



Retrospective Study

# HIPPO artificial intelligence: Correlating automated radiographic femoroacetabular measurements with patient-reported outcomes in developmental hip dysplasia

Ahmed Alshaikhsalama, Holden Archer, Yin Xi, Richard Ljuhar, Joel E Wells, Avneesh Chhabra

**Specialty type:** Medicine, research and experimental

**Provenance and peer review:** Invited article; Externally peer reviewed.

**Peer-review model:** Single blind

**Peer-review report's classification**

**Scientific Quality:** Grade A, Grade B

**Novelty:** Grade A, Grade B

**Creativity or Innovation:** Grade A, Grade B

**Scientific Significance:** Grade A, Grade B

**P-Reviewer:** Chouffani El Fassi S; Li J

**Received:** July 20, 2024

**Revised:** September 23, 2024

**Accepted:** October 24, 2024

**Published online:** December 20, 2024

**Processing time:** 102 Days and 15.2 Hours



**Ahmed Alshaikhsalama, Holden Archer,** Department of Radiology, University of Texas Southwestern, Dallas, TX 75390, United States

**Yin Xi, Avneesh Chhabra,** Department of Radiology, University of Texas Southwestern Medical Center, Dallas, TX 75235, United States

**Richard Ljuhar,** Department of Radiology, Image Biopsy, Vienna 1190, Austria

**Joel E Wells,** Department of Orthopedic Surgery, Baylor Scott and White, Dallas, TX 75235, United States

**Corresponding author:** Ahmed Alshaikhsalama, BSc, Research Associate, Department of Radiology, University of Texas Southwestern, 5323 Harry Hines Blvd, Dallas, TX 75390, United States. [ahmed.alshaikhsalama@utsouthwestern.edu](mailto:ahmed.alshaikhsalama@utsouthwestern.edu)

## Abstract

### BACKGROUND

Hip dysplasia (HD) is characterized by insufficient acetabular coverage of the femoral head, leading to a predisposition for osteoarthritis. While radiographic measurements such as the lateral center edge angle (LCEA) and Tönnis angle are essential in evaluating HD severity, patient-reported outcome measures (PROMs) offer insights into the subjective health impact on patients.

### AIM

To investigate the correlations between machine-learning automated and manual radiographic measurements of HD and PROMs with the hypothesis that artificial intelligence (AI)-generated HD measurements indicating less severe dysplasia correlate with better PROMs.

### METHODS

Retrospective study evaluating 256 hips from 130 HD patients from a hip preservation clinic database. Manual and AI-derived radiographic measurements were collected and PROMs such as the Harris hip score (HHS), international hip outcome tool (iHOT-12), short form (SF) 12 (SF-12), and Visual Analogue Scale of the European Quality of Life Group survey were correlated using Spearman's rank-order correlation.

## RESULTS

The median patient age was 28.6 years (range 15.7-62.3 years) with 82.3% of patients being women and 17.7% being men. The median interpretation time for manual readers and AI ranged between 4-12 minutes per patient and 31 seconds, respectively. Manual measurements exhibited weak correlations with HHS, including LCEA ( $r = 0.18$ ) and Tönnis angle ( $r = -0.24$ ). AI-derived metrics showed similar weak correlations, with the most significant being Caput-Collum-Diaphyseal (CCD) with iHOT-12 at  $r = -0.25$  ( $P = 0.042$ ) and CCD with SF-12 at  $r = 0.25$  ( $P = 0.048$ ). Other measured correlations were not significant ( $P > 0.05$ ).

## CONCLUSION

This study suggests AI can aid in HD assessment, but weak PROM correlations highlight their continued importance in predicting subjective health and outcomes, complementing AI-derived measurements in HD management.

**Key Words:** Hip dysplasia; Patient reported outcome measures; Deep-learning; Artificial intelligence; Radiographs; Lateral center edge angle

©The Author(s) 2024. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** In this study, we compared an artificial intelligence (AI) tool measuring anteroposterior hip radiographs against manual readers for assessing hip dysplasia (HD) associations with patient-reported outcome measures (PROMs). The AI tool, HIPPO, efficiently generated radiographic measurements but showed poor correlations with PROMs, highlighting its current limitations in predicting clinical outcomes solely from radiological data. This indicates that while AI can aid radiographic assessments, PROMs remain crucial for capturing subjective patient experiences. The findings underscore the importance of integrating PROMs as an additional element in the clinical decision-making processes for HD, while also incorporating efficient radiographic assessment by AI tools.

**Citation:** Alshaikhsalama A, Archer H, Xi Y, Ljuhar R, Wells JE, Chhabra A. HIPPO artificial intelligence: Correlating automated radiographic femoroacetabular measurements with patient-reported outcomes in developmental hip dysplasia. *World J Exp Med* 2024; 14(4): 99359

**URL:** <https://www.wjgnet.com/2220-315x/full/v14/i4/99359.htm>

**DOI:** <https://dx.doi.org/10.5493/wjem.v14.i4.99359>

## INTRODUCTION

Acetabular or hip dysplasia (HD) is a developmental condition that is characterized by a shallow or upsloping acetabulum that can be accompanied by femoral head incongruency[1]. HD often presents in the pediatric and adult population with symptoms of hip pain and/or instability. When left untreated, it can lead to hip osteoarthritis (OA) due to stress overload, shear forces, and improper mechanics progressively affecting joint cartilage[2]. Several conservative and surgical treatment options currently exist; among them, the most used modalities include physical therapy and lifestyle modifications, periacetabular osteotomy, hip arthroscopy, and total hip arthroplasty. The treatment modality chosen depends upon the time of discovery, symptom severity, and status of the hip labrum and cartilage, and functional disability[3-5].

Hip radiographs are the current gold standard for the initial screening and assessment of HD[6]. There are a multitude of validated diagnostic radiographic measurements employed to assist the diagnosis of HD. Among them, lateral center edge angle (LCEA) is most commonly used, as measured on a standing anteroposterior (AP) pelvis radiograph[7]. Additionally, the Tönnis angle and extrusion index are also commonly used in clinical practice[8]. Following radiographic assessment, advanced imaging such as magnetic resonance imaging or computed tomography can be used for pre-operative planning and further assessment of the health of the labrum or hyaline cartilage[9].

