Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Knowledge-guided artificial intelligence supports radiographic surgical planning for acetabular reconstruction in developmental hip dysplasia

Shihao Chen
West China Hospital, Sichuan University
Junqing Wang
West China Hospital, Sichuan University
Biao Wang
West China Hospital, Sichuan University
Guo Chen
Sichuan Provincial Orthopedic Hospital
Tao Deng
West China Hospital, Sichuan University
Liang Zhao
Dyania Health
Qiang Zhang
West China Hospital, Sichuan University
Jiaqu Sun
West China Hospital, Sichuan University
Bingxing Huo
West China Hospital, Sichuan University
Yi Zeng
West China Hospital, Sichuan University
Kang Li
West China Hospital, Sichuan University
Zongke Zhou
West China Hospital, Sichuan University
Yong Nie
nieyong1983@wchscu.cn

West China Hospital, Sichuan University

Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11591828/v1

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

Additional Declarations: No competing interests reported.

Knowledge-guided artificial intelligence supports radiographic surgical planning for acetabular reconstruction in developmental hip dysplasia

Shihao Chen[1†], Junqing Wang[1,2†], Biao Wang[2†], Guo Chen[3†], Tao Deng[1], Liang Zhao[4], Qiang Zhang[1], Jiaqu Sun[5], Bingxing Huo[6], Yi Zeng[1*], Kang Li[2,7,8*], Zongke Zhou[1*], Yong Nie[1*]

† These authors contributed equally and share first authorship.

* Corresponding authors.

1 Department of Orthopedic Surgery and Orthopedic Research Institute, West China Hospital, Sichuan University, Chengdu, China

2 West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China

3 Department of Geriatric Orthopedics, Sichuan Provincial Orthopedic Hospital, Chengdu, Sichuan, China

4 Dyania Health, New Jersey, USA

5 Department of Radiology, West China Hospital, Sichuan University, Chengdu, China

6 West China Hospital, Sichuan University, Chengdu, China

7 Med-X Center for Informatics, Sichuan University, Chengdu, China

8 Sichuan University-Pittsburgh Institute, Sichuan University, Chengdu, China

Correspondence: Yong Nie (nieyong1983@wchscu.cn); Kang Li (likang@wchscu.cn); Zongke Zhou (zhouzongke@scu.edu.cn); Yi Zeng (zengyi51@scu.edu.cn).

Abstract

Acetabular planning for developmental dysplasia of the hip (DDH) commonly uses two-dimensional templating or CT-based workflows. We developed Knowledge-base-guided Acetabular Planning for Developmental Dysplasia of the Hip (KAP-DDH), which integrates automated two-dimensional radiographic analysis with paired historical surgical reconstructions to generate patient-specific plans. In this retrospective two-centre study, development, internal-validation and external-validation cohorts comprised 2,760, 691 and 422 records. Evaluation used two expert references and two blinded surgeon planners. External cup-diameter mean absolute errors were 3.18 mm for KAP-DDH, 4.37 mm for Surgeon 1 and 3.25 mm for Surgeon 2; both surgeons had smaller errors internally. Historical-case guidance reduced cup-diameter error by 0.54 mm internally and 0.62 mm externally versus Crowe-specific defaults and reduced positional errors, with little change in inclination or coverage errors. All evaluated KAP-DDH plans met prespecified inclination and coverage criteria. These findings support translating historical reconstruction experience into patient-specific radiographic planning targets.

Introduction

Developmental dysplasia of the hip (DDH) is a structural hip disorder that accounts for a substantial proportion of arthroplasties in younger adults. A systematic review of 10,998 hips from 5,506 asymptomatic adults reported a pooled prevalence of radiographic hip dysplasia of 2.3%.[1] Among patients younger than 55 years undergoing total hip arthroplasty (THA), DDH accounted for 627 of 3,219 THAs (19.5%) in another systematic review.[2] A meta-analysis of nine observational studies including 575,255 THAs found a younger mean age at surgery in DDH than in primary osteoarthritis (50.6 versus 62.1 years) and higher odds of revision (odds ratio 1.66, 95% confidence interval 1.11–2.48).[3] These findings emphasise the clinical importance of planning a durable reconstruction in patients with dysplastic anatomy.

DDH ranges from a shallow acetabulum to high dislocation, with variable acetabular dimensions, orientation, bone distribution and hip-centre position.[4–6] The surgeon must accommodate a small or deficient acetabulum while selecting component size and orientation,

restoring the hip centre and obtaining host-bone support.[7] These decisions are coupled: cup size affects bone coverage and removal, while centre position changes both coverage and reconstruction geometry. Poor cup sizing or positioning can sacrifice bone stock, compromise fixation and hip biomechanics, and contribute to instability or loosening.[4] Superior cup placement has been associated with long-term loosening in Crowe II hips.[8] Planning must therefore extend beyond implant sizing to balance component position, coverage and bone preservation for each hip. Crowe severity provides useful clinical context, but hips within the same category can have different local bone deficiencies and require different solutions.[9]

Standardised anteroposterior pelvic radiographs provide an accessible basis for this assessment. Pelvic reference lines, acetabular margins and the hip centre support measurements of displacement and component size and position.[10–12] Proximal femoral-head migration underlies the Crowe classification.[13] Three-dimensional studies show additional variation in acetabular orientation, medial-wall thickness and segmental bone deficiency across Crowe types.[5,6] Radiographic coverage captures one projection of this anatomy: comparisons with intraoperative and three-dimensional measurements support its use for describing projected support, while demonstrating that the measures are not interchangeable.[14–16] Paired preoperative and postoperative radiographs contain a further source of information. They link the anatomy encountered before surgery to the cup size and position subsequently achieved, providing a record of how surgeons have addressed related reconstruction problems.

Artificial intelligence (AI) has extended radiographic assessment from automated measurements[17] to component templating and size prediction,[18] CT-based planning,[19–21] landmark localisation[22] and postoperative component assessment.[23] External validation has also demonstrated the feasibility of automated hip measurements.[24] In 103 patients with DDH, Xie and colleagues reported exact cup-size prediction in 84.1% using AI-assisted three-dimensional planning, compared with 64.0% using two-dimensional templates.[20] More recent approaches generate simulated postoperative THA radiographs[25] or combine keypoint detection with geometric rules for paediatric DDH osteotomy planning.[26] These advances establish that AI can support

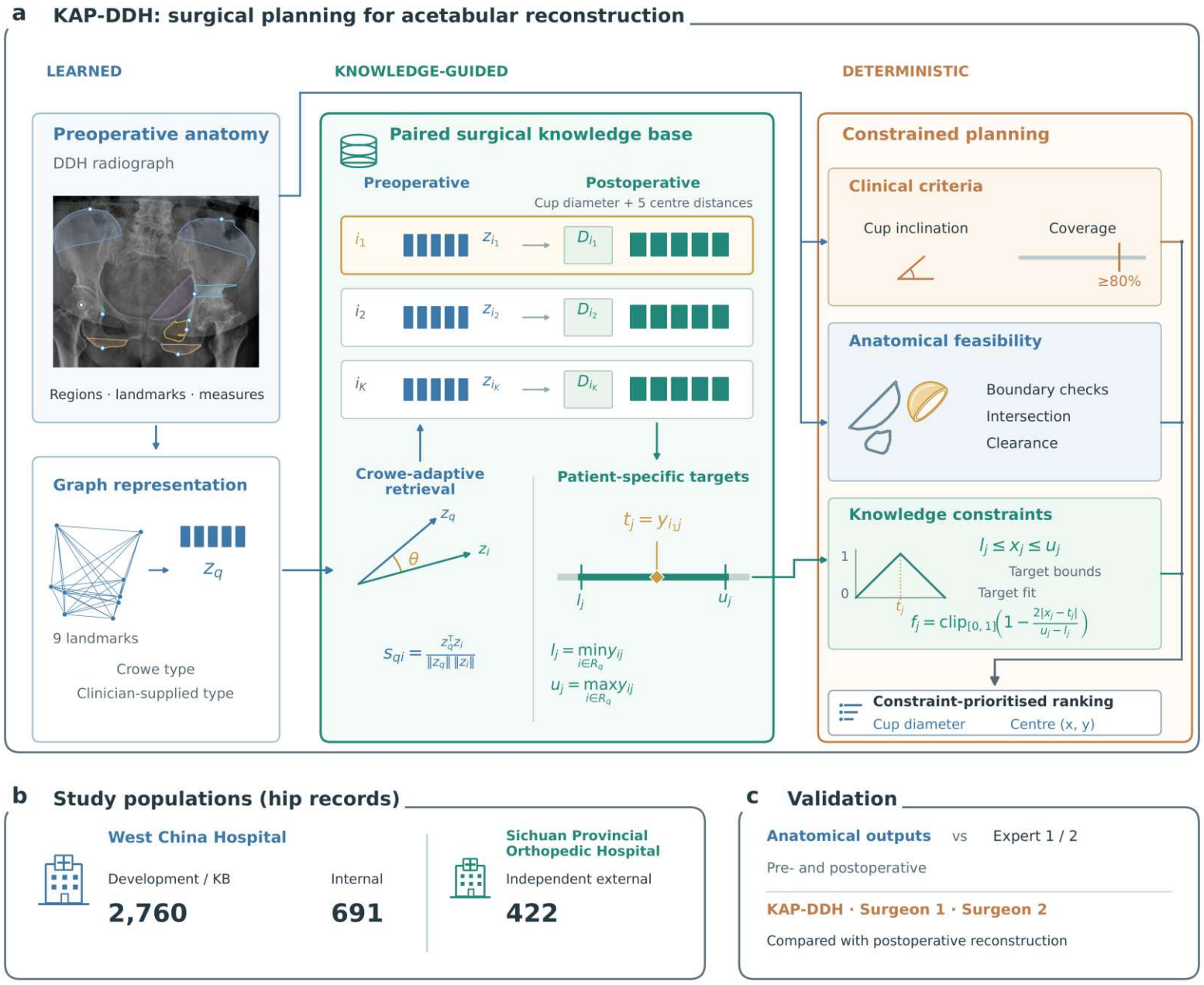
planning beyond anatomical recognition. A specific opportunity in adult dysplastic THA is to make the contribution of prior reconstructions explicit: identify anatomically relevant operated hips, translate their reconstruction measurements into patient-specific targets, and evaluate those targets against the current hip's geometry.

Knowledge-based planning and case-based reasoning offer a way to use this information. Previous cases inform a new decision through representation, retrieval and adaptation, as illustrated by knowledge-based radiotherapy planning and treatment-plan retrieval.[27–29] In DDH, paired records can connect preoperative anatomy to postoperative cup size and position. A retrieved neighbourhood can describe a range of historical solutions, with a selected record providing a preferred target. Geometric evaluation then determines how those targets apply to the current hip. This approach gives the surgeon access to both the proposed measurements and the historical reconstructions that informed them, making the use of prior experience inspectable.

Here we develop Knowledge-base-guided Acetabular Planning for Developmental Dysplasia of the Hip (KAP-DDH), a two-dimensional radiograph-based system for acetabular reconstruction planning (Fig. 1). Learned perception identifies anatomical regions from which landmarks and measurements are derived. Nine preoperative landmarks and clinician-supplied Crowe type guide retrieval of historical cases; their postoperative reconstructions define six target intervals and preferred values for cup size and centre position. Deterministic candidate generation and ranking evaluate proposed reconstructions in the current anatomy and select the automated Top-1 output. Historical postoperative records supply planning knowledge, while the query record's postoperative information is reserved for evaluation.

Development used 2,760 operated-hip records from West China Hospital; validation used 691 internal and 422 separate-institution records. We assessed preoperative perception, postoperative measurement and retrieval-derived targets, then compared automated plans with two independent surgeon plans. Two expert annotations assessed reference sensitivity. Crowe-stratified displays characterised the validation spectrum, and a paired comparison with Crowe-specific default targets assessed historical guidance within the same pipeline.

Figure 1 | Knowledge-guided surgical planning for acetabular reconstruction



a, Preoperative radiographs provide anatomical regions, landmarks and measurements. Nine-landmark graph representations incorporate clinician-supplied Crowe type; graph edges are schematic. Crowe-adaptive retrieval matches query and historical preoperative embeddings, zq and zi, using cosine similarity sqi. Indexed paired records link historical preoperative representations to postoperative cup diameter Di and five cup-centre distances. Rq denotes the retrieved record set for query q; target bounds lj and uj are derived from available values yij for each field within this set. For j = 1, …, 6, fields comprise cup diameter and centre distances to the inter-teardrop, Köhler, bi-iliac, bi-ischial and iliac–ischial midpoint references. When a non-degenerate interval and the first-ranked identified record's value are available, tj = $y_{i_1 j}$; otherwise, configured field-specific values are retained. Gold marks the preferred-record relationship. Preoperative anatomical outputs support candidate construction and anatomical checks; retrieved targets provide knowledge constraints. Clinical, anatomical and knowledge assessments inform constraint-prioritised ranking. The illustrated fj, defined for uj > lj, is a target-fit component rather than the complete ranking score. xj denotes a candidate measurement; clip[0,1] limits the value to [0,1]. Its curve illustrates the response around tj, with zeros at tj ± (uj − lj)/2, rather than necessarily at lj and uj. Clinical criteria comprise cup inclination relative to the inter-teardrop reference and radiographic cup coverage. Fallback branches do not guarantee that all constraints will be satisfied. The radiograph, segmentation and landmarks are derived from saved case assets. The gold cup shell is schematic; the two blue outlines retain saved contour shapes. Their juxtaposition illustrates anatomical checks, not an observed or planned cup position. Embedding blocks, record rows, vector directions and the target interval illustrate principles without numerical retrieval results. b, Development/knowledge-base, internal-validation and independent external-validation populations; counts denote included hip records. The knowledge base contains 2,760 paired records. Eight external-cohort patients contributed two records from the same side. c, Anatomical outputs are compared with expert references; KAP-DDH and two surgeon plans are compared with postoperative reconstruction.

