥28)</td>
<td>7</td>
<td>1 (14.29%)</td>
<td>6 (85.71%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

Preoperative planning for THA has evolved from the widely used 2D digital template to 3D software planning, with a gradual improvement in the accuracy of prediction. However, most of the software operation procedures of 3D planning are still complicated and time-consuming at present. In this study, the AI HIP was used for preoperative planning of patients undergoing primary THA. The accuracy of acetabular cup and femoral stem prediction was 74.58% (44 of 59) and 71.19% (42 of 59), respectively, and the average operation time was 3.91±0.64 min. Compared with the 3D mimics, the accuracy was similar, but the operation time was greatly shortened, making the operation more convenient. Compared with the traditional 2D digital template, its accuracy was much higher, with a slightly longer operation time than that of the digital template, but it can provide a 3D perspective that cannot be provided by the 2D digital template. In this study, AI technology was used for preoperative planning of THA for the first time, and was compared with the accuracy and time efficiency of 3D and 2D planning. Combined with the current era of big data, the clinical application value of this technology in preoperative planning was established.

In order to highlight the objectivity of AI HIP, 3D mimics and 2D digital templates were planned and compared. In previous studies on 3D planning, Inoue et al. [12] conducted a retrospective study on 65 hips of 57 patients with a total of 65 hips using 3D planning software-zed-hip. The results showed that the complete accuracy of the femoral stem was 65%, 98% in one size, and 92% in the acetabular cup, and 100% in one size. Similarly, using the 3D planning software-zed-hip, Erik et al. conducted a retrospective preoperative planning study on 116 patients. The results showed that the complete accuracy of the
femoral stem was 58.6%, 94% in one size, and 56.9% in the acetabular cup and 86.2% in one size [14]. The accuracy rate of 3D planning in the preoperative planning of complex hip joint diseases is also high. For example, Zeng et al. 3 D mimics was used to study the preoperative planning of 20 cases of DDH patients with high dislocation. The results showed that the complete accuracy rate of acetabular cup was 70%, and that of one size was 100%, while the results of the traditional 2D template showed that the complete accuracy rate of the acetabular cup was 25%, and that of one size was 45% [10]. Similarly, peihui et al. [15] also used mimics to plan the acetabular cup in 41 patients with 49 hip DDH. The results showed that the complete accuracy was 71%, and 100% within a size. It can be seen that the above 3D preoperative planning related research results show a high accuracy rate, and has a great advantage in the planning of complex hip diseases. This is because 3D planning software provides different 3D perspectives. In our study, the results of 3D mimics planning showed that the complete accuracy of acetabular cup was 71.19%, and that of one size was 93.22%. The complete accuracy of the femoral stem was 76.27%, and that of one size was 93.22%, which was similar to the previous 3D planning results. The accuracy of the AI HIP also reached the prediction level of this 3D software (cup: complete accuracy was 74.58%, in one size was 94.92%; stem complete accuracy was 71.19%, in one size 93.22%). In previous studies on 2D digital templates, the accuracy of prosthesis prediction was low due to the influence of many factors. Strøm et al. [16] used a double-blind method to perform preoperative planning for 41 patients. The results showed that the complete accuracy rate of acetabular cup was 7.3%, that of one size was 41%, that of femoral stem was 34%, and that of one size was 76%. Shaarani et al. [17] carried out preoperative planning for 100 patients with the same digital template. The results showed that the complete accuracy rate of the acetabular cup was 38%, one size was 80%, and the complete accuracy rate of the femoral stem was 36%. Since the adjacent model of the accolade stem was 0.5, the internal size of one was 75%. Schiffner et al. [14] compared 3D and 2D planning methods, and the results showed that the complete accuracy rate of 2D planning acetabular cup was 44.8%, 45.7% for femoral stem, 56.9% for 3D planning of acetabular cup, and 58.6% for femoral stem. The above studies on 3D and 2D showed that the accuracy of 3D planning was significantly higher than that of 2D planning. In our 2D digital template planning, the results showed that the complete accuracy of the acetabular cup was 42%, including one size was 75%, the femoral stem was 50%, and including one size was 77%. The results of this study were similar to those of previous relevant literature. All plannings were checked by two experienced surgeons. If there was a disagreement, the planning was repeated by all the three surveyors agreed upon. Therefore, our results are both effective and reliable.