While a diagnosis of HD is established by a combination of clinical presentation, examination findings and radiographic measurements, patient-reported outcome measures (PROMs) are equally important to illustrate the perception of patients' subjective hip health status[10]. These are gleaned from different surveys administered at the time of clinical presentation, such as the Harris hip score (HHS), international hip outcome tool (iHOT-12), Visual Analogue Scale (VAS) for Pain, VAS of the European Quality of Life Group (EQ-VAS) (health status), and short form (SF) 12 (SF-12) (quality of life), among others. Each patient reported outcome survey provides a different evaluation of the patient's condition. For instance, the HHS is a reliable indicator for patient function, while iHOT-12 provides a good indication for quality-of-life changes[11-13].

PROMs have become increasingly important in evaluating indications for treatment and prognosis for HD patients[14-16]. Despite their common use in the clinical evaluation of patients with HD and pain, the International Hip-related Pain Research Network meeting in 2018 ruled that more studies are needed to further evaluate the usefulness of PROMS[17].

Thus, it is important to examine the relationships between validated radiographic HD measurements and PROMs[11]. One prior study evaluated the by Takegami *et al*[18] evaluated the relationship between manual individual radiographic parameters with the patient-reported outcome measurements in Japanese patients. However, it is time consuming to routinely measure the above-described parameters, let alone control for the associated inherent reader variance and need to remember how to obtain such parameters. If these measurements could be automatically produced by machine learning using artificial intelligence (AI), the clinical note and/or radiographic interpretation report could be auto-populated. In addition, the correlations between radiographic parameters and PROMs can be studied in a more standardized manner and for longitudinal data collection. To that end, AP radiographic measurements can be auto-evaluated by HIPPO software, which is a validated AI hip measurement tool validated in a European study and Conformite Europeenne certified [ImageBiopsy Lab Inc. (Vienna, Austria)][19]. Yet, it is not known how these standardized deep-learning software generated measurements obtained in the United States population correlate with their PROMs data. Additionally, it is not known if a validated AI tool can assist in predicting PROMs data and providing comprehensive evaluation for HD patients.

Our hypothesis was that AI-generated HD measurements indicating less severe dysplasia correlate with better PROMs. Thus, the aim was to assess the correlation between AI-derived hip measurement and initial PROMs in a consecutive series of patients. This is the first study to evaluate manual and AI measures of radiographs in patients with HD and associate radiographic findings with preoperative PROMs data.

The SF-12 questionnaire is a short form of the SF-36, where a patient provides a subjective assessment of their own health status and its influence on their respective lifestyle; it reports on psychological features of the condition[20]. Another tool to assess patient outcomes is the iHOT-12 adapted from the 33-question survey that defines changes in quality of life due to hip pathology[21]. An additional meaningful measure is the EQ-VAS-a visual analog scale from 0 to 100-through which the general overall health status of the patient can be observed[22]. In terms of radiographic assessment, multiple parameters provide an indication of the hip's mechanical profile. For instance, the Caput-Collum-Diaphyseal (CCD) angle between the femoral neck and shaft axes contributes to the evaluation of femoral alignment[23]. Additionally, the Sharp's angle, LCEA, Tönnis angle, and the extrusion index represent important radiographic parameters that help assess acetabular coverage, which is important in assessing the severity of dysplasia[23,24]. These PROMs, in concert with the described standardized radiographic measurements, enable the clinician to have a comprehensive understanding of the severity and impact of HD on patients.

## MATERIALS AND METHODS

Institutional Review Board approval was received for retrospective use of a longitudinally gathered patient registry data and surveys. Anonymous survey data involving PROMs was collected in our institutional hip preservation practice. All Health Insurance Portability and Accountability Act regulations were followed.

### Patients

Using our anonymized electronic database of patients who visited the institutional hip preservation clinic, we identified 325 hips from 276 patients with a complete radiographic series from December 2016 to December 2021. Each patient had a reference final HD diagnosis based on consensus radiographic opinions of an independent fellowship trained musculoskeletal radiologist and hip preservation surgeon using the 4-view radiographic series (AP pelvis, 45° Dunn, Frog-leg lateral, and false profile views) and clinical findings. Only patients with a concordant final diagnosis of HD were included in this study, resulting in 256 hips from 130 patients. Six of the 136 patients did not return an output from HIPPO (Figure 1). The hips with prior surgical interventions or avascular necrosis were excluded. Patient demographic data including age, gender, and body mass index (BMI) were extracted from the electronic health records. Additionally, dates of the patient's first office visit and survey, along with the dates and details of any surgeries were collected. The surveys were obtained at the time of the initial clinic visit when the radiograph was obtained to avoid delay between imaging and initial PROM survey.

### PROMs

The patients were surveyed at the time of their initial office visit, which included HHS, iHOT-12, SF-12, and EQ-VAS as shown in Table 1. Survey data was obtained using an online REDCap form and was retrieved into an excel document for each of the included deidentified study patients. Each survey result was manually calculated and normalized to 100% by two medical students under the training and supervision of the senior orthopedic hip specialist.

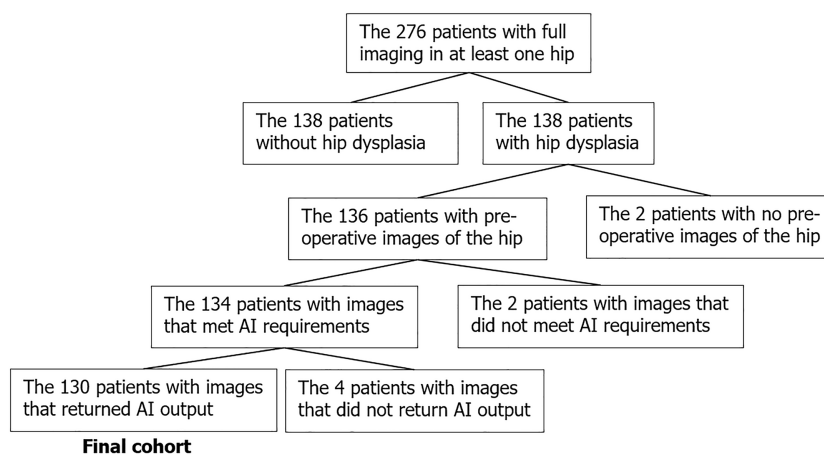
### Manual measurements

Tönnis grade of hip OA was evaluated in all cases by the senior orthopedic surgeon. Manual HD measurements were obtained as a control for the AI measurements. Measurements were taken for each patient by three readers under the supervision and training of a senior musculoskeletal (MSK) radiologist. The three readers underwent extensive training under the MSK radiologist and were assessed for accuracy on a series of training images before obtaining the measurements for the study. The study measurements were then averaged and correlated with PROMs (Table 1)[16,20-22]. Time required to assess these measurements was recorded using a stopwatch from the time images were loaded on IntelliSpace Picture Archiving and Communication System (Philips, Best, Netherlands) to completion of the reads using a built-in measurement tool. Measurement data from the AI algorithm and manual measurements with their detailed inter-reader and inter-modality correlations between manual measurements and AI was published and showed good to