Results

Study cohorts and the knowledge-guided planning framework

KAP-DDH converts a preoperative pelvic radiograph into a ranked acetabular reconstruction plan through anatomical perception, retrieval of historical records and deterministic candidate evaluation (Figs. 1 and 2). The study comprised 3,873 records from 3,393 unique patients. Development and knowledge-base construction used 2,760 records from 2,492 patients at West China Hospital; internal validation included 691 records from 672 patients at the same institution. External validation included 422 records from 362 patients at Sichuan Provincial Orthopedic Hospital. Figure 2 details the data flow, the information available at each stage and the evaluation components.

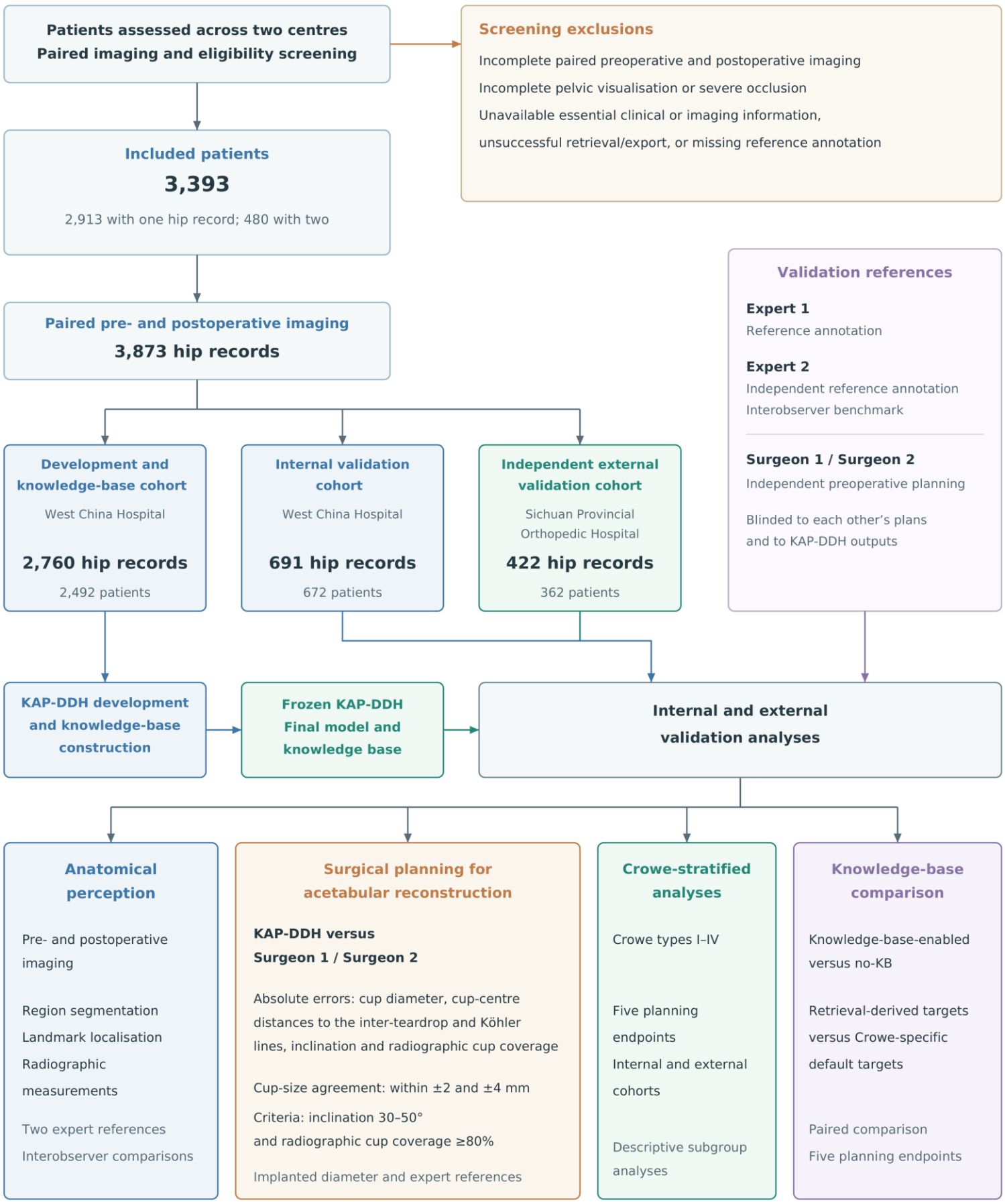
Median age was 60.0 years (interquartile range, 52.0–68.0) in both the development and internal cohorts, and 59.5 years (54.0–67.0) in the external cohort (Table 1). Median body mass index was 23.8 kg/m² in the development and internal cohorts and 24.6 kg/m² in the external cohort. Women accounted for 2,082/2,492 development patients, 568/672 internal patients and 291/362 external patients. Age and body mass index were available for all patients. Table 1 distinguishes patient-level characteristics from record-level operated side and Crowe type. Both centres contributed the full range of recorded Crowe types, although the proportion of severe dysplasia differed.

Table 1 | Cohort characteristics

Characteristic	Development	Internal validation	Independent external validation
Cohort size			
Centre	West China Hospital	West China Hospital	Sichuan Provincial Orthopedic Hospital
Patients, n	2,492	672	362
Included hip records, n	2,760	691	422
Patient-level characteristics			
Age, years, median [IQR]	60.0 [52.0–68.0]	60.0 [52.0–68.0]	59.5 [54.0–67.0]
BMI, kg/m², median [IQR]	23.8 [21.7–26.2]	23.8 [21.6–26.3]	24.6 [22.3–26.7]
Sex: Female	2,082/2,492 (83.5%)	568/672 (84.5%)	291/362 (80.4%)
Sex: Male	410/2,492 (16.5%)	104/672 (15.5%)	71/362 (19.6%)
Record-level characteristics			
Operated side: Left	1,248/2,760 (45.2%)	321/691 (46.5%)	181/422 (42.9%)
Operated side: Right	1,512/2,760 (54.8%)	370/691 (53.5%)	241/422 (57.1%)
Crowe type I	2,080/2,760 (75.4%)	521/691 (75.4%)	336/422 (79.6%)
Crowe type II	400/2,760 (14.5%)	99/691 (14.3%)	75/422 (17.8%)
Crowe type III	111/2,760 (4.0%)	29/691 (4.2%)	7/422 (1.7%)
Crowe type IV	169/2,760 (6.1%)	42/691 (6.1%)	4/422 (0.9%)

Values are median [interquartile range] or n/N (%). Age, body mass index (BMI) and sex are summarised once per patient within each cohort, using the included record with the lowest surgery sequence number. Side and Crowe type are summarised per hip record. The 2,760 development records form the knowledge base.

Figure 2 | Study design and validation workflow



Patients from West China Hospital and Sichuan Provincial Orthopedic Hospital were screened for eligibility based on the availability of paired preoperative and postoperative radiographs and the required clinical, imaging and reference information. The included dataset comprises 3,393 unique patients and 3,873 hip records: 2,913 patients contributed one record and 480 contributed two. Eight external-cohort patients contributed two records from the same side. The development/knowledge-base, internal-validation and independent external-validation cohorts contain 2,760, 691 and 422 hip records, respectively. The development cohort supports KAP-DDH development and knowledge-base construction. The frozen model and knowledge base undergo internal and external validation. Expert 1 provides reference annotations, while Expert 2 provides an independent reference and an interobserver benchmark. Surgeon 1 and Surgeon 2 perform independent preoperative planning, blinded to each other's plans and KAP-DDH outputs. The evaluation covers four analysis domains. Anatomical perception combines preoperative and postoperative region segmentation, landmark localisation and radiographic measurement assessment. Acetabular planning compares KAP-DDH with the two surgeon planners using absolute errors for cup diameter, cup-centre distance to the inter-teardrop line, cup-centre distance to the Köhler line, cup inclination and radiographic cup coverage. Recorded implanted diameter is the cup-size reference; Expert 1 and Expert 2 provide separate postoperative references for the four image-derived endpoints. Additional planning analyses assess cup-size agreement within ±2 and ±4 mm and satisfaction of the inclination criterion of 30–50°, the coverage criterion of at least 80%, and both criteria. Crowe-stratified analyses describe the five planning errors across Crowe types I–IV in the two validation cohorts. These subgroup displays are descriptive, with limited external representation of Crowe III and IV. The knowledge-base comparison contrasts retrieval-derived targets in knowledge-base-enabled planning with Crowe-specific default targets in the no-KB configuration. The two configurations are paired on the same hip records and evaluated using the same five planning endpoints and postoperative references.

The evaluation followed four linked stages: preoperative anatomical perception, postoperative reconstruction measurement, retrieval-derived target construction and planning comparison. The current record's preoperative anatomy supplied the retrieval query and candidate geometry, historical paired records supplied reconstruction targets, and the current record's postoperative reconstruction supplied the evaluation reference.

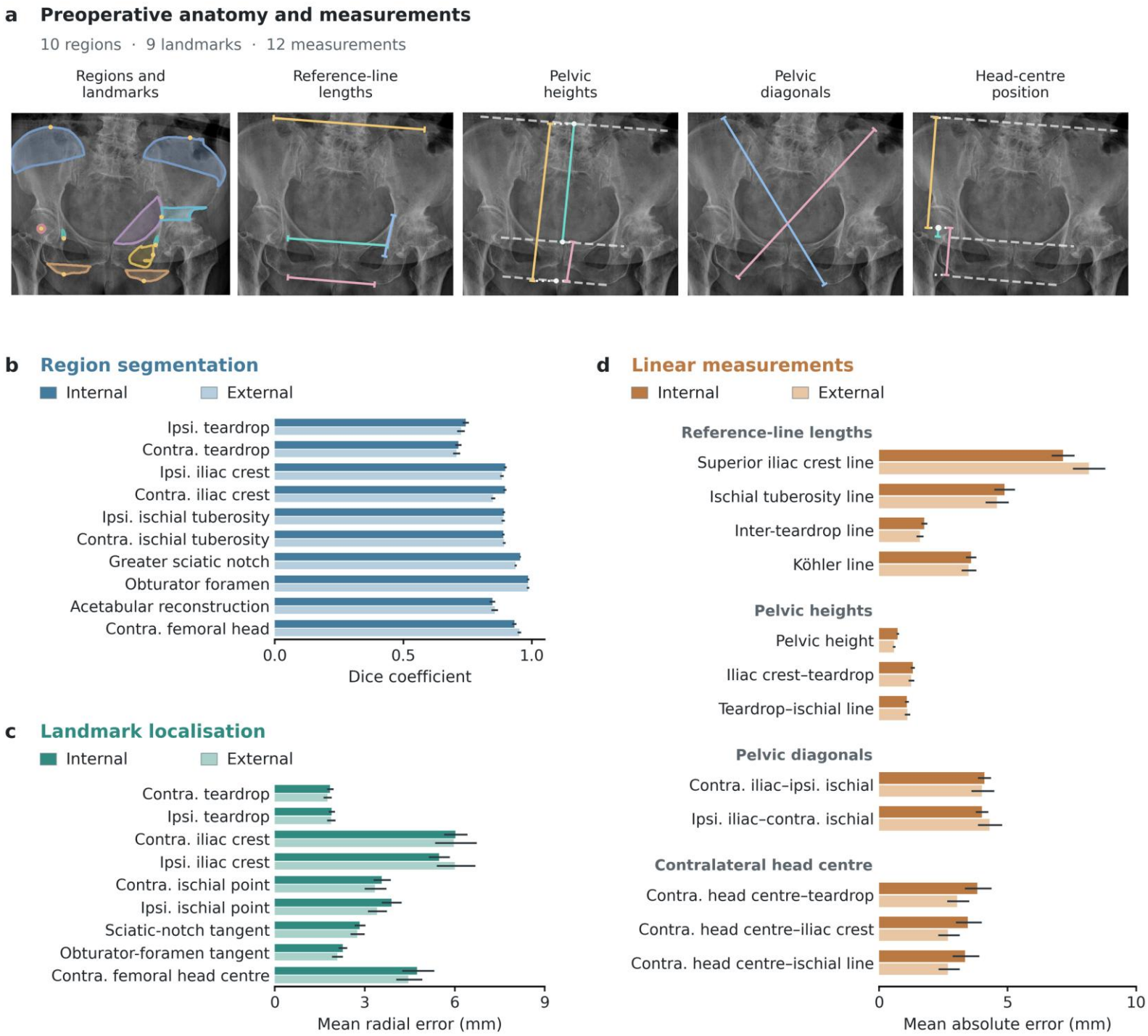
Preoperative anatomical perception across validation cohorts

Preoperative anatomical perception comprised ten regions, nine landmarks and twelve derived measurements (Fig. 3). Against Expert 1, macro-average Dice was 0.874 (95% confidence interval [CI], 0.871–0.877) internally and 0.867 (0.863–0.871) externally. Mean landmark localisation error was 3.61 mm (3.50–3.73) and 3.51 mm (3.37–3.67), respectively. Macro summaries were calculated by first averaging the applicable structures within each record, and confidence intervals accounted for patient clustering. The nine landmarks formed the structured anatomical representation subsequently used to retrieve historical cases.