The AI HIP was found in 10 consecutive planned hip joints that the prediction of acetabular cup and femoral stem fluctuated in one size. In the first 10 patients planning, the accuracy of the acetabular cup and femoral stem was 70%. In the planning of 10-20 patients, the accuracy rates of acetabular cup and femoral stem were 80% and 70%, respectively. It shows that AI HIP was relatively stable, and there was no obvious deviation in the automatic learning in the early stage, so the planning could continue without elimination. In the whole study, there were 11 hips in one size, 2 hips in two sizes, and 1 hip in more than three sizes. In two hips with a difference of 2 sizes, one was osteoarthritis due to malunion of a chronic femoral neck fracture. It became the shallow acetabulum fossa edge on the part of the deformation,
when the acetabulum cup matches 54 mm bone coverage rate of 87%, while the actual intraoperative application of 50 mm acetabulum cup, the reason may be that AI planned the position and size of acetabular cups in the ideal state. However, when the acetabular cup was ground to 50mm during the operation, it was found that the coverage rate of the acetabular cup was sufficient and the stability of the cup was satisfactory. So, the grinding was not deepened. The other was the same AI software that planned a 54mm acetabular cup, and a 50mm acetabular cup was used in the operation. The possible reason was that in patients with ankylosing spondylitis because of ankylosis of the hip joint, the acetabulum was fused with the femoral head, and the software could not accurately segment the acetabulum and femoral head, resulting in a large deviation. Thus, the application of AI HIP in preoperative planning for serious AS needs further investigation. There was one case with a difference of more than 3 sizes. This patient had severe hip osteoarthritis with osteophytes around the hip joint. However, the software matched a 52mm acetabular cup according to the true anterior and posterior diameter of the acetabulum, including the upper and lower diameter of the acetabulum. Perhaps the surrounding osteophytes interfered with the software in recognizing the acetabulum morphology. This led to a large error in the planning results. In the planning of the femoral stem, there were 10 hips in one size, 3 hips in two sizes, and 1 hip in three sizes. In 3 hips with a difference of 2 sizes, 2 hips were DDH, which may be due to: the variation of femoral neck anteversion angle and the increase of neck stem angle, the enlargement and upward movement of the femoral anterior arch, the enlargement of the anteroposterior diameter of the proximal femoral medullary cavity, and the reduction of transverse diameter. This may lead to the deviation of AI in identifying: the epiphysis of the femoral shaft, the medial and lateral cortex of the distal femur, resulting in large planning errors, which requires AI HIP to make sufficient planning after a large number of DDH cases. It can determine the femoral variation of DDH, so that it can be more accurate in DDH. In the case of 3 sizes, the anterior femoral arch of the patient was significantly enlarged, which may have led to a bias towards matching the smaller femur stem during AI planning.

Our study also compared the operation time of the three planning methods. The results showed that the operation time of AI HIP was shorter than that of 3D mimics. The study of preoperative planning of the operation time has also been reported before, Erik Schiffner et al. [14] compared the operation time of 3D and 2D for the first time. The results showed that the average operation time of the 2D template was 12 minutes (range: 8-23 min), while that of the 3D template was 17 minutes (range: 10-25 min). The operation time of 2D was faster than that of 3D, with an average of 5 minutes interval. When sariali et al. used hip-plan 3D planning software, the operation time was 10-15minutes. Similarly, Inoue [12] et al. used a zed hip for preoperative planning, and the operation time was also 10-15 minutes. The reason for which their 3D planning time was faster than our 3D mimics planning time was because the algorithms of the pelvic plane, femoral morphology, and anatomical markers in their 3D software were automatic, while our 3D mimics was manual in pelvic segmentation, modeling, prosthesis implantation, and other operations. However, the AI HIP in our study was 3-5 times faster than the 3D software planning reported before, and 10 times faster than 3D mimics. This is because the software uses 3D segmentation neural network and 3D anatomical recognition neural network technology, which can quickly identify, segment, correct, and measurement by artificial intelligence, greatly shortens the planning time. Petretta [7] compared the time
efficiency of the acetate template and digital template, and the results obtained showed that the
operation time of the digital template was 154 s (range: 73-343 s). In our study, the operation time of the
2D digital template was 2.96±0.48 minutes, which was similar to the results of petretta.