**Table 1 Patient-reported outcome measures surveys**

Survey	Description
HHS	The HHS is a joint-specific 10-question survey evaluating hip function. The survey parameters include- ability to climb stairs, take public transport, and put on shoes and socks. The test has been shown to have strong construct validity, and thus would be appropriate as a comprehensive assessment of the affected joint's impact on the patient[16]
SF-12	The SF-12 survey, which was adapted from the SF-36 survey, assesses the patient's view of their own health and how it relates to their lifestyle. It includes questions, such as asking the patient if they achieved as much as they have liked and whether they have felt calm and peaceful. Thus, the SF-12 can provide insight into the psychological aspect of the patient's condition[20]
IHOT-12	The iHOT-12 is a 12-question survey adapted from the 33-question survey. The survey evaluates quality of life changes[21]
EQ-VAS	EQ-VAS is a scale from 0 (worst health) to 100 (best health) that allows the patient to indicate their overall perspective of their health state[22]

IHOT: International hip outcome tool; EQ-VAS: Visual Analogue Scale of the European Quality of Life Group; HHS: Harris hip score; SF: Short form.



**Figure 1** It Shows the final cohort for hip dysplasia patients with patient-reported outcome measures data and compatible imaging.

excellent inter-method reliability for common HD landmarks including LCEA and Tönnis angle[19].

### AI measurement tool–HIPPO

‘HIPPO’ is an AI deep-learning software [ImageBiopsy Lab Inc. (Vienna, Austria)] that automatically locates anatomical landmarks on AP full leg standing radiographs. Using these landmarks, the tool measures various radiographic parameters. These parameters are LCEA, Tönnis Angle, Sharp Angle, CCD angle and pelvic obliquity (Table 2 and Figure 2)[12,23,24]. The software accepts images in Digital Imaging and Communications in Medicine (DICOM) format and returns a DICOM compatible AI report. When the software returns an error report or does not return a report at all, a software failure is indicated. A software failure could be due to errors in the software itself or anatomical subtleties in the radiograph that heavily affected how the software interprets the images. All images in the study were securely transferred to the picture archiving and communication system server at our institution, and from there were pushed to a local installation of the AI software. Measurements were then downloaded onto an excel document after being processed through the software (Windows 11, Microsoft, Redmond, WA). In our study, the median HIPPO reading time per patient was 41 seconds.

### Statistical analysis

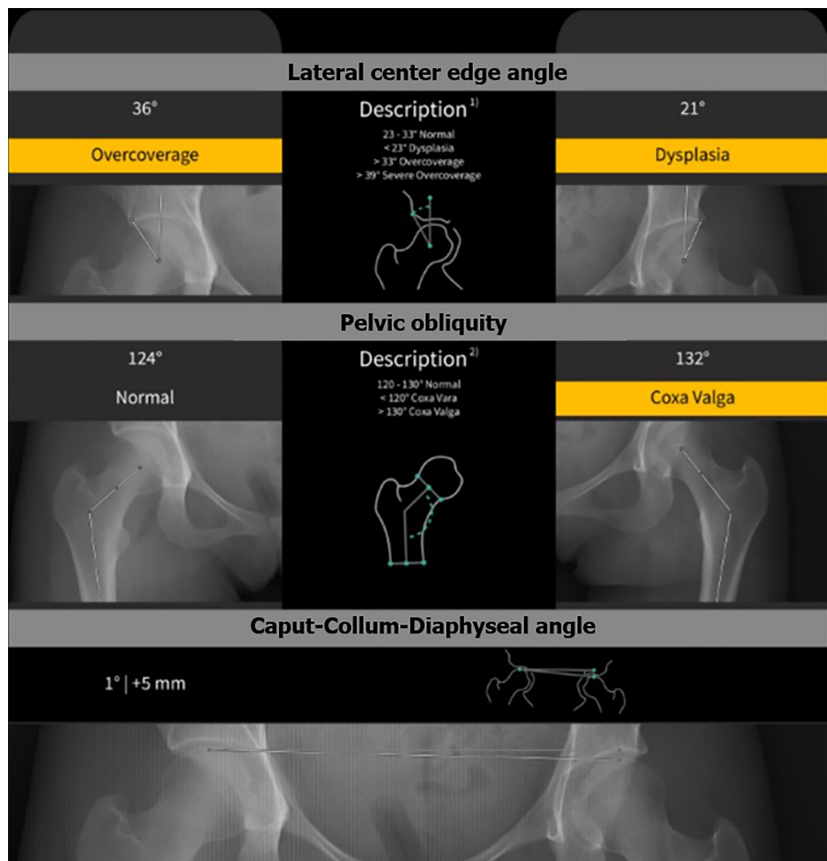
Descriptive statistics were calculated for patient demographics. All hip measurements were on per-hip level while PROMs except HHS were on per-patient level. Therefore, one hip from each patient was selected when comparing hip measurements to iHOT-12, SF-12, and EQ-VAS. The hip with the worst mean LCEA score from the 3 readers was selected. Correlations between hip measurements and HHS were calculated on the same selected hips. Spearman’s rank correlation coefficients were reported with corresponding 95%CI. Hypothesis tests for non-zero correlation were conducted at a 0.05 significance level. *P*-values were adjusted for false discovery rate *via* the Benjamini and Hochberg method for each PROM. Correlation coefficients were interpreted as negligible: 0-0.1, weak: 0.1-0.39, moderate 0.4-0.69, strong: 0.7-0.89 and very strong: 0.9-1[25]. With 80% power to detect a correlation of at least 0.26 at 0.05 significance level, the study needed 130 patients before adjustments for multiple comparisons.