The derived measurements described reference-line lengths, pelvic heights, pelvic diagonals and contralateral femoral-head position. Internal and external mean absolute errors (MAEs) were 0.72 and 0.58 mm for pelvic height, 1.31 and 1.25 mm for iliac crest–teardrop distance, and 1.08 and 1.10 mm for teardrop–ischial distance. These measurements assess the geometric quantities used downstream, complementing region overlap and landmark localisation.

Agreement varied with the reference annotator. Against Expert 2, macro-average Dice was 0.822 internally and 0.822 externally; corresponding Expert 1–Expert 2 values were 0.804 and 0.813. The paired comparisons place model agreement alongside interobserver agreement on the same records. Structure-specific estimates and measurement-agreement summaries are presented in Supplementary Figs. 1 and 2, Supplementary Table 11. Anatomical and measurement definitions are provided in Supplementary Tables 2–4.

Figure 3 | Reliability of preoperative anatomical perception and radiographic measurements



a, A representative preoperative radiograph from one external-validation case, showing automated region masks, mask-derived landmarks and all 12 measurement constructions. From left: regions and landmarks, reference-line lengths, pelvic heights, pelvic diagonals and contralateral head-centre distances. For reference-line lengths, gold, pink, teal and blue denote iliac-crest, ischial, inter-teardrop and Köhler segments, respectively. For centre-to-line distances, gold, teal and pink denote iliac-crest, inter-teardrop and ischial references; for pelvic heights, they denote total pelvic, iliac–teardrop and teardrop–ischial distances. Blue and pink distinguish the contralateral-iliac to ipsilateral-ischial diagonal from the opposite diagonal. Dashed white lines indicate references or extensions; dotted connectors link offset dimension annotations to their endpoints. b, Segmentation Dice coefficients for 10 regions. c, Mean radial localisation errors for nine landmarks. d, Mean absolute errors for all 12 linear measurements. Reference: Expert 1. Bars show record-weighted means, and error bars show 95% percentile confidence intervals from 10,000 patient-cluster bootstrap resamples, retaining all records within each sampled patient. Dark and light shades denote internal validation (691 hip records from 672 patients) and independent external validation (422 hip records from 362 patients), respectively. Blue, teal and ochre distinguish segmentation, localisation and measurement tasks. Ipsi., ipsilateral; contra., contralateral.

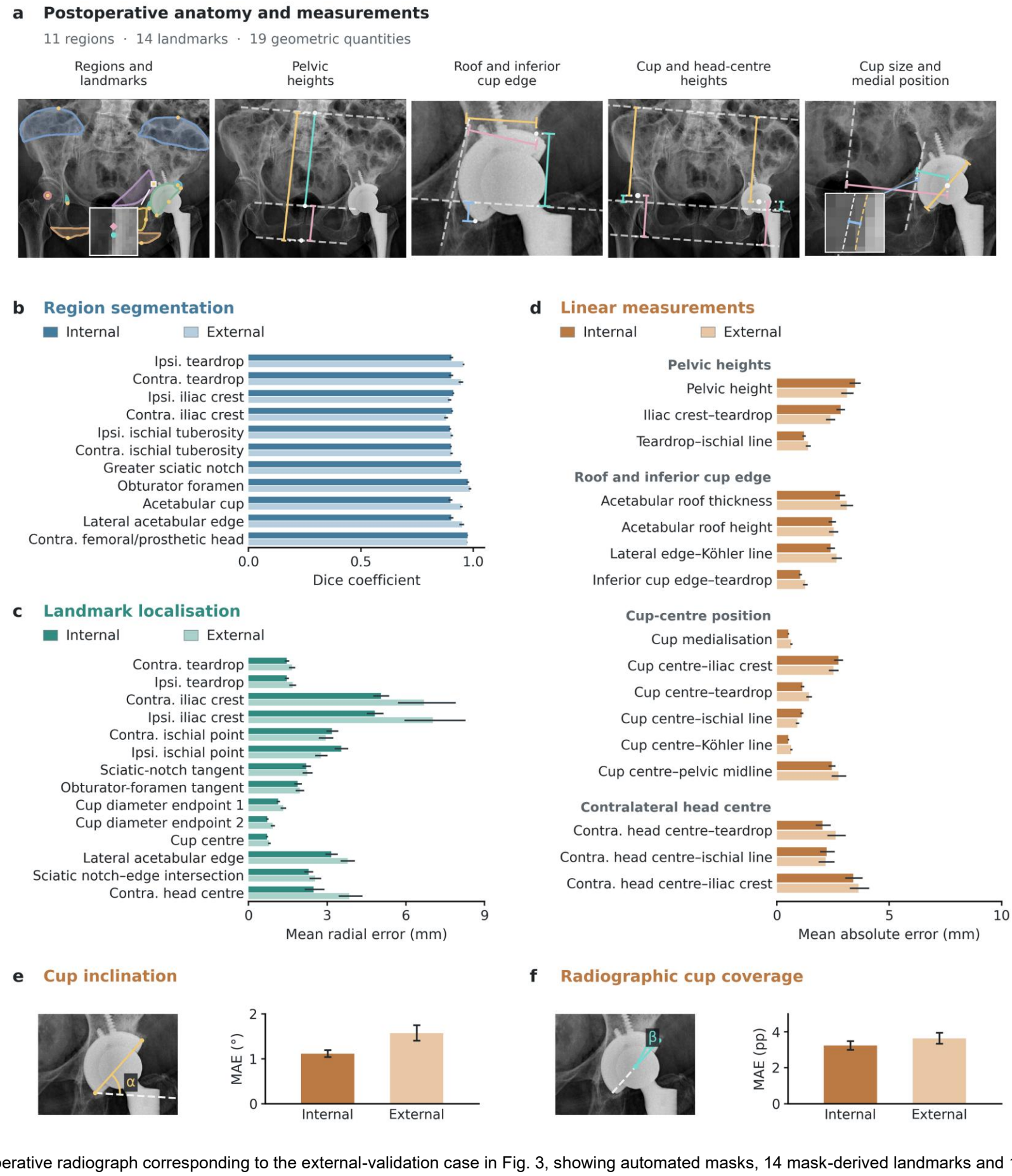
Postoperative anatomy and reconstruction measurements

Postoperative perception covered eleven regions and fourteen landmarks (Fig. 4). Against Expert 1, macro-average Dice was 0.919 (95% CI, 0.916–0.922) internally and 0.934 (0.932–0.937) externally. Mean landmark localisation error was 2.43 mm (2.37–2.50) and 2.87 mm (2.72–3.02), respectively. Against Expert 2, macro-average Dice was 0.850 internally and 0.852 externally, compared with interobserver values of 0.841 and 0.856.

The postoperative geometry comprised seventeen linear or geometric constructions, cup inclination and radiographic cup coverage. Sixteen linear endpoints, inclination and coverage were evaluated for measurement accuracy. These endpoints described pelvic reference geometry, acetabular-roof dimensions, cup-edge and cup-centre position, and contralateral head-centre position. The additional image-derived cup-diameter construction was retained for geometric illustration; recorded implanted diameter provided the separate reference for evaluating planned cup size.

For the four image-derived endpoints subsequently used to evaluate plans, internal MAEs were 1.14 mm for cup-centre distance to the inter-teardrop line, 0.51 mm for distance to the Köhler line, 1.12° for inclination and 3.24 percentage points for coverage. Corresponding external MAEs were 1.43 mm, 0.64 mm, 1.57° and 3.64 percentage points. Inclination was defined relative to the inter-teardrop line, and coverage was calculated using the implemented two-dimensional angular measure. Positional, angular and coverage errors were reported in their respective units. Observer comparisons and agreement summaries are presented in Supplementary Figs. 3–5 and Supplementary Table 12. These analyses quantify postoperative measurement accuracy; the planning analyses below compare proposed plans with the observed reconstruction.

Figure 4 | Reliability of postoperative anatomical perception and radiographic measurements



a, The postoperative radiograph corresponding to the external-validation case in Fig. 3, showing automated masks, 14 mask-derived landmarks and 17 linear geometric constructions. From left: regions and landmarks, pelvic heights, acetabular roof and inferior cup-edge geometry, cup- and contralateral head-centre distances, and cup size and medial position. Insets clarify adjacent landmarks and cup medialisation without changing source coordinates. The teal circle and pink diamond distinguish the sciatic-notch tangent from the lateral-edge-level intersection. Cup medialisation is the perpendicular cup-centre–Köhler distance minus the cup radius; its blue segment connects the Köhler reference to a dashed gold auxiliary radius-construction arc. Dashed white lines mark references; dotted connectors locate offset dimension annotations. b, Dice coefficients for 11 regions. c, Mean radial localisation errors for 14 landmarks. d, Mean absolute errors for 16 linear measurements. e, f, Cup-inclination (e) and radiographic cup-coverage (f) definitions and errors. α is the acute angle between cup-diameter and inter-teardrop directions; its white reference is parallel to the inter-teardrop line. β is the acute angle between cup-diameter and cup-centre–lateral-edge directions. Radiographic cup coverage is (180° − β)/180° × 100%. Together, a, e and f display all 19 geometric quantities; image-derived cup diameter is a construction quantity rather than an additional accuracy endpoint. Reference: Expert 1. Bars show record-weighted means, and error bars show 95% percentile confidence intervals from 10,000 patient-cluster bootstrap resamples, retaining all records within each sampled patient. Dark and light shades indicate internal and external validation, respectively. Blue, teal and ochre distinguish segmentation, localisation and measurement tasks. Ipsi., ipsilateral; contra., contralateral; MAE, mean absolute error; pp, percentage points.

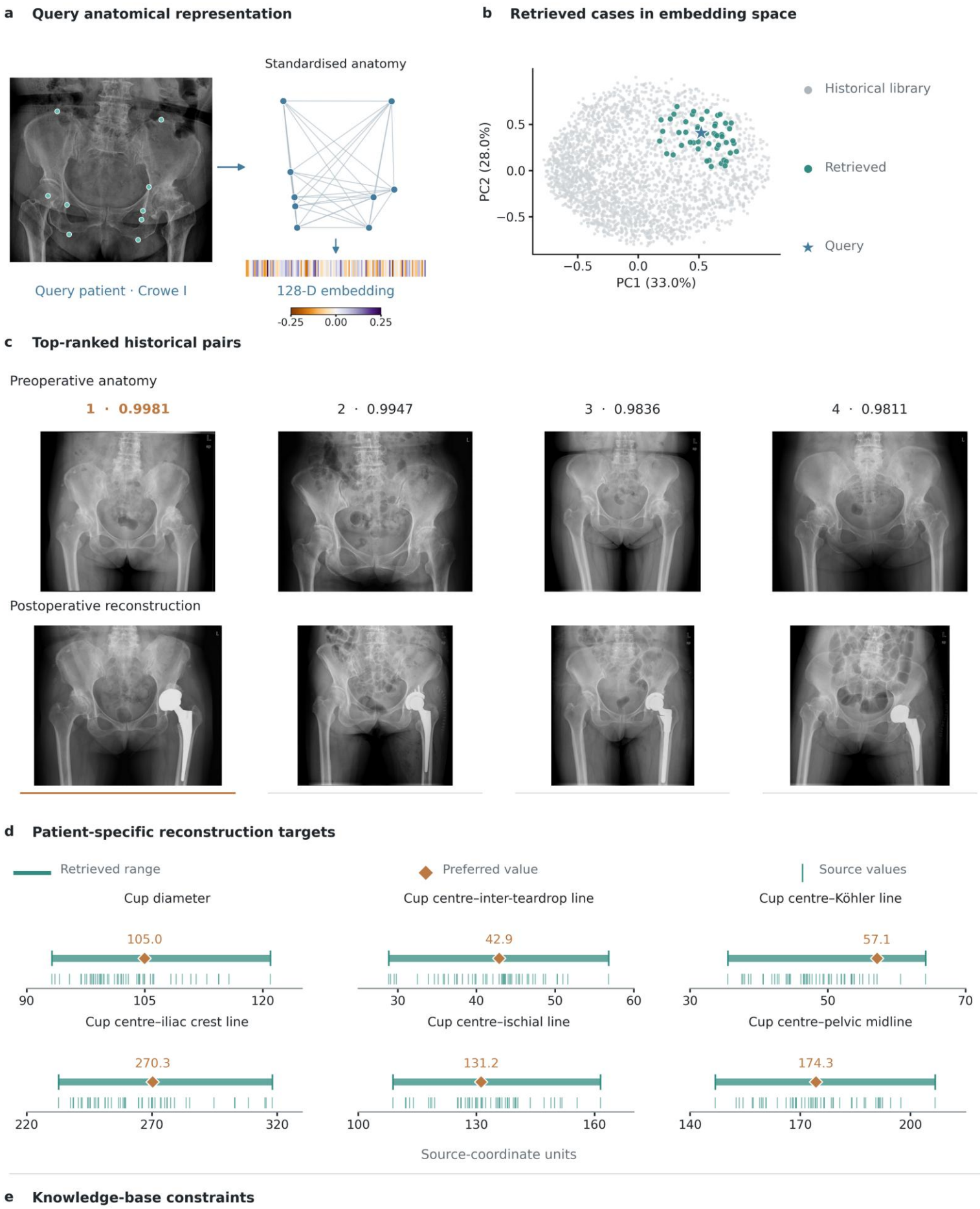
Historical retrieval constructs patient-specific targets

A representative query illustrates the transition from anatomical representation to reconstruction targets (Fig. 5). Its nine-landmark graph was encoded as a 128-dimensional representation and compared with historical representations using cosine similarity. The exported neighbourhood contained fifty ranked records, with the three highest similarities of 0.9981, 0.9947 and 0.9836. Ranking used the original embedding space; the principal-component projection visualises the query and neighbourhood within the historical distribution. Paired preoperative and postoperative radiographs for the four highest-ranked cases show the anatomy retrieved and the corresponding reconstructions used to construct targets.