In the analysis of factors influencing the accuracy of AI HIP, we found that sex had no influence on the
software. Holzer et al., studied the accuracy of digital template in cementless prosthesis in 2D
preoperative planning. The results obtained also showed that sex had no influence on the accuracy of
digital template prediction [8] because of little variation and limited differences in sex in the amplification
rate of markers. [22]. Previous studies on the influencing factors of BMI on the accuracy of 3D planning
also found that obese patients were not affected by the accuracy of CT navigation to locate the
acetabular cup[23]. In our study, AI HIP adds AI technology based on the CT data of patients. It is also
based on the CT data for 3D planning, which is not affected by the magnification. Therefore, the accuracy
of its prediction is not affected by BMI. In the analysis of hip dysplasia, we found that the accuracy of AI
HIP in the DDH group was lower than that in the non-DDH group, and the difference was statistically
significant. However, there was no significant difference in the accuracy between the DDH and non-DDH
groups in the planning of the femoral stem. The main reason for these results was that 16 hips in the
DDH group were Crowe I-II type, and the degree of femoral deformity variation was small. So, there was
no significant difference in accuracy between the DDH and non-DDH groups. On the acetabular side,
although the deformity variation is relatively light, which is not as severe as Crowe III-IV high dislocation
deformity, AI HIP was in situ reconstruction during the planning of acetabular cup, without proper upward
or inward movement according to the degree of variation of acetabular deformity, and the placement
angle is 40° abduction angle and 20° anteversion angle. In the actual operation, 13 (81.25%) of the 16
hips in the DDH group had the acetabular cup appropriately moved upward or inward in order to achieve
greater bone coverage. This may be the reason why the software was inaccurate in predicting acetabular
cups in patients with DDH. Therefore, the software needs to expand the planning of the number of cases
of each type of DDH, continuously strengthen the deep learning of the angle and position of the
placement of each type of DDH acetabular cup, so as to master the placement rules of each type of DDH
prosthesis, and achieve accurate planning in DDH planning.

Innovation:

Based on AI technology and big data, compared with previous 2D and 3D studies, AI HIP has the
following innovations: (1) Speed: one-button intelligent operation, simple operation; (2) Visualization: 3D
visualized image output is immersive, providing realistic 3D visual anatomy; (3) Pelvic correction:
Automatically corrects the position of the pelvis and lower limbs, without constant manual adjustment;
(4) automatic planning: artificial intelligence planning can avoid planning deviation caused by personal
experience, while the results of other 2D template methods and 3D template measurements are related to
the experience level of the surveyors, and the more skilled the operation, the more accurate the results; (5)
output scheme: generate a multi-perspective observation planning scheme, which is easy to realize in
operation; (6) continuing education; further support and improvement of continuing education to shorten
the doctor's learning curve. At the same time, in this study, in order to avoid surgeon selection deviation, the blind method was adopted for the surgeon because if the surgeon participated in preoperative planning, the surgeon might change the grinding technique and insert a larger or smaller part by removing more or less bones when necessary, which will lead to errors.

**Limitations:**

(1) The prosthesis used in this study was limited to depuy products, and the range of prosthesis selection was small. (2) The sample size of the study was small, and the influencing factors of diseases were only divided into DDH and non-DDH groups, without further subdividing the difference in accuracy of the software in different diseases. (3) Only the accuracy of prosthesis size was evaluated, but important parameters such as: rotation center, offset, and leg length discrepancy were not further evaluated. (4) Because of the need for CT scanning, the X-ray radiation exposure of patients was large, and the economic cost was also high. (5) CT data need to be copied, whether the data transmission process can be optimized or needs no further research. However, this is not the scope of this experiment. Next, we will continue to: expand the cases, further study the accuracy of AI HIP in different diseases, evaluate its reproducibility in biomechanical planning and its impact on postoperative clinical efficacy.

**Conclusions**

AI HIP showed excellent reliability for component size in THA. Acetabular dysplasia may affect the accuracy of AI HIP templating. Therefore, it is necessary to continue to expand the number of DDH cases and continuously strengthen the AI learning process.

**Abbreviations**

AI: artificial intelligence; THA: total hip arthroplasty; 2D: two-dimensional; 3D: three-dimensional; BMI: body mass index; DDH: developmental dysplasia of the hip; OA: osteoarthritis; ON: osteonecrosis; AS: ankylosing spondylitis; RA: rheumatoid arthritis.

**Declarations**

**Acknowledgements**

We would like to thank Depuy and Changmugu for their assistance in providing AI HIP software. We would like to thank Editage (www.editage.cn) for English language editing.

**Authors’ contributions**
WXJ, BYF and CY collected the data. HJB performed all preoperative planning and was the major contributors in writing the manuscript. HGX and ZC analyzed and checked all preoperative planning results. HD was responsible for analyzing the data. CDZ and ZC reviewed the manuscript. All authors read and approved the final manuscript.

**Fundings**

This work were supported by grants obtained from the Science and Technology Project of Guangdong Province (No. 2016B090916003), the Natural Science Foundation of Guangdong Province (No. 2018030310355).

**Availability of data and materials**

The datasets during and/or analyzed during the current study available from the corresponding author on reasonable request.

**Ethics approval and consent to participate**

The study was approved by the local institutional review board.

**Consent for publication**

Written consents for publication were obtained from all study participants.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' information**

1Department of Orthopedics, Orthopedic Hospital of Guangdong Province, The Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China.

2Department of quality management and evaluation, Orthopedic Hospital of Guangdong Province, The Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China.

**References**