**Table 2 HIPPO radiographic hip measurements and landmarks**

Measurement	Description
CCD	The CCD angle was measured as the angle between the femur neck and shaft axis[23]
Pelvic obliquity	The pelvic obliquity was measured with a tangential line from the apex of the femoral heads and a line parallel to the horizontal plane as in Figure 1
Sharps angle	Sharp's angle was measured with a line connecting the inferior ischial tuberosities and a line connecting the lower medial edge of the acetabular teardrop and the lateral edge of acetabular sourcil[23]
LCEA	The LCEA was measured as a perpendicular line through the center of the femoral head perpendicular to the acetabular tuberosities and the angle between a line from the center of the femoral head to the lateral acetabular sourcil[24]
Extrusion index	The extrusion index was measured by the difference of medial and lateral femoral head and the lateral edge sourcil with three vertical lines at edge aspect. The femoral head coverage was represented by the percentage of femoral head covered: Lateral femoral head to lateral edge sourcil distance minus the total horizontal head diameter[23]
Tönnis angle	The Tönnis angle was measured as the angle between a line connecting the inferior and lateral aspects of the acetabular sourcil and a line connecting the inferior portion of the ischial tuberosities[12]

CCD: Caput-Collum-Diaphyseal; LCEA: Lateral center edge angle.



**Figure 2** HIPPO Digital Imaging and Communications in Medicine output showing lateral center edge angle, Caput-Collum-Diaphyseal angle, and pelvic obliquity as measured by HIPPO on anteroposterior radiograph.

## RESULTS

### Patients

Descriptive statistics were calculated for appropriate demographic factors. The median patient age was 28.6 years with a maximum of 62.3 years and a minimum of 15.7 years. The 82.3% of patients were women and 17.7% were men. The BMI ranged from 17 kg/m<sup>2</sup> to 38 kg/m<sup>2</sup>, with 24 kg/m<sup>2</sup> as the median. An orthopedic surgeon classified the hips according to the Tönnis grade. The median Tönnis grade was 0 with the majority (204 hips, 79.7%) having Tönnis grade 0, 51 hips (19.9%) with Tönnis grade 1, and 1 hip (0.4%) with Tönnis grade 2.

### Manual measurements

Measurement data from the AI algorithm and manual measurements showed good to excellent inter-method reliability for common HD landmarks including LCEA and Tönnis angle. The median read time for manual readers ranged between 4 and 12 minutes per patient[19].

### Manual hip measurements vs PROMs

The largest estimated correlation coefficients were between LCEA and HHS [0.18 (0.00, 0.35)], Tönnis Angle and HHS [-0.24 (-0.40, -0.06)], CCD and SF-12 [0.19, (0.01, 0.36)], and CCD and iHOT-12-12 [-0.19, (-0.36, 0.00)]; however, these weak correlations were not significant at a 0.05 level after adjustment for multiple comparisons (Table 3). No other significant correlation was observed between the remaining manual measurements and PROMs. A scatter plot is shown in Figure 3A.

### AI hip measurements vs PROMs

CCD were significantly correlated with iHOT-12 and SF12, but the correlation strength was weak [CCD *vs* iHOT-12: -0.25 (-0.42, -0.07),  $P_{adj} = 0.042$ ; CCD *vs* SF12: 0.25 (0.07, 0.42),  $P_{adj} = 0.048$ ]. Other notable correlations of similar magnitude were estimated for Obliquity and EQ-VAS [-0.22, (-0.39, -0.4)], as well as Tönnis angle and HHS [-0.20, (-0.36, -0.02)]; however, these estimates were not significant at a 0.05 level after adjustment for multiple comparisons (Table 4 and Figure 3B).

### PROMs

HD patients before intervention had an average survey scores of 69% EQ-VAS suggesting moderate pain[26] and 63% SF-12, which is slightly above the depression threshold[27]. They also had 61% iHOT-12, which is nominally above the acceptable symptom threshold (pass) of 59% indicating the patients had a greatly affected quality of life[28], and 62% HHS, which is poor function as defined by the standard less than < 70%[29].

## DISCUSSION

This study aimed to evaluate the correlation between AI-generated radiographic measurements and PROMs in individuals with HD. Our findings suggest that while there is a presence of weak correlations between certain AI-derived radiographic measurements and PROMs, these relationships did not achieve statistical significance after adjustments for multiple comparisons. This indicates that the current capacity of AI, specifically the HIPPO deep-learning software, to predict clinical outcomes based on radiological data is limited, although not entirely negligible.

HIPPO is a novel tool for acquiring rapid hip measurements, successfully processing most cases with notable efficiency as reported previously[30]. Where manual readers required a median time of 6 minutes and 48 seconds per hip, and trained radiologists require on average 83 seconds per AP hip radiograph, the AI completed the same task in an average of 41 seconds, highlighting a significant reduction in time and cost per radiograph[19,30]. In this study, HIPPO AI found a significant association between the CCD angle and iHOT-12/SF-12 PROMs compared to manual readers. An elevated CCD angle (Coxa Valga) has been associated with HD, although it is a less commonly used measurement diagnostically [6,31]. While the exact reason for this significant association is not known, the authors hypothesize that the difficulty of measuring CCD among manual readers compared to a standardized AI tool introduced sufficient variation to prevent an observed association[32]. These findings further highlight the importance of standardization in assessment and interpretation of radiographic measurements. The results of this study differ from those of Takegami *et al*[18], where the LCEA angle in 108 Japanese HD patients was independently associated with the Japanese Orthopaedic Association's hip disease questionnaire. However, the end point PROMs examined in our study were different and applied to a heterogeneous United States population, limiting direct comparison. Despite the potential for AI to streamline clinical workflow, our study highlights the difficulty and current unfeasibility of correlating radiographic findings with patient-centric outcomes such as PROMs. Although HIPPO is efficient at measuring, it may require more training to recognize patterns that better match patients experience. This highlights an area where AI can develop to become more clinically meaningful.

An additional consideration is our patient cohort. Overall, the study's patient cohort was symptomatic, presenting with moderate pain, slightly above the depression threshold, and poor functional scores as per EQ-VAS, SF-12, iHOT-12, and HHS, respectively[26-29]. The homogeneity of this group may have diluted the potential to discern a stronger correlation between radiographic measurements and PROMs. Including asymptomatic individuals in future studies may provide a broader spectrum of disease and potentially unveil more defined associations.

It is important to note the subjective nature of PROMs and their potential to be affected by factors beyond the HD diagnosis. For instance, while HHS mainly measures hip function, SF-12 encompasses wider quality of life and mental health parameters, which can be affected by multiple socio-economic and demographic factors[33]. Similarly, individual variability in physical fitness and factors such as hamstring strength play a role in hip stability and perceived symptoms and functionality, contributing to an observed variability in PROMs that may make it difficult to correlate any radiographic measurement, no matter the tool used[34,35].