Fifty numeric values were available for each of six target variables: cup diameter and cup-centre distances to the inter-teardrop, Köhler, iliac crest and ischial lines and the implemented iliac–ischial midpoint reference. This yielded 300 historical values for the illustrated query. The minimum and maximum retrieved values defined each target interval, while the highest-ranked record supplied the preferred value. For example, the cup-diameter interval was 93.227–120.983 with a preferred value of 104.99, and the inter-teardrop-distance interval was 28.85–56.80 with a preferred value of 42.90. These values are expressed in source coordinates, not millimetres. All six variables had numeric observations for every retrieved record, and the exported target summaries recorded no default-target substitution for this example.

The retrieved range and rank-1 preferred value provide complementary guidance: the range describes variation among historical reconstructions, and the preferred value identifies a specific solution within that neighbourhood. The interval is a historical range, not a confidence interval. Candidates are then generated and assessed in the current anatomy. The example therefore traces how historical reconstructions inform target construction and candidate evaluation, with the query's postoperative measurements reserved for evaluation.

Figure 5 | Retrieval-derived targets and knowledge-base constraints for acetabular reconstruction planning



$L_i = \min_j y_{ji}, \quad U_i = \max_j y_{ji}$

$t_i = y_{1i}$

$L_i \le x_i \le U_i \quad (i = 1, \ldots, 6)$

$f_i = \mathrm{clip}\left(1 - \frac{2|x_i - t_i|}{U_i - L_i}, 0, 1\right)$

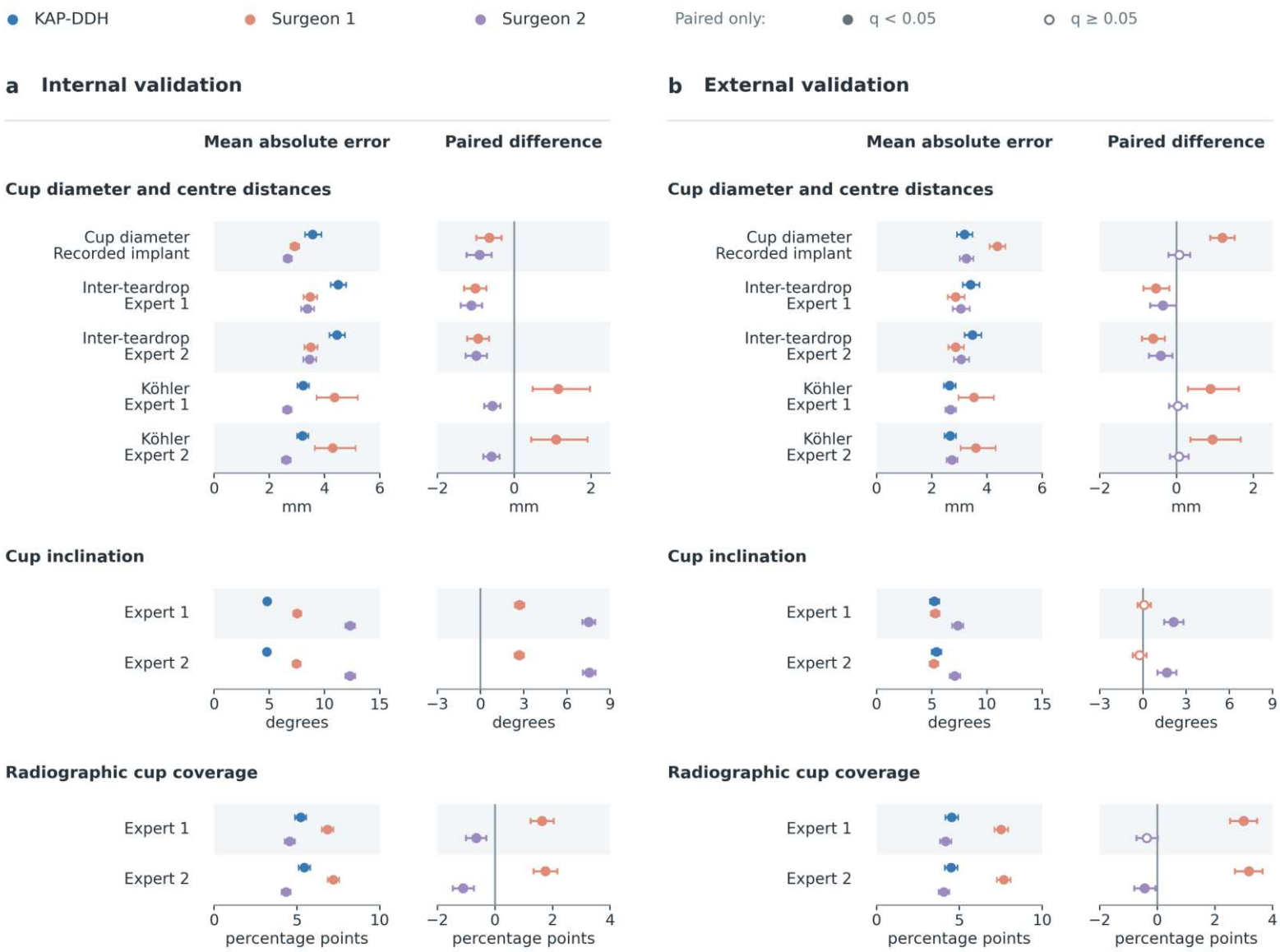
$K = 0.25 f_1 + 0.15(f_2 + \cdots + f_6)$

a, Query radiograph with supplied project landmarks, standardised anatomical graph and 128-dimensional embedding. b, Principal-component projection of 2,760 historical records and the query; 50 retrieved records are highlighted. Retrieval uses cosine similarity in the original 128-dimensional space. c, Preoperative and postoperative image pairs for retrieval ranks 1–4; column labels give rank and 128-dimensional cosine similarity. d, Six target distributions from all 50 retrieved records. Lines show minimum-to-maximum ranges, ticks show individual values, and diamonds show rank-1 preferred values. The six target parameters retain the raw image-coordinate values stored in the postoperative parameter table; no conversion to millimetres or pelvic-height normalisation is applied. e, For parameter i and retrieval rank j, y_{ji} defines bounds L_i and U_i and preferred value t_i; x_i denotes the candidate value. The six-parameter range check and clipped fit f_i are shown. K is the knowledge-base fit component, with diameter weighted 0.25 and five distances 0.15 each; it is not the full planning score.

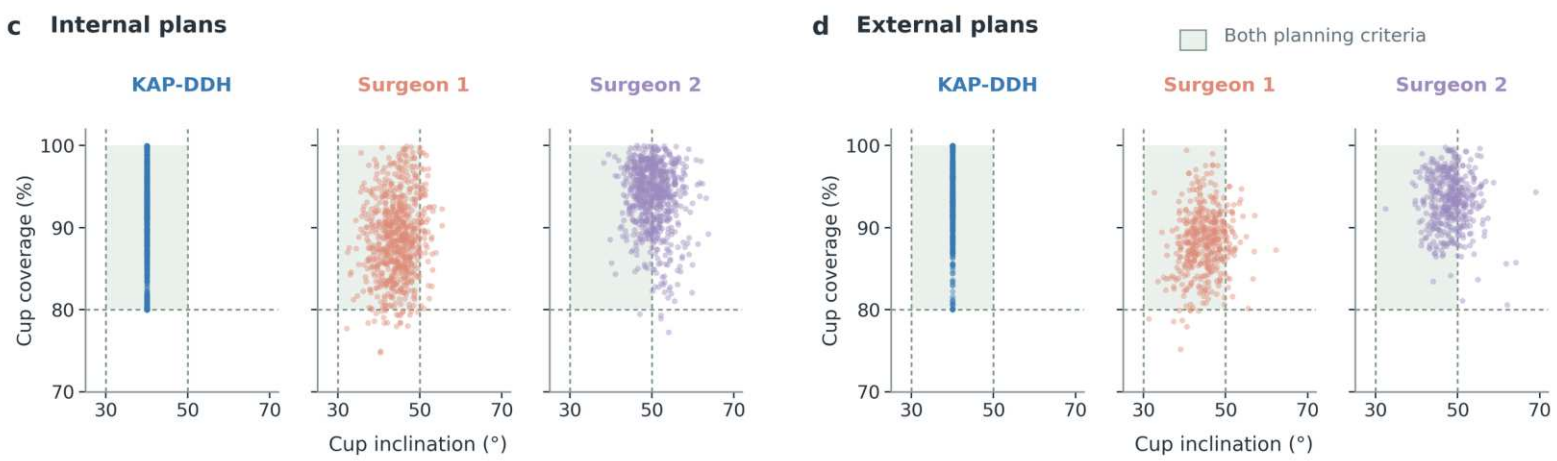
Planning agreement with postoperative reconstruction

We compared KAP-DDH with Surgeon 1 and Surgeon 2 on the same records for each endpoint, using recorded implanted diameter or independently measured postoperative references (Fig. 6). All continuous and criterion-based planning analyses included 691 internal and 422 external records. Positive paired differences denote surgeon absolute error minus KAP-DDH absolute error, indicating closer agreement between KAP-DDH and the reference. Confidence intervals accounted for patient clustering. This paired evaluation distinguished relative performance from each planner's MAE.

Figure 6 | Planning agreement with postoperative reconstruction and satisfaction of planning criteria



Paired difference = surgeon MAE − KAP-DDH MAE; positive values favour KAP-DDH



e Agreement and planning-criterion rates

		Cup diameter agreement		Planning criteria			
		Within ±2 mm	Within ±4 mm	Inclination 30–50°	Coverage ≥80%	Both	%
Internal	KAP-DDH	36.8	64.5	100.0	100.0	100.0	100
	Surgeon 1	38.8	71.6	92.3	96.7	89.0	50
	Surgeon 2	43.7	75.3	46.3	99.4	46.2	0
External	KAP-DDH	38.6	68.7	100.0	100.0	100.0	
	Surgeon 1	22.5	49.1	90.3	98.1	88.4	
	Surgeon 2	37.4	65.6	71.1	100.0	71.1	

a,b, Internal validation (691 hip records from 672 patients; a) and external validation (422 hip records from 362 patients; b). Left columns show mean absolute errors (MAEs) for KAP-DDH, Surgeon 1 and Surgeon 2. Cup diameter is compared with the recorded implant; cup-centre distances to the inter-teardrop and Köhler lines, inclination and radiographic coverage are compared with the indicated postoperative Expert 1 or Expert 2 measurements. Right columns show the mean within-record paired difference in absolute error (surgeon minus KAP-DDH); positive values indicate smaller KAP-DDH errors, and negative values indicate smaller surgeon errors. Whiskers are pointwise 95% confidence intervals from 10,000 patient-cluster bootstrap resamples. Filled and open markers in the paired-difference columns denote exploratory Benjamini–Hochberg-adjusted $q < 0.05$ and $q \geq 0.05$, respectively, across 36 contrasts. These q values are based on two-sided patient-block sign-flip tests using 20,000 randomisations and a plus-one correction; the confidence intervals are not multiplicity-adjusted. c,d, Planned inclination and radiographic cup coverage for each hip record, separated by planner, for internal (c) and external (d) validation. Each point represents one hip record; overlapping points are retained, and no jitter is applied. Shading indicates simultaneous satisfaction of inclination 30–50° inclusive and coverage ≥80%; dashed lines mark the thresholds. The near-vertical KAP-DDH distributions reflect planned inclination values close to 40°. e, Descriptive percentages of hip records meeting cup-diameter agreement thresholds or planning criteria. Cell values are percentages, displayed to one decimal place on a common 0–100% colour scale. No significance tests are encoded for these percentages. KAP-DDH satisfies both planning criteria in 691/691 internal and 422/422 external records. All outcomes use the same complete 691 internal and 422 external hip records across planners. All errors, paired differences and threshold-based rates use full-precision inputs; rounding is applied only for display. Satisfaction of planning criteria and agreement with postoperative reconstruction assess distinct aspects of performance.

For cup diameter, the internal KAP-DDH MAE was 3.56 mm, compared with 2.92 and 2.66 mm for Surgeon 1 and Surgeon 2. Paired differences were −0.65 mm (95% CI, −0.99 to −0.33) and −0.90 mm (−1.24 to −0.59), respectively. Externally, MAEs were 3.18, 4.37 and 3.25 mm. The corresponding differences were 1.19 mm (0.87–1.51) versus Surgeon 1 and 0.07 mm (−0.21 to 0.35) versus Surgeon 2.

Internal positional and angular results also varied by endpoint. KAP-DDH MAEs against Expert 1 were 4.49 mm for inter-teardrop distance, 3.22 mm for Köhler distance, 4.80° for inclination and 5.21 percentage points for coverage. Inter-teardrop-distance errors were greater for KAP-DDH than for either surgeon, with paired differences of −1.01 mm (95% CI, −1.30 to −0.73) and −1.12 mm (−1.40 to −0.84). Köhler-distance error was smaller than Surgeon 1's error (1.14 mm; 0.48–1.97) but greater than Surgeon 2's (−0.57 mm; −0.78 to −0.36). Inclination errors were smaller for KAP-DDH than for both surgeons, with differences of 2.70° (2.39–3.02) and 7.51° (7.07–7.96). Coverage differences also varied by comparator: 1.63 percentage points (1.23–2.03) versus Surgeon 1 and −0.65 percentage points (−1.01 to −0.31) versus Surgeon 2.

Against Expert 1, external KAP-DDH MAEs were 3.41 mm for inter-teardrop distance, 2.65 mm for Köhler distance, 5.23° for inclination and 4.53 percentage points for coverage. Compared with Surgeon 1, KAP-DDH had smaller errors for Köhler distance (0.88 mm; 95% CI, 0.29–1.62) and coverage (2.99 percentage points; 2.52–3.46), but greater error for inter-teardrop distance (−0.54 mm; −0.86 to −0.19). Inclination errors differed little between them (0.08°; −0.38 to 0.55). Compared with Surgeon 2, KAP-DDH had smaller inclination error (2.12°; 1.48–2.80) and greater inter-teardrop-distance error (−0.35 mm; −0.69 to −0.01; q = 0.049). Intervals for Köhler distance and coverage included zero.

Using Expert 2 as the reference preserved the direction of all internal image-derived comparisons, including lower KAP-DDH inclination error and higher inter-teardrop-distance error than both surgeons. Externally, the near-zero inclination difference versus Surgeon 1 changed sign, from 0.08° with Expert 1 to −0.24° (95% CI, −0.70 to 0.25) with Expert 2; neither interval excluded zero. The external coverage comparison with Surgeon 2 was −0.44 percentage points (−0.81 to

−0.07) with Expert 2, with an exploratory adjusted q value of 0.028. With Expert 1, the estimate was −0.37 percentage points (−0.73 to 0.01; q = 0.070). The choice of reference thus particularly affected the interpretation of differences near zero, while the larger internal inclination differences retained their direction (Fig. 6a,b).

Cup-size agreement within ±4 mm was 64.5%, 71.6% and 75.3% internally, and 68.7%, 49.1% and 65.6% externally, for KAP-DDH, Surgeon 1 and Surgeon 2, respectively. Externally, the paired difference versus Surgeon 1 was 19.67 percentage points (95% CI, 13.89–25.48): 125 records met this threshold only with KAP-DDH and 42 only with Surgeon 1. Versus Surgeon 2, the difference was 3.08 percentage points (−2.66 to 8.92). At the ±2-mm threshold, external KAP-DDH agreement was 38.6%; the difference from Surgeon 1 was 16.11 percentage points (10.24–22.03), whereas the difference from Surgeon 2 was 1.18 percentage points (−4.68 to 6.94). The two external surgeon comparisons thus differed at both thresholds. The tighter threshold also showed why similar aggregate rates need not imply the same record-level outcomes. Internally, KAP-DDH and Surgeon 1 rates were 36.8% and 38.8%, yet 156 records met the threshold only with KAP-DDH and 170 only with Surgeon 1 (Supplementary Table 13).

The inclination criterion of 30–50° and coverage criterion of at least 80% were jointly met in 691/691 internal and 422/422 external KAP-DDH plans. Corresponding counts were 615/691 and 373/422 for Surgeon 1, and 319/691 and 300/422 for Surgeon 2. Internally, 76 records met both criteria with KAP-DDH alone in the comparison with Surgeon 1, and 372 in the comparison with Surgeon 2, giving paired differences of 11.00 and 53.84 percentage points. Externally, the corresponding counts were 49 and 122. No record met both criteria only with a surgeon. Empirical-bootstrap intervals were omitted where a marginal rate was at the boundary, according to the analysis rule described in Methods. These rates describe fulfilment of the planning thresholds; errors relative to postoperative reconstruction are reported separately.

Exploratory multiplicity-adjusted comparisons supported the internal cup-diameter differences favouring the surgeons (both $q < 0.001$) and the external difference favouring KAP-DDH over Surgeon 1 ($q < 0.001$), but not the external comparison with Surgeon 2 ($q = 0.655$). Continuous and

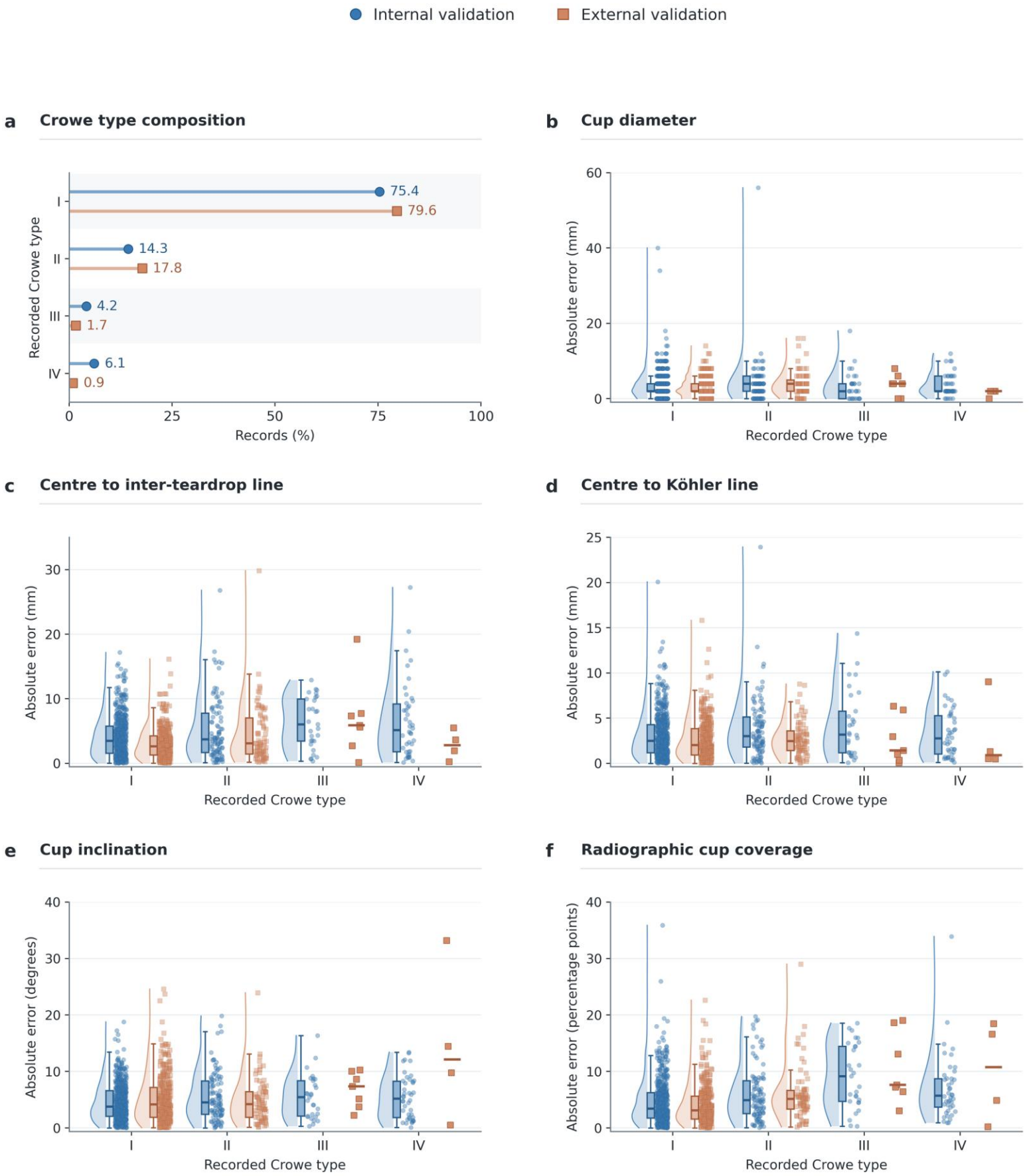
criterion-based analyses used the same full-precision measured outputs. Planned cup diameters were compared with recorded implant sizes without further rounding to standard implant sizes.

Crowe-stratified evaluation delineates the validation spectrum

The internal cohort included 521 Crowe I, 99 Crowe II, 29 Crowe III and 42 Crowe IV records; the corresponding external counts were 336, 75, 7 and 4 (Fig. 7a). Both cohorts therefore covered the full classification range but provided different amounts of evidence for severe dysplasia. Figure 7b–f shows separate cohort distributions of record-level cup-size, positional, inclination and coverage errors. For groups with at least ten patients, distribution shapes and box plots accompany individual observations; smaller groups show individual points and medians.

External Crowe III and IV results comprised seven and four records, respectively, and are presented as individual observations. Supplementary Fig. 6 repeats the stratified assessment using Expert 2 as the postoperative reference, retaining the same records and Crowe assignments.

Figure 7 | KAP-DDH planning errors across recorded Crowe types

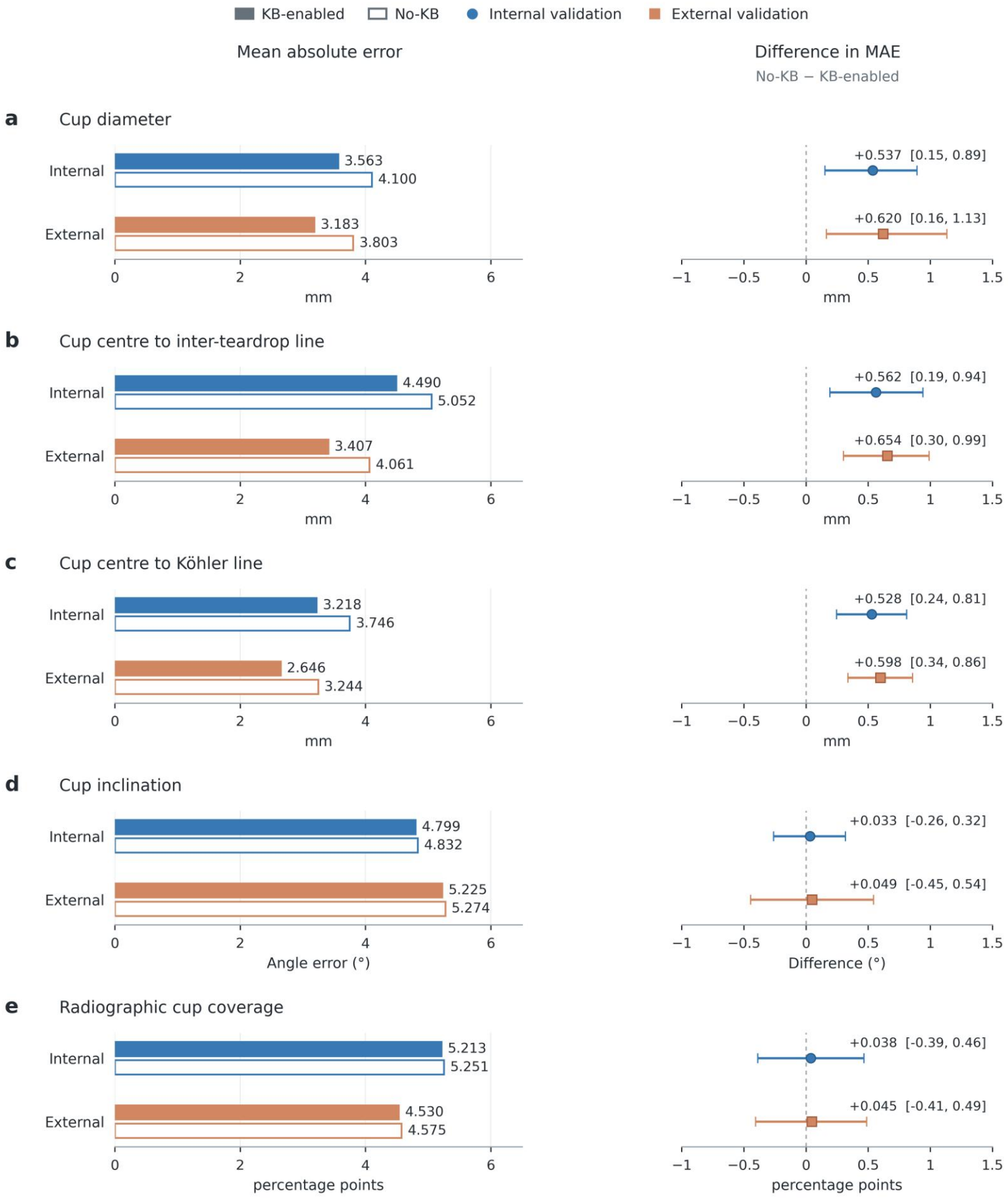


Internal validation comprised 691 hip records from 672 patients, and external validation comprised 422 hip records from 362 patients. a, Horizontal lollipop plots show within-cohort percentages of records in Crowe types I–IV. The corresponding counts are 521, 99, 29 and 42 internally, and 336, 75, 7 and 4 externally. b–f, Absolute differences between KAP-DDH plans and the reference for cup diameter (b), cup-centre distance to the inter-teardrop line (c), cup-centre distance to the Köhler line (d), cup inclination (e) and radiographic cup coverage (f). Diameter uses the recorded implant as the reference; the remaining endpoints use postoperative Expert 1 measurements. Blue circles and orange squares represent individual internal and external hip records, respectively. Horizontal jitter separates overlapping observations without changing their error values. Half-violins show descriptive kernel density estimates (Scott bandwidth), each scaled to the same maximum width and restricted to its observed range. Boxes show the median and interquartile range, with whiskers extending to the most extreme observations within 1.5 interquartile ranges of the box. Strata with fewer than ten patients show individual records and the median only. All eligible records and the full error range are retained. Absolute errors are calculated from full-precision planned and reference values; rounding is applied only for display. These stratified distributions are descriptive. The small external Crowe III and IV groups do not establish equivalent performance across types. The corresponding Expert 2 analysis is provided in Supplementary Figure 6.