The weak correlations observed challenge our initial hypothesis that improvements in HD radiographic measures would linearly correlate with better PROMs. The authors do not believe that these weak correlations are due to inaccuracies in the AI measurement tool, which was previously validated by Archer *et al*[19] revealing moderate to strong associations with trained manual readers. Additionally, the vast majority of observed correlations were nonsignificant and contained similar results to the manual readers, with exception of CCA angle and certain PROMs on AI reads, thus

**Table 3 Spearman correlation between manual hip measurements and various patient-reported outcome measures surveys**

Patient-reported outcome measures	Hip measures	Estimate	Lower 95%CI	Upper 95%CI	Raw P value	Adjusted P value
Visual Analogue Scale of the European Quality of Life Group	CCD	0.07	-0.11	0.25	0.450	0.802
	Extrusion index	0.02	-0.16	0.20	0.823	0.823
	LCEA	-0.04	-0.22	0.15	0.688	0.823
	Obliquity	-0.17	-0.34	0.01	0.063	0.378
	Sharp	0.06	-0.13	0.24	0.535	0.802
	Tönnis	-0.08	-0.25	0.11	0.419	0.802
Harris hip score	CCD	0.02	-0.16	0.20	0.791	0.791
	Extrusion index	-0.14	-0.31	0.04	0.122	0.183
	LCEA	0.18	0.00	0.35	0.049	0.147
	Obliquity	-0.16	-0.33	0.02	0.081	0.162
	Sharp	-0.06	-0.24	0.12	0.493	0.592
	Tönnis	-0.24	-0.40	-0.06	0.009	0.054
International hip outcome tool	CCD	-0.19	-0.36	0.00	0.045	0.270
	Extrusion index	-0.03	-0.21	0.16	0.764	0.999
	LCEA	0.00	-0.18	0.18	0.999	0.999
	Obliquity	0.13	-0.06	0.30	0.183	0.549
	Sharp	0.00	-0.18	0.18	0.998	0.999
	Tönnis	0.07	-0.12	0.25	0.469	0.938
Short form 12	CCD	0.19	0.01	0.36	0.042	0.252
	Extrusion index	0.03	-0.16	0.21	0.778	0.870
	LCEA	-0.03	-0.22	0.15	0.720	0.870
	Obliquity	-0.13	-0.30	0.06	0.186	0.558
	Sharp	0.06	-0.13	0.24	0.530	0.870
	Tönnis	-0.02	-0.20	0.17	0.870	0.870

CCD: Caput-Collum-Diaphyseal; LCEA: Lateral center edge angle.

suggesting a similar radiographic accuracy between groups as previously described. These results call into question the clinical utility of radiographic measurements alone in predicting patient-reported outcomes and highlights the complexity of HD as a disease entity. While AI can rapidly provide quantitative data valuable for initial screenings and monitoring disease progression, it should complement-not replace-PROMs, which encapsulate the patient's subjective experience and the functional impact of the disease. PROMs remain essential for capturing the holistic impact on quality of life, guiding more personalized treatment approaches. Therefore, clinicians are encouraged to use various means of information-gathering including the use of PROMs. They capture a spectrum of patient experiences and outcomes that are not obvious through radiographic data, reinforcing their role in comprehensive care for patients with HD.

Our study has several limitations. The gender distribution in our study was predominantly female, reflecting the higher incidence of HD in women[36]. This distribution may influence the correlations observed and thus may not be generalizable to a male population. Additionally, most participants were middle-aged adults, so our results might not reflect the bone density and joint health variations found in older patients, and thus may affect the generalizability of this study[37]. Finally, the manual measurements, while performed by medical students under the supervision of an MSK radiologist, are not immune to human error. Anatomical variability might have led to inaccuracies; however, extensive training aimed to mitigate such errors, and their impact on the study's validity is considered minimal. Future studies should also incorporate prospective clinical validation studies to assess AI tools against traditional radiographic measurements, post-implementation in patient care settings. Additionally, randomized controlled trials comparing patient outcomes using AI-derived data with those using manual radiographic assessments are critical to establish the

**Table 4 Spearman correlation between artificial intelligence hip measurements and various patient-reported outcome measures surveys**

Patient-reported outcome measures	Hip measures	Estimate	Lower 95%CI	Upper 95%CI	Raw P value	Adjusted P value
Visual Analogue Scale of the European Quality of Life Group	CCD	0.09	-0.09	0.27	0.330	0.509
	Extrusion index	0.07	-0.11	0.25	0.424	0.509
	LCEA	-0.09	-0.27	0.10	0.354	0.509
	Obliquity	-0.22	-0.39	-0.04	0.015	0.090
	Sharp	0.06	-0.13	0.23	0.557	0.557
	Tönnis	-0.11	-0.29	0.07	0.235	0.509
Harris hip score	CCD	0.11	-0.07	0.28	0.238	0.286
	Extrusion index	-0.15	-0.32	0.03	0.112	0.255
	LCEA	0.13	-0.05	0.30	0.170	0.255
	Obliquity	-0.13	-0.30	0.05	0.154	0.255
	Sharp	-0.03	-0.21	0.15	0.723	0.723
	Tönnis	-0.20	-0.36	-0.02	0.033	0.198
International hip outcome tool	CCD	-0.25	-0.42	-0.07	0.007	0.042 <sup>a</sup>
	Extrusion index	0.03	-0.15	0.22	0.718	0.718
	LCEA	-0.05	-0.23	0.14	0.608	0.718
	Obliquity	0.17	-0.02	0.34	0.079	0.237
	Sharp	0.04	-0.15	0.22	0.703	0.718
	Tönnis	0.05	-0.13	0.24	0.565	0.718
Short form 12	CCD	0.25	0.07	0.42	0.008	0.048 <sup>c</sup>
	Extrusion index	0.00	-0.18	0.19	0.972	0.972
	LCEA	-0.07	-0.25	0.12	0.476	0.714
	Obliquity	-0.16	-0.34	0.02	0.088	0.264
	Sharp	0.08	-0.10	0.26	0.385	0.714
	Tönnis	0.02	-0.16	0.21	0.817	0.972

<sup>a</sup> $P < 0.05$ .<sup>c</sup> $P < 0.001$ .