Knowledge-base-enabled and no-KB comparison

Knowledge-base-enabled and no-KB plans were compared on the same 691 internal and 422 external records (Fig. 8). The no-KB configuration substituted Crowe-specific default targets for retrieved targets, retaining the perception, candidate-generation, geometric-checking and ranking procedures (Methods, Knowledge-base comparison configuration; Supplementary Tables 9 and 10). Positive paired differences denote no-KB absolute error minus knowledge-base-enabled absolute error.

Figure 8 | Planning errors with and without knowledge-base guidance



a-e, Cup diameter (a), cup-centre distance to the inter-teardrop line (b), cup-centre distance to the Köhler line (c), cup inclination (d), and radiographic cup coverage (e). Left, filled and open bars show mean absolute error (MAE) for KB-enabled and no-KB planning, respectively. Blue represents internal validation, and orange represents external validation. Right, circles (internal) and squares (external) show the mean paired difference in absolute error, calculated as no-KB minus KB-enabled. Whiskers and bracketed values show pointwise 95% confidence intervals (CIs). A positive difference favours KB-enabled planning; the dashed line marks zero difference. Units are millimetres for a-c, degrees for d, and percentage points for e. Cup diameter uses the recorded implant diameter as the reference; panels b-e use postoperative Expert 1 measurements. Internal validation includes 691 hip-level observations from 672 patients; external validation includes 422 hip-level observations from 362 patients.

Absolute errors and paired differences were calculated using full-precision planning and reference values. CIs were obtained from 10,000 patient-cluster bootstrap resamples within each cohort, retaining all records and matched planning/reference values for each sampled patient, with percentile limits at 2.5% and 97.5%. Estimates are weighted by hip record. Exploratory two-sided patient-block sign-flip tests used 20,000 randomisations with a plus-one correction; the Benjamini-Hochberg procedure was applied separately within each cohort across nine comparisons (one implant-based diameter comparison and four outcomes with each of two expert references). These tests assume joint sign symmetry within patient blocks.

For cup diameter, internal MAEs were 3.56 mm with knowledge-base guidance and 4.10 mm without it; the paired difference was 0.54 mm (95% CI, 0.15–0.89; q = 0.009). External MAEs were 3.18 and 3.80 mm, respectively, with a difference of 0.62 mm (0.16–1.13; q = 0.017). Against Expert 1, internal inter-teardrop and Köhler distance errors were 4.49 and 3.22 mm with guidance, compared with 5.05 and 3.75 mm without it. The paired improvements were 0.56 mm (0.19–0.94; q = 0.009) and 0.53 mm (0.24–0.81; q = 0.003). Corresponding external errors were 3.41 versus 4.06 mm and 2.65 versus 3.24 mm, giving differences of 0.65 mm (0.30–0.99; q < 0.001) and 0.60 mm (0.34–0.86; q < 0.001).

The Expert 2 comparisons also showed smaller positional errors with knowledge-base guidance in both cohorts (Supplementary Fig. 7). Inclination and coverage errors changed little: against Expert 1, paired differences were 0.03° and 0.05° for inclination and 0.04 and 0.05 percentage points for coverage internally and externally, respectively. Their confidence intervals included zero with either expert reference. The improvement was concentrated in cup-size and centre-position agreement, the outputs directly addressed by the retrieved targets.

Discussion

KAP-DDH connects radiographic anatomy to an explicit reconstruction proposal through historical-case retrieval and geometric planning. Across internal and external cohorts, the evaluation traced this process from anatomical perception to comparison with postoperative reconstruction and independent surgeon plans. The results identify where historical guidance contributes most: replacing Crowe-specific defaults with retrieved targets reduced cup-size and centre-position errors, whereas inclination and coverage errors changed little. Comparisons with surgeons also varied by endpoint and cohort. Together, these findings support a planning approach in which prior reconstructions inform patient-specific targets and geometric rules determine how those targets can be used in the current hip.

The anatomical evaluations provide a basis for interpreting those planning results (Figs. 3 and 4). Segmentation supplies the geometry used to derive landmarks, reference lines and distances; errors can therefore propagate beyond the region in which they arise. A small boundary

displacement near a landmark may affect a cup-centre distance even when overlap across the larger region remains high. Conversely, a measurement defined by two nearby structures can remain relatively stable if their localisation errors occur in a similar direction. These are geometric explanations for why Dice, point error and measurement error should be assessed separately, consistent with recommendations for image-analysis validation.[30,31] In our data, pelvic-height and related vertical-distance measurements had relatively small errors despite larger average point-localisation errors. Direct validation of the derived quantities is therefore particularly relevant to a system that uses anatomical geometry for subsequent decisions.

Postoperative perception achieved higher region-overlap estimates than preoperative perception against Expert 1. The more distinct implant contour and regular cup geometry may contribute to this pattern, although this study did not isolate the effects of image contrast or implant shape. Earlier work on automated postoperative component measurement and arthroplasty radiography registries supports the feasibility of extracting reconstruction geometry from these images.[23,32] In our evaluation, image-derived postoperative measurement errors were smaller than the corresponding planning-to-postoperative discrepancies. This distinction helps locate the task being assessed: recovering the observed reconstruction and proposing a reconstruction from preoperative information require different inferences. This validation establishes the performance of automated postoperative assessment; the subsequent planning comparisons use the separate Expert 1 and Expert 2 measurements as their image-derived references.

The second expert reference adds clinical context to this evaluation. Expert 1 provided development annotations as well as the primary reference, so closer agreement with that expert can partly reflect consistency with the annotation conventions learned during development. Against Expert 2, segmentation agreement was generally close to the interobserver benchmark, supporting useful transfer beyond the primary annotation source. Measurement bias and limits of agreement complement these overlap comparisons by showing systematic differences and their spread in clinically interpretable units. Reference choice mattered most for planning contrasts near zero, including external inclination versus Surgeon 1 and coverage versus Surgeon 2. The larger internal

inclination differences retained their direction. This pattern supports interpreting the magnitude and stability of each comparison alongside its statistical significance.

The retrieval example shows how anatomical measurements become planning information (Fig. 5). Crowe type describes displacement severity, but it does not fully represent the shape and distribution of acetabular deficiency. Morphological studies describe variation within low and high dislocations,[9] and three-dimensional analyses identify distinct patterns of anterior, posterior and global deficiency.[33] In implantation simulations, Crowe II/III hips had lower integrated coverage than Crowe I and IV hips, further showing that displacement severity alone does not specify available support.[5] A landmark representation can refine the search within this broad clinical classification by encoding spatial relationships across the pelvis. Historical pairing then adds the reconstruction associated with each retrieved anatomy. This is the relevant link to knowledge-based planning and case-based reasoning: the value of retrieval lies in supplying information that can be adapted to a new decision.[27–29] In KAP-DDH, the retrieved intervals and preferred values make that adaptation explicit, while the associated image pairs allow a clinician to inspect the examples behind the numerical targets.

This design complements existing AI planning approaches. CT-based studies have evaluated implant sizing and three-dimensional reconstruction in DDH and other THA populations,[19–21,34,35] while THA-Net generates plausible postoperative radiographs from preoperative inputs.[25] A recent paediatric DDH system combines keypoint detection with rule-based osteotomy and plate recommendations.[26] KAP-DDH addresses adult acetabular reconstruction through a different intermediate representation: historical cup-size and positional measurements define targets, and deterministic geometry generates and ranks candidates. The distinction matters clinically because the anatomical input, retrieved examples and candidate assessment can each be reviewed. The framework also separates tasks that may require different improvements: better perception for uncertain boundaries, a more representative historical collection for uncommon anatomy, or revised constraints for a different reconstruction objective.

The surgeon comparisons illustrate the competing objectives of dysplastic acetabular reconstruction (Fig. 6). Both surgeons matched implanted cup diameter more closely internally, whereas externally KAP-DDH had smaller cup-diameter errors than Surgeon 1 and differed little from Surgeon 2. KAP-DDH also showed greater inter-teardrop-distance error than both surgeons against Expert 1 in both cohorts, despite smaller errors for some other endpoints. Hip-centre height is a particularly consequential choice: moving the cup changes bone support and reconstruction biomechanics, while medialisation can affect the available host-bone coverage.[4,8,36–38] Thus, a plan prioritising one set of size and coverage targets may differ from the operative centre selected for another balance of objectives. Differences in case mix and surgeons' reconstruction preferences could also contribute to the cohort-specific pattern, but were not tested here. The clinical message is to inspect the proposed size and centre jointly, rather than infer overall planning quality from either endpoint alone.

Inclination and coverage show a different aspect of the system. The near-vertical inclination distributions in Fig. 6 reflect candidate construction close to 40°, and coverage is constrained during candidate evaluation. Joint fulfilment of the prespecified criteria in all evaluated KAP-DDH plans is therefore consistent with the implemented geometry. It demonstrates reproducible application of those rules, alongside the separate analysis of agreement with postoperative reconstruction. Actual operative orientation may also reflect considerations outside an anteroposterior image, including component impingement and functional pelvic position.[39–43] For example, changes between supine and standing posture can alter apparent cup orientation and acetabular coverage.[42,43] A radiographic proposal can establish an initial set of measurements for review, but orientation decisions in clinical use also require the relevant functional and three-dimensional information.

The Crowe-stratified results help define the population to which these findings apply (Fig. 7). Most validation records were Crowe I or II, whereas the external cohort contained only seven Crowe III and four Crowe IV records. Severe displacement is not simply a larger version of mild dysplasia: studies of high dislocation describe distinctive acetabular dimensions and bone distribution that complicate cup placement.[6,44,45] Sparse examples may also limit the variety of historical solutions

available for retrieval, a possibility that requires direct testing. The current displays show the range of individual errors across the represented spectrum, but larger severe-dysplasia cohorts are needed to estimate type-specific performance and to examine retrieval and fallback behaviour in those hips.

The within-pipeline comparison helps explain the endpoint-specific results (Fig. 8). Historical targets directly specify cup diameter and five centre-position distances. The observed reductions in diameter and positional errors are consistent with those targets capturing reconstruction choices that Crowe-specific defaults cannot represent for every hip. Inclination and coverage, by contrast, are governed largely by procedures shared between the two configurations, providing a plausible explanation for their small changes. Because retrieved targets can change which candidates are feasible and ultimately selected, the measured effect concerns the complete target-guided planning process. The similar direction of the positional findings with both expert references strengthens this interpretation. The absolute improvements were modest and should be understood as closer retrospective reconstruction agreement; their value for operative decisions remains to be established.

Radiographic planning offers accessibility, but its geometry must be interpreted at the appropriate scale. Standardising a landmark graph allows anatomical comparisons; physical calibration separately determines how image measurements translate into millimetres.[46] In the present implementation, historical targets retain their source-coordinate convention, so consistent target and query scaling remains important when applying the workflow to new acquisition settings. Projection-based coverage also describes a different quantity from three-dimensional bone contact. Comparisons against intraoperative measurements and three-dimensional reconstructions demonstrate the importance of that distinction.[14–16] CT-based analysis remains valuable when acetabular version, regional bone stock or fixation cannot be adequately characterised on the radiograph.[5,6,33,34] The useful role of the present approach is to organise a radiographic reconstruction proposal and its supporting cases within that broader clinical assessment.

Several features of the design govern interpretation. The study was retrospective, and postoperative reconstruction records the treatment actually performed, including intraoperative judgement and implant availability.[47] It is a practical templating reference rather than an independently established optimal plan. Development and internal-validation cohorts shared 133 patients; patient-cluster resampling accounts for dependence within the evaluated records but does not remove potential information overlap with development.[48] The separate-institution cohort therefore provides the stronger assessment of transfer. The historical collection also reflects the practice of its contributing institution over a long acquisition period. Evaluation at additional centres, with clearly separated development populations and more severe dysplasia, would clarify how case mix, imaging conventions and surgical practice affect the proposed targets. The present study assessed radiographic outputs, without measuring planning time, operative changes or patient outcomes.

The next clinical question is how surgeons use the proposal. Studies of clinical decision aids and human–AI collaboration show that the effect of assistance depends on how information is presented and incorporated into decisions.[49,50] A prospective KAP-DDH evaluation could examine planning time, agreement before and after assistance, the reasons for modifying or rejecting a proposal, and whether those changes influence operative reconstruction. DECIDE-AI and FUTURE-AI provide frameworks for studying such use and its failure modes.[51,52] The current validation used automatic Top-1 outputs; interactive refinement demonstrated in the interface was a separate capability. By linking current anatomy, historical examples and explicit planning targets, KAP-DDH provides a concrete basis for evaluating this clinician–system workflow in DDH reconstruction.

Methods

Study design, cohorts and ethics

We conducted a retrospective two-centre development and validation study in patients undergoing total hip arthroplasty for osteoarthritis secondary to developmental dysplasia of the hip (DDH). Paired preoperative and postoperative anteroposterior pelvic radiographs were obtained from West China Hospital (January 2000–January 2025) and Sichuan Provincial Orthopedic Hospital

(January 2021–April 2026). Eligibility required paired images, adequate pelvic visualisation, essential clinical information and reference annotations. Exclusion criteria were incomplete paired imaging, severe occlusion or incomplete pelvic visualisation, missing essential information, unsuccessful image retrieval or export, and missing required annotations.

West China Hospital provided 2,760 operated-hip records from 2,492 patients for development and the historical-knowledge workflow, and 691 records from 672 patients for internal validation. External validation comprised 422 records from 362 patients at Sichuan Provincial Orthopedic Hospital. Across cohorts, 3,873 records represented 3,393 unique patients. The development and internal-validation cohorts shared 133 patients. Records and patients were counted separately, and uncertainty estimates accounted for clustering of evaluated records within patients. Cohort membership and endpoint-specific denominators are summarised in Table 1 and Supplementary Table 1.

The record-to-patient crosswalk preserves paired preoperative and postoperative radiographs and links records from the same patient. Eight patients in the external cohort each contributed two same-side records. Original cohort membership and patient linkage were retained. The hip record is the unit of observation for anatomical and planning analyses; resampling preserves patient-level dependence.

Each patient was counted once per cohort when summarising baseline characteristics. Age and body-mass index were taken from the operation with the smallest recorded sequence number; BMI was calculated as weight in kilograms divided by height in metres squared. Sex was summarised by patient, and side and Crowe type[13] by record. Age and body-mass index data were complete after source-data verification. Ethics approval was recorded under number 2026 No. 727, with a waiver of individual informed consent. Records were de-identified before analysis.

Radiographic annotation and measurement definitions

Expert 1 provided manual region annotations for development and as the primary anatomical reference. Expert 2 independently annotated both validation cohorts without access to Expert 1 annotations or KAP-DDH outputs. Region labels provided supervision for the image model;

landmarks and radiographic measurements were derived through deterministic processing of region and implant geometry. The automated outputs used for evaluation were not manually corrected.

Preoperative perception comprised ten regions, nine landmarks and twelve linear measurements. Postoperative representation comprised eleven regions, fourteen geometric point inputs and nineteen constructions: seventeen linear quantities, inclination and radiographic coverage. Eighteen postoperative measurements were evaluated for image-measurement accuracy; geometric cup-diameter construction M7 was excluded. The clinically recorded implanted diameter served as a separate reference for planning evaluation. Region, point and measurement definitions are provided in Supplementary Tables 2–5. Preoperative measurements quantify pelvic reference-line lengths, pelvic height, distances between pelvic levels, pelvic diagonals and contralateral head-centre distances to the three bilateral reference lines. Postoperative constructions also describe acetabular roof geometry, cup position and contralateral head-centre position. Point numbers therefore identify task-dependent geometry and are not interchangeable anatomical labels before and after surgery.

The postoperative calculation inputs Q1–Q14 map to validation identifiers P1–P14 in Supplementary Table 3; Q11/P11 denotes the cup centre. Geometric cup diameter is the distance between Q9 and Q10. Medialisation is calculated by subtracting the cup radius from the cup-centre-to-Köhler-line distance. Inclination is measured relative to the inter-teardrop line. Coverage uses the implemented two-dimensional angular measure. Exported linear measurements use image-specific scale factors for reporting in millimetres; angles are expressed in degrees, coverage on a 0–100 percentage scale and coverage errors in percentage points. Physical calibration is distinct from computational PNG grid spacing. Missing point inputs retain their index positions, and dependent calculations remain blank. The geometric code rounds Euclidean and midpoint distances, medialisation, midline distance and acute angles to three decimal places; direct point-to-line distances are not rounded at this stage. Coverage is calculated from the rounded angle and rounded to two decimal places. These geometric calculation rules determine the stored measurements; subsequent statistical comparisons use their full stored precision without further rounding.

Postoperative information from a validation record was used for evaluation, whereas postoperative data from retrieved historical records provided planning knowledge.

Anatomical perception architecture and implementation

Separate right- and left-sided preoperative and postoperative tasks used nnUNetv2 [53,54] with the project-specific nnUNetTrainerSPMImproved. Saved training plans and matching internal/external inference configurations specify three-channel PNG input, cropping to the nonzero bounding box, per-channel z-score normalisation without a foreground normalisation mask, and resampling according to the plans. Preoperative training patches are 512 × 512 pixels with a batch size of 12; postoperative patches are 448 × 512 with a batch size of 14. Patch dimensions refer to training inputs rather than original radiograph dimensions. Saved dataset metadata specify 1,512 right-sided and 1,248 left-sided training images for each temporal task, consistent with the 2,760-record development population; preoperative and postoperative images belong to the same paired records.

The backbone uses eight encoder stages preoperatively and seven postoperatively, with feature channels increasing from 32 to a maximum of 512. Each encoder and decoder stage contains two 3 × 3 convolutions, instance normalisation and LeakyReLU activation. ImprovedSPM introduces learned class-specific shape priors at the bottleneck and second decoder stage. Global and local priors are represented on 16 × 16 grids and refined through two self-attention[55] update blocks. Channel projections, element-wise feature–prior modulation and residual updates couple the priors to image features. Attention gates regulate the first two decoder skip connections; a learned sigmoid weight combines global and local priors for the second decoder-stage update.

The trainer configuration specifies 700 epochs, with 250 training iterations and 50 validation iterations per epoch. Optimisation uses stochastic gradient descent with Nesterov momentum of 0.99, an initial learning rate of 0.01, weight decay of 3×10^{-5} and polynomial decay with an exponent of 0.9. Equally weighted cross-entropy and soft-Dice losses supervise mutually exclusive label maps. Deep-supervision weights decrease with output resolution and are normalised after the coarsest output is omitted. The trainer inherits the no-data-augmentation variant, uses validation transforms during training and disables inference mirroring. These schedules specify the configured

training procedure; completion of every scheduled epoch in each historical run was not established from the retained records. Supplementary Table 6 lists the dataset identifiers and saved training and inference configurations. Segmentation outputs undergo deterministic geometric postprocessing without independent landmark or measurement prediction heads.

Anatomical representation and similar-case retrieval

The historical knowledge base links each operated-hip record's preoperative anatomy to its postoperative reconstruction. The development cohort defines the source population; searchable-index membership is recorded separately. Query inputs comprise nine preoperative landmarks and clinician-supplied Crowe type. Landmarks are centred on the bilateral teardrop midpoint, rotated to align the inter-teardrop line horizontally, and divided by the distance between the bilateral superior-iliac and ischial midpoints. Reflection and paired-landmark reordering standardise laterality. These dimensionless representation coordinates are distinct from both historical reconstruction-target coordinates and calibrated clinical outputs.

The input graph contains nine nodes and 72 directed non-self edges. The sixteen node features comprise two coordinates, a nine-dimensional landmark-identity one-hot vector and a five-dimensional anatomical-group one-hot vector. Four edge features encode horizontal and vertical displacement, Euclidean distance and shared anatomical group. An eight-dimensional trainable Crowe embedding is concatenated to every node. A GATv2 encoder [56] projects node features to 64 dimensions, applies four 64-dimensional attention heads followed by one 128-dimensional head, and combines four attention-pooling heads through a linear layer. The resulting case embedding has 128 dimensions and is L2-normalised; layer-level self-loops follow the implementation.

In the representation-training implementation, two independently augmented views of each record form positive pairs for a contrastive objective[57] with a temperature of 0.07. A training-only projection head has 128-dimensional hidden and output layers. Coordinate jitter of ±0.02 and scaling from 0.9 to 1.1 precede geometric standardisation. Sampling weights are proportional to the inverse square root of Crowe-type frequency. The configured batch size is 64 and the maximum training duration is 500 epochs; the Adam[58] learning rate decreases from 10^{-3} to 10^{-5} with 5% warm-up and

cosine decay. Checkpoint selection maximises the mean difference between top same-class similarity and its 90th percentile for up to three development queries per class, excluding self-matches. Evaluation occurs every five epochs and at the first and final epochs; early stopping follows eight evaluations without improvement after epoch 30.

Retrieval ranks embeddings by cosine similarity, a score in [−1, 1] rather than a probability. Same-Crowe and global search results are merged, deduplicated by case path and reranked without an absolute similarity threshold. The same-class search requests 50, 30, 20 or 10 records for class buckets of ≥500, 200–499, 100–199 or <100 records; the global search requests 50, 30 or 20 for buckets of ≤60, 61–150 or >150. Returned-case caps are 100, 50, 30 and 30 for Crowe I–IV, although actual counts can be smaller. Principal-component projection is used only to visualise retrieval results; ranking uses the original 128-dimensional embeddings (Supplementary Table 7).

For Fig. 5, the historical embedding matrix was projected onto principal components and the query was placed in the same fitted space. The four image pairs illustrate the highest-ranked historical cases; all 50 returned records contribute to the displayed target summaries. Retrieval ranking uses the original embeddings.

Historical reconstruction targets

Six quantities define patient-specific targets: cup diameter and unsigned perpendicular cup-centre distances to the inter-teardrop (IT), superior iliac (SI), ischial (IS), Köhler (KL) and iliac–ischial midpoint (IM) lines. The IM line joins the bilateral superior-iliac midpoint to the bilateral ischial midpoint and is not assumed to coincide with a separately defined pelvic midline. For each variable, the minimum and maximum available numeric values among retained historical records define the target interval, and the highest-ranked retrieved record provides the preferred value. The interval is therefore a historical range, not a confidence interval, and the preferred value is not a neighbourhood mean. A missing or zero-width interval, an unavailable top-ranked preferred value or retrieval failure triggers the corresponding prespecified Crowe-specific default; target weights remain unchanged.

Target construction and fit calculations preserve the historical table's source-coordinate convention. The implementation uses a target normalisation denominator of 1.0, independently of graph standardisation. Physical cup-size reporting uses image-specific calibration. Historical target variables and code mappings are listed in Supplementary Table 8, and defaults in Supplementary Table 10.

Candidate generation and deterministic ranking

Complementary angular-geometry and horizontal-position pathways generate candidate cup reconstructions. The implementation uses an orientation of approximately 40° relative to the inter-teardrop line. The angular pathway scans a secondary angle from 36° to 0° in 0.1° steps while sampling the lateral acetabular-edge curve; the horizontal pathway evaluates medial–lateral centre positions. Candidate assessment checks effective target intervals and pathway-specific geometry. Image calibration converts geometry to standard even cup diameters, with a configured minimum of 40 mm. The active candidate pathway applies an 80% threshold using its angle-based or horizontal projected-coverage calculation. Coverage and inclination are not weighted ranking components; inclination of 30–50° is an evaluation criterion. Saved planning outputs were evaluated as continuous measured values; statistical analysis did not convert them to nominal even implant sizes. For candidate value x_i, effective preferred value t_i and interval $[l_i, u_i]$, target fit is calculated after replacing a retrieved zero-width interval with its corresponding default:

$$f_i = \mathrm{clip}\left(1 - \frac{2|x_i - t_i|}{u_i - l_i}, 0, 1\right)$$

Clipping restricts fit to [0, 1]. Deviation from the preferred value is scaled by half the effective interval width. Interval membership is checked separately and does not confer a uniform fit.

The six fits define the target-fit component:

$$K = 0.25 f_{\mathrm{diam}} + 0.15\left(f_{\mathrm{IT}} + f_{\mathrm{SI}} + f_{\mathrm{IS}} + f_{\mathrm{KL}} + f_{\mathrm{IM}}\right)$$

Diameter has a weight of 0.25 and each positional target a weight of 0.15; the weights sum to 1. Subscripts follow Supplementary Table 8.

Positional stability uses the centres supplied to the scoring function:

$$P=\min\left\{1,\frac{N_5}{\max(1,0.2N)}\right\}$$

N_5 counts centres at a Euclidean distance strictly below 5 pixels, including the candidate itself; N is the number of centres supplied. Regular candidates are scored as singletons, so P = 1 and does not affect their ordering.

Geometric balance uses four positional fits:

$$B=1-\min\{1,\sigma_0(f_{\mathrm{IT}},f_{\mathrm{SI}},f_{\mathrm{IS}},f_{\mathrm{KL}})\}$$

Here σ_0 is the population standard deviation (ddof = 0); diameter and IM fits are excluded.

Candidates that share a sampled lateral acetabular-edge point form a local group of size n_g. Their standard even diameters determine the diameter bonus:

$$D=\begin{cases}\mathrm{clip}\left(\dfrac{d-d_{\min}}{d_{\max}-d_{\min}+\varepsilon},0,1\right), & n_g>1 \text{ and } d_{\max}-d_{\min}>\varepsilon,\\ 0, & \text{otherwise.}\end{cases}$$

Here $\varepsilon = 10^{-6}$; singleton or equal-diameter groups receive zero. Regular-candidate scores are (K + 0.30P + 0.20B)/1.50 + 0.20D. Removing the constant contribution of P = 1 gives the equivalent ordering expression:

$$S_{\mathrm{rank}}=\frac{K+0.20B}{1.50}+0.20D$$

This constant-shift equivalence applies only to regular candidates. Special fallback candidates retain the scores assigned by their implemented scoring paths. Scores are used for ranking and do not represent probabilities, confidence measures or accuracy measures.

Selection prioritises valid regular candidates. If none are available, selection from the full regular pool is permitted, and a separate special fallback may be generated. Ranking, merging of similar results and output selection follow the corresponding software path. Validity and fallback flags

remain distinct from rank. The evaluated output is the automatically saved Top-1 plan, without clinician adjustment.