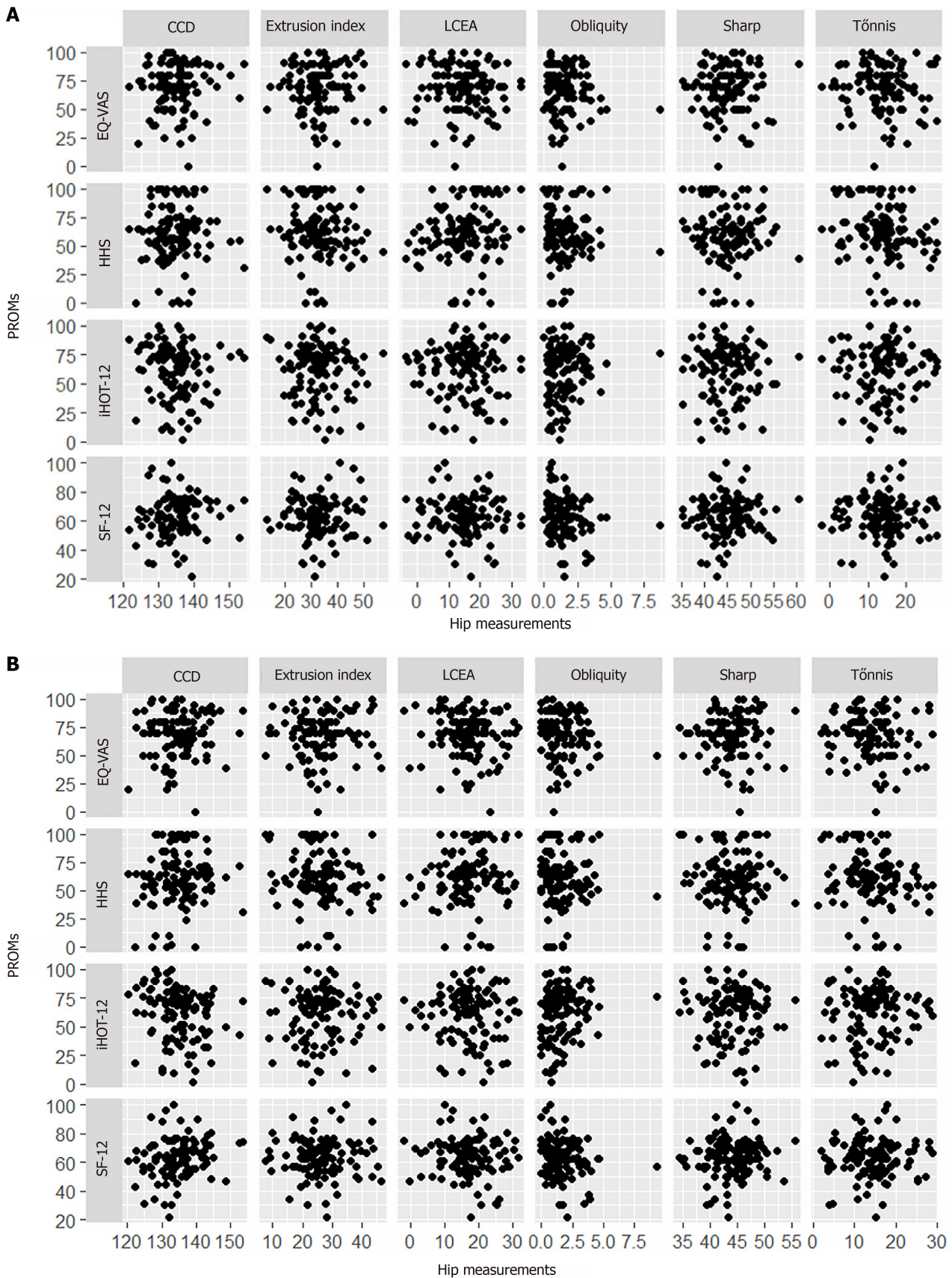
CCD: Caput-Collum-Diaphyseal; LCEA: Lateral center edge angle.

effectiveness of AI in clinical decision-making for HD.

## CONCLUSION

In conclusion, this study validated fast measurements using AI-software. Some correlations between AI-derived radiographic measurements and PROMs were seen in HD patients but these findings are mostly insignificant and weak, with most of the associations mirroring that of manual readers. Thus, at present, AI interpretations of radiographic data should be used with caution when predicting patient-reported outcomes. The potential of AI in clinical decision-making for HD patients remains promising in providing quick and accurate radiographic hip measurements. AI software has massive potential in streamlining physician workflow and in performing measurements that can have influence on the clinical decision-making process for patients with HD. It is through these continued efforts that we may fully realize the role of AI in the management of HD, while PROMs will continue to play a crucial role in assessing the broader implications of treatment on patient quality of life.





**Figure 3 Scatterplot.** A: Manual reader measurements and patient-reported outcome measures correlations; B: Artificial intelligence measurements and patient-reported outcome measures correlations. SF-12: Short form 12; iHOT-12: International hip outcome tool; HHS: Harris hip score; EQ-VAS: Visual Analogue Scale of the European Quality of Life Group; CCD: Caput-Collum-Diaphyseal; LCEA: Lateral center edge angle.

## FOOTNOTES



**Author contributions:** Alshaikhsalama A and Archer H were involved in the conception, design, data collection, writing and editing of the manuscript; Xi Y supervised and edited the manuscript and performed the statistical analysis; Ljuhar H, Wells J, and Chhabra A involved in the conception, design, and supervision of the manuscript; all of the authors read and approved the final version of the manuscript to be published.

**Institutional review board statement:** The study was reviewed and approved by the University of Texas Southwestern Institutional Review Board (approval No. Stu-2022-1014).

**Informed consent statement:** The University of Texas Southwestern Institutional Review Board determined informed consent was not required for this study since the data is fully anonymized.

**Conflict-of-interest statement:** Wells JE had received fees for serving as a consultant for Ethicon; Ljuhar R was an employee of Image Biopsy Labs that developed HIPPO AI software; Chhabra A had received fees for serving as a consultant for ICON Medical and TREACE Medical Concepts Inc and for serving as a Siemens Medical advisor for Image Biopsy Inc.

**Data sharing statement:** Technical appendix, statistical code, and dataset available from the corresponding author at [ahmed.alshaikhsalama@utsouthwestern.edu](mailto:ahmed.alshaikhsalama@utsouthwestern.edu). Consent was not obtained but the presented data are anonymized and risk of identification is low.

**Open-Access:** This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <https://creativecommons.org/licenses/by-nc/4.0/>

**Country of origin:** United States

**ORCID number:** Ahmed Alshaikhsalama 0000-0001-9465-8994; Avneesh Chhabra 0000-0002-0718-3802.

**Corresponding Author's Membership in Professional Societies:** American Medical Association, No. 672399.

**S-Editor:** Luo ML

**L-Editor:** A

**P-Editor:** Yu HG

## REFERENCES

- 1 Tannast M, Hanke MS, Zheng G, Steppacher SD, Siebenrock KA. What are the radiographic reference values for acetabular under- and overcoverage? *Clin Orthop Relat Res* 2015; **473**: 1234-1246 [PMID: 25384429 DOI: 10.1007/s11999-014-4038-3]
- 2 Bedi A, Dolan M, Leunig M, Kelly BT. Static and dynamic mechanical causes of hip pain. *Arthroscopy* 2011; **27**: 235-251 [PMID: 21035993 DOI: 10.1016/j.arthro.2010.07.022]
- 3 Livermore AT, Anderson LA, Anderson MB, Erickson JA, Peters CL. Correction of mildly dysplastic hips with periacetabular osteotomy demonstrates promising outcomes, achievement of correction goals, and excellent five-year survivorship. *Bone Joint J* 2019; **101-B**: 16-22 [PMID: 31146564 DOI: 10.1302/0301-620X.101B6.BJJ-2018-1487.R1]
- 4 Ömeroglu H. Treatment of developmental dysplasia of the hip with the Pavlik harness in children under six months of age: indications, results and failures. *J Child Orthop* 2018; **12**: 308-316 [PMID: 30154920 DOI: 10.1302/1863-2548.12.180055]
- 5 Rogers BA, Garbedian S, Kuchinad RA, Backstein D, Safir O, Gross AE. Total hip arthroplasty for adult hip dysplasia. *J Bone Joint Surg Am* 2012; **94**: 1809-1821 [PMID: 23032592 DOI: 10.2106/JBJS.K.00779]
- 6 Beltran LS, Rosenberg ZS, Mayo JD, De Tuesta MD, Martin O, Neto LP, Bencardino JT. Imaging evaluation of developmental hip dysplasia in the young adult. *AJR Am J Roentgenol* 2013; **200**: 1077-1088 [PMID: 23617493 DOI: 10.2214/AJR.12.9360]
- 7 Kraeutler MJ, Safran MR, Scillia AJ, Ayeni OR, Garabekyan T, Mei-Dan O. A Contemporary Look at the Evaluation and Treatment of Adult Borderline and Frank Hip Dysplasia. *Am J Sports Med* 2020; **48**: 2314-2323 [PMID: 31725329 DOI: 10.1177/0363546519881411]
- 8 Dwan LN, Gibbons P, Jamil K, Little D, Birke O, Menezes MP, Burns J. Reliability and sensitivity of radiographic measures of hip dysplasia in childhood Charcot-Marie-Tooth disease. *Hip Int* 2023; **33**: 323-331 [PMID: 34180253 DOI: 10.1177/11207000211027591]
- 9 Barrera CA, Cohen SA, Sankar WN, Ho-Fung VM, Sze RW, Nguyen JC. Imaging of developmental dysplasia of the hip: ultrasound, radiography and magnetic resonance imaging. *Pediatr Radiol* 2019; **49**: 1652-1668 [PMID: 31686171 DOI: 10.1007/s00247-019-04504-3]
- 10 Tamaki T, Oinuma K, Nakakita Y, Miura Y, Higashi H, Shiratsuchi H. Patient-Reported Outcomes and Perioperative Complications of Total Hip Arthroplasty Following Joint Preservation Surgery for Hip Dysplasia. *J Arthroplasty* 2020; **35**: 1622-1626 [PMID: 32088057 DOI: 10.1016/j.arth.2020.01.072]
- 11 Nwachukwu BU, Chang B, Beck EC, Neal WH, Movassaghi K, Ranawat AS, Nho SJ. How Should We Define Clinically Significant Outcome Improvement on the iHOT-12? *HSS J* 2019; **15**: 103-108 [PMID: 31327939 DOI: 10.1007/s11420-018-9646-0]
- 12 Clohisy JC, Carlisle JC, Beaulé PE, Kim YJ, Trousdale RT, Sierra RJ, Leunig M, Schoenecker PL, Millis MB. A systematic approach to the plain radiographic evaluation of the young adult hip. *J Bone Joint Surg Am* 2008; **90** Suppl 4: 47-66 [PMID: 18984718 DOI: 10.2106/JBJS.H.00756]
- 13 Christensen CP, Althausen PL, Mittleman MA, Lee JA, McCarthy JC. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res* 2003; **75**: 75-83 [PMID: 12579003 DOI: 10.1097/01.blo.0000043047.84315.4b]
- 14 Wells J, Schoenecker P, Duncan S, Goss CW, Thomason K, Clohisy JC. Intermediate-Term Hip Survivorship and Patient-Reported Outcomes of Periacetabular Osteotomy: The Washington University Experience. *J Bone Joint Surg Am* 2018; **100**: 218-225 [PMID: 29406343 DOI: 10.2106/JBJS.17.00337]

- 15 **Matheney T**, Kim YJ, Zurakowski D, Matero C, Millis M. Intermediate to long-term results following the Bernese periacetabular osteotomy and predictors of clinical outcome. *J Bone Joint Surg Am* 2009; **91**: 2113-2123 [PMID: [19723987](#) DOI: [10.2106/JBJS.G.00143](#)]
- 16 **Fan Y**, Li W, Wu Y, Li R, Hong G, Li Z, Chen L, Fang H, Zhou C, He W, Chen Z. The association the patient-reported outcomes after periacetabular osteotomy with radiographic features: a short-term retrospective study. *J Orthop Surg Res* 2021; **16**: 718 [PMID: [34924013](#) DOI: [10.1186/s13018-021-02858-9](#)]
- 17 **Impellizzeri FM**, Jones DM, Griffin D, Harris-Hayes M, Thorborg K, Crossley KM, Reiman MP, Scholes MJ, Ageberg E, Agricola R, Bizzini M, Bloom N, Casartelli NC, Diamond LE, Dijkstra HP, Di Stasi S, Drew M, Friedman DJ, Freke M, Gojanovic B, Heerey JJ, Hölmich P, Hunt MA, Ishoi L, Kassarian A, King M, Lawrenson PR, Leunig M, Lewis CL, Warholm KM, Mayes S, Moksnes H, Mosler AB, Risberg MA, Semciw A, Serner A, van Klij P, Wörner T, Kemp J. Patient-reported outcome measures for hip-related pain: a review of the available evidence and a consensus statement from the International Hip-related Pain Research Network, Zurich 2018. *Br J Sports Med* 2020; **54**: 848-857 [PMID: [32066573](#) DOI: [10.1136/bjsports-2019-101456](#)]
- 18 **Takegami Y**, Seki T, Osawa Y, Kusano T, Ishiguro N. The relationship between radiographic findings and patient-reported outcomes in adult hip dysplasia patients: A hospital cross-sectional study. *J Orthop Sci* 2020; **25**: 606-611 [PMID: [31607517](#) DOI: [10.1016/j.jos.2019.09.007](#)]
- 19 **Archer H**, Reine S, Alshaikhsalama A, Wells J, Kohli A, Vazquez L, Hummer A, DiFranco MD, Ljuhar R, Xi Y, Chhabra A. Artificial intelligence-generated hip radiological measurements are fast and adequate for reliable assessment of hip dysplasia : an external validation study. *Bone Jt Open* 2022; **3**: 877-884 [PMID: [36373773](#) DOI: [10.1302/2633-1462.311.BJO-2022-0125.R1](#)]
- 20 **Gandek B**, Ware JE, Aaronson NK, Apolone G, Björner JB, Brazier JE, Bullinger M, Kaasa S, Leplege A, Prieto L, Sullivan M. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; **51**: 1171-1178 [PMID: [9817135](#) DOI: [10.1016/s0895-4356\(98\)00109-7](#)]
- 21 **Griffin DR**, Parsons N, Mohtadi NG, Safran MR; Multicenter Arthroscopy of the Hip Outcomes Research Network. A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthroscopy* 2012; **28**: 611-6; quiz 616 [PMID: [22542434](#) DOI: [10.1016/j.arthro.2012.02.027](#)]
- 22 **Balestroni G**, Bertolotti G. [EuroQol-5D (EQ-5D): an instrument for measuring quality of life]. *Monaldi Arch Chest Dis* 2012; **78**: 155-159 [PMID: [23614330](#) DOI: [10.4081/monaldi.2012.121](#)]
- 23 **Isaac B**, Vettivel S, Prasad R, Jeyaseelan L, Chandi G. Prediction of the femoral neck-shaft angle from the length of the femoral neck. *Clin Anat* 1997; **10**: 318-323 [PMID: [9283729](#) DOI: [10.1002/\(SICI\)1098-2353\(1997\)10:5<318::AID-CA5>3.0.CO;2-M](#)]
- 24 **Hanson JA**, Kapron AL, Swenson KM, Maak TG, Peters CL, Aoki SK. Discrepancies in measuring acetabular coverage: revisiting the anterior and lateral center edge angles. *J Hip Preserv Surg* 2015; **2**: 280-286 [PMID: [27011850](#) DOI: [10.1093/jhps/hnv041](#)]
- 25 **Schober P**, Boer C, Schwarte LA. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg* 2018; **126**: 1763-1768 [PMID: [29481436](#) DOI: [10.1213/ANE.0000000000002864](#)]
- 26 **Jensen MP**, Chen C, Brugger AM. Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of postoperative pain. *J Pain* 2003; **4**: 407-414 [PMID: [14622683](#) DOI: [10.1016/s1526-5900\(03\)00716-8](#)]
- 27 **Yu DS**, Yan EC, Chow CK. Interpreting SF-12 mental component score: an investigation of its convergent validity with CESD-10. *Qual Life Res* 2015; **24**: 2209-2217 [PMID: [25786886](#) DOI: [10.1007/s11136-015-0959-x](#)]
- 28 **Robinson PG**, Maempel JF, Rankin CS, Gaston P, Hamilton DF. Evaluation of the patient acceptable symptom state following hip arthroscopy using the 12 item international hip outcome tool. *BMC Musculoskelet Disord* 2020; **21**: 5 [PMID: [31900150](#) DOI: [10.1186/s12891-019-3026-x](#)]
- 29 **Harris WH**. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969; **51**: 737-755 [PMID: [5783851](#)]
- 30 **Forsberg D**, Rosipko B, Sunshine JL. Radiologists' Variation of Time to Read Across Different Procedure Types. *J Digit Imaging* 2017; **30**: 86-94 [PMID: [27714473](#) DOI: [10.1007/s10278-016-9911-z](#)]
- 31 **Clohisey JC**, Nunley RM, Carlisle JC, Schoenecker PL. Incidence and characteristics of femoral deformities in the dysplastic hip. *Clin Orthop Relat Res* 2009; **467**: 128-134 [PMID: [19034600](#) DOI: [10.1007/s11999-008-0481-3](#)]
- 32 **Boese CK**, Frink M, Jostmeier J, Haneder S, Dargel J, Eysel P, Lechler P. The Modified Femoral Neck-Shaft Angle: Age- and Sex-Dependent Reference Values and Reliability Analysis. *Biomed Res Int* 2016; **2016**: 8645027 [PMID: [28070521](#) DOI: [10.1155/2016/8645027](#)]
- 33 **Ellis C**, Grubaugh AL, Egede LE. Factors associated with SF-12 physical and mental health quality of life scores in adults with stroke. *J Stroke Cerebrovasc Dis* 2013; **22**: 309-317 [PMID: [22005038](#) DOI: [10.1016/j.jstrokecerebrovasdis.2011.09.007](#)]
- 34 **Kraeutler MJ**, Garabekyan T, Pascual-Garrido C, Mei-Dan O. Hip instability: a review of hip dysplasia and other contributing factors. *Muscles Ligaments Tendons J* 2016; **6**: 343-353 [PMID: [28066739](#) DOI: [10.11138/mltj/2016.6.3.343](#)]
- 35 **Nunley RM**, Prather H, Hunt D, Schoenecker PL, Clohisey JC. Clinical presentation of symptomatic acetabular dysplasia in skeletally mature patients. *J Bone Joint Surg Am* 2011; **93** Suppl 2: 17-21 [PMID: [21543683](#) DOI: [10.2106/JBJS.J.01735](#)]
- 36 **Kaya O**, Gonder N, Demir DA, Demir IH, Akay O. Patients with developmental dysplasia of the hip shows higher digit ratio (2D:4D). *Early Hum Dev* 2023; **185**: 105857 [PMID: [37677892](#) DOI: [10.1016/j.earlhumdev.2023.105857](#)]
- 37 **Ishidou Y**, Matsuyama K, Sakuma D, Setoguchi T, Nagano S, Kawamura I, Maeda S, Komiya S. Osteoarthritis of the hip joint in elderly patients is most commonly atrophic, with low parameters of acetabular dysplasia and possible involvement of osteoporosis. *Arch Osteoporos* 2017; **12**: 30 [PMID: [28332125](#) DOI: [10.1007/s11657-017-0325-4](#)]



Published by **Baishideng Publishing Group Inc**  
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

**Telephone:** +1-925-3991568

**E-mail:** [office@baishideng.com](mailto:office@baishideng.com)

**Help Desk:** <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