Expert references and evaluation endpoints

Surgeon 1 and Surgeon 2 planned independently from preoperative radiographs alone, blinded to KAP-DDH outputs, one another's plans, postoperative images and postoperative reconstruction. Their planning role was distinct from that of the two expert reference annotators. Interactive refinement shown in the interface demonstration was not used to define the evaluated KAP-DDH outputs.

Anatomical perception was evaluated using the Dice similarity coefficient for regions, mean radial error for landmarks and mean absolute error (MAE) for measurements. Radial error was the Euclidean distance between predicted and reference locations in millimetres. For predicted and reference regions A and B, Dice was twice their intersection divided by the sum of their areas. For each measurement, MAE was the arithmetic mean of the absolute differences between predicted and reference values across evaluable records. Dice approaches one as overlap increases, whereas localisation and measurement errors approach zero as agreement increases. Error summaries for lengths, angles and coverage were expressed in their respective units rather than pooled across incompatible quantities. Overall region and landmark summaries were calculated by first averaging applicable structures within each record and then averaging across records. Comparisons included KAP-DDH versus each expert and Expert 1 versus Expert 2. Supplementary agreement analyses used signed bias (KAP-DDH minus Expert 1), Bland–Altman limits of agreement (mean signed bias ± 1.96 sample standard deviations of the paired differences) [59], Lin's concordance correlation [60] and the two-way absolute-agreement single-measure intraclass correlation, ICC(A,1) [61].

The limits of agreement describe the spread of paired differences, rather than the confidence interval for mean bias. Coverage differences are expressed in percentage points. Independent-expert paired differences are presented in Supplementary Figs. 1, 3 and 4, and measurement

agreement with Expert 1 in Supplementary Figs. 2 and 5. Supplementary Tables 11 and 12 report both Expert 2-referenced comparisons and their saved paired estimates.

Planning-to-postoperative agreement was assessed for cup diameter, cup-centre-to-IT distance, cup-centre-to-KL distance, inclination and radiographic coverage. The reference for diameter was the clinically recorded implanted size; the other endpoints used postoperative measurements from each expert separately. The two annotation references assess sensitivity to reference choice on the same records. We also assessed cup-size agreement within ±2 mm and ±4 mm, planned inclination of 30–50°, coverage ≥80% and simultaneous inclination and coverage criterion fulfilment. All three planners were evaluated on the same 691 internal and 422 external records for every planning endpoint and reference, including coverage.

Criterion rates describe the planned reconstruction; a rate of 100% means that every evaluated plan met the stated threshold.

Knowledge-base comparison configuration

The paired knowledge-base comparison evaluated two target-construction configurations on the same 691 internal and 422 external records. Knowledge-base-enabled planning uses historical retrieval to define patient-specific intervals and preferred values. The no-KB configuration disables retrieval and substitutes Crowe-specific defaults while retaining perception, candidate-generation procedures, geometric checks and ranking; changes in targets may alter candidate pools and fallback status. Both configurations used the same calibration and postoperative references, and the query record's postoperative information was used only for evaluation. Positive effects were defined as no-KB absolute error minus KB absolute error. Patient-cluster intervals and patient-block tests followed the statistical procedures described below; Benjamini–Hochberg adjustment covered nine unique continuous contrasts within each cohort, with implanted diameter counted once and four image-derived endpoints evaluated against both experts. Figure 8, Supplementary Fig. 7 and Supplementary Table 14 report the paired results; Supplementary Tables 9 and 10 define the configuration. The default targets define a within-pipeline comparator and do not represent a separate clinical planner.

Statistical analyses

Baseline continuous variables were summarised as medians and interquartile ranges, and categorical variables as counts and percentages, without baseline hypothesis testing. Anatomical-perception results are reported as record-weighted means, sample standard deviations calculated with n − 1 and pointwise 95% patient-cluster percentile bootstrap confidence intervals. Independent-observer comparisons used Expert 2 as the common reference: the paired difference was KAP-DDH–Expert 2 agreement minus Expert 1–Expert 2 agreement. Positive Dice differences favoured KAP-DDH, whereas negative localisation- or measurement-error differences favoured KAP-DDH. Benjamini–Hochberg adjustment[62] was applied separately to six phase-by-domain families, each pooling both cohorts: preoperative regions, landmarks and measurements comprised 20, 18 and 24 contrasts, and the corresponding postoperative families comprised 22, 28 and 36 contrasts.

All inferential analyses used the verified record-to-patient crosswalk. Point estimates weighted eligible records equally; 10,000 patient-cluster bootstrap samples retained all eligible records and all paired values belonging to each sampled patient.[63] Confidence limits were the 2.5th and 97.5th percentiles, calculated using linear quantiles. This procedure preserved dependence between records from the same patient without giving each patient equal weight in the estimand. The planning effect was defined as surgeon absolute error minus KAP-DDH absolute error against the same reference; positive values favoured KAP-DDH. A common set of sampled patients was used across planning endpoints, planners and expert references within each cohort. Separate streams were used for each anatomical phase/domain and cohort. Random streams were generated from NumPy SeedSequence with seed 20261004 and its default PCG64 generator; patient identifiers were sorted before cluster construction. Separate bootstrap and sign-flip streams were used for anatomical analyses. The analysis implementation records the complete stream allocation.

Exploratory two-sided tests used 20,000 patient-block sign flips, with one random sign assigned to all paired differences from the same patient within an analysis.[64] The test statistic was the absolute record-weighted mean. Monte Carlo P values were calculated as (b + 1)/20,001, where b is the number of randomised statistics at least as large as the observed statistic, with an absolute

numerical tie tolerance of 10^{-12}. The minimum attainable P value was therefore 1/20,001.[65] For surgeon comparisons, Benjamini–Hochberg q values covered 36 continuous contrasts: two cohorts, two surgeons and nine endpoint-reference combinations, comprising one implanted-diameter reference and two expert references for each remaining endpoint. Confidence intervals were pointwise and were not multiplicity-adjusted. Sign-flip tests assume joint sign symmetry under the null within independent patient blocks; planning methods were not randomly assigned.

Twenty paired criterion-rate comparisons report KAP-DDH minus surgeon criterion-fulfilment proportions in percentage points, retaining denominators and discordant record pairs. Empirical-bootstrap intervals were omitted for rates of exactly 0% or 100% and for paired rate differences when either marginal rate was at a boundary; descriptive counts, rates and differences were retained. The omission reflects the boundary behaviour of the empirical bootstrap. The 20 comparisons comprise five binary criteria, two cohorts and two surgeon comparators; they were descriptive, with no hypothesis tests or multiplicity-adjusted q values, and were not included in the continuous-testing family. All calculations used the full stored numerical precision. Values were rounded only for display, and threshold classification used exact inclusive comparisons without rounding or numerical tolerance. No observed outlier was removed from the primary analysis. Crowe-stratified analyses were descriptive, with individual observations and medians, without density curves or box estimates, retained for groups containing fewer than ten patients. All recorded Crowe types and subgroup denominators were retained. The criterion-rate and subgroup results are presented in Fig. 6e, Fig. 7, Supplementary Fig. 6 and Supplementary Table 13.

Software and reproducibility

The research implementation separates perception (nnUNetv2), representation/retrieval (PyTorch[66] and PyTorch Geometric[67]) and planning (NumPy[68] and OpenCV). The public software release covers the preoperative planning module.

The public release includes a fully synthetic example that runs in fixed-constraint mode and demonstrates planning-module execution. Release validation comprised synthetic-demo execution,

Python compilation and SHA256 integrity checks; installation, input/output, integration and validation instructions accompany the code.

Numerical analyses cover anatomical perception, planning, threshold, subgroup and ablation summaries, as well as the Fig. 5 target summaries. Statistical reanalysis used Python 3.12.14, NumPy 2.3.5 and openpyxl 3.1.5; these versions refer to statistical reanalysis, not model training or planning inference.

Supplementary Video 1 demonstrates automated planning and interactive refinement within the research workflow. Only automatic Top-1 outputs were evaluated.

Data availability

Access to source radiographs and study-specific historical resources is governed by participant-privacy and institutional data-governance requirements. Requests for controlled access to de-identified materials should be addressed to Yong Nie (nieyong1983@wchscu.cn) and are subject to institutional and ethics approval and a data-use agreement.

Code availability

The source code and graphical user interface for the KAP-DDH preoperative planning module are publicly available at https://github.com/muhuohuo1128/KAP-DDH-Planning. Version v1.0.0 is permanently archived in Zenodo at https://doi.org/10.5281/zenodo.23163959 under the GNU General Public License v3.0.[69] The release includes the preoperative planning engine, graphical user interface, configuration, study-specific constraint integration interface, fully synthetic demonstration data, and reproducibility and validation utilities. Study-specific resources used in the full research workflow are managed separately in accordance with institutional data-governance requirements. Relevant supporting materials can be made available to editors and reviewers for manuscript evaluation where appropriate.

Acknowledgements

This work was supported by the Advanced Materials—National Science and Technology Major Project (No. 2025ZD0619901) and the 1.3.5 project for disciplines of excellence, West China Hospital, Sichuan University (26HXJS004). The funders had no role in the study design; data collection, analysis, or interpretation; writing of the report; or the decision to submit the manuscript for publication. The authors had final responsibility for the decision to submit.

Author contributions

S.C. contributed to conceptualisation, methodology, data curation, formal analysis, visualisation, and writing of the original draft. J.W. contributed to methodology, data curation, validation, expert annotation, and review and editing of the manuscript. B.W. contributed to methodology, software implementation, formal analysis, and review and editing. Y.Z. contributed to conceptualisation, clinical planning validation as Surgeon 1, supervision, and review and editing. G.C. contributed to clinical investigation, external validation, clinical planning validation as Surgeon 2, data curation, and review and editing. T.D., L.Z., Q.Z., J.S., and B.H. contributed to data acquisition or curation, interpretation of study findings, and review and editing of the manuscript. K.L. contributed to conceptualisation, methodology, computational supervision, and review and editing. Z.Z. contributed to conceptualisation, methodology, clinical interpretation, supervision, and review and editing of the manuscript. Y.N. contributed to conceptualisation, expert annotation, supervision, project administration, funding acquisition, and review and editing. S.C. and J.W. directly accessed and verified the underlying data. All authors reviewed and approved the final manuscript and accept responsibility for the work.

Competing interests

The authors declare no competing interests.

References

1. O'Connor, K. P., Marshall, B. J., Davison, J., Clohisy, J. C. & Willey, M. C. Prevalence of radiographic hip dysplasia in the general adult population: a systematic review. *Iowa Orthop. J.* **44,** 145–149 (2024).

2. Mei, X. Y., Gong, Y. J., Safir, O., Gross, A. & Kuzyk, P. Long-term outcomes of total hip arthroplasty in patients younger than 55 years: a systematic review of the contemporary literature. *Can. J. Surg.* **62,** 249–258 (2019).

3. Salman, L. A. et al. The outcomes of total hip arthroplasty in developmental dysplasia of hip versus osteoarthritis: a systematic review and meta-analysis. *Eur. J. Orthop. Surg. Traumatol.* **34,** 1–8 (2024).

4. Dantas, P., Clode Araujo, T., Medeiros Grenho, A., Gonçalves, S. R. & Tsiridis, E. THA for DDH: replacement principles and techniques - acetabular side. *EFORT Open Rev.* **11,** 696–709 (2026).

5. Yang, Y. et al. Morphological analysis of true acetabulum in hip dysplasia (Crowe classes I–IV) via 3-D implantation simulation. *J. Bone Joint Surg. Am.* **99,** e92 (2017).

6. Yang, Y. et al. Three-dimensional morphological analysis of true acetabulum in Crowe type IV hip dysplasia via standard-sized cup-simulated implantation. *Quant. Imaging Med. Surg.* **12,** 2904–2916 (2022).

7. Hartofilakidis, G., Stamos, K., Karachalios, T., Ioannidis, T. T. & Zacharakis, N. Congenital hip disease in adults. Classification of acetabular deficiencies and operative treatment with acetabuloplasty combined with total hip arthroplasty. *J. Bone Joint Surg. Am.* **78**, 683–692 (1996).

8. Pagnano, M. W., Hanssen, A. D., Lewallen, D. G. & Shaughnessy, W. J. The effect of superior placement of the acetabular component on the rate of loosening after total hip arthroplasty. Long-term results in patients who have Crowe type-II congenital dysplasia of the hip. *J. Bone Joint Surg. Am.* **78**, 1004–1014 (1996).

9. Hartofilakidis, G., Yiannakopoulos, C. K. & Babis, G. C. The morphologic variations of low and high hip dislocation. *Clin. Orthop. Relat. Res.* **466**, 820–824 (2008).

10. Colombi, A., Schena, D. & Castelli, C. C. Total hip arthroplasty planning. *EFORT Open Rev.* **4,** 626–632 (2019).

11. Lourens, E., Kurmis, A. P. & Lim, W. Y. Clinical impact of pelvic malrotation on radiograph-based preoperative planning for total hip arthroplasty: a proof-of-concept and prudent prediction of acceptable rotation. *Indian J. Orthop.* **56,** 1053–1060 (2022).

12. Zhao, X. et al. The utility of digital templating in total hip arthroplasty with Crowe type II and III dysplastic hips. *Int. Orthop.* **35**, 631–638 (2011).

13. Crowe, J. F., Mani, V. J. & Ranawat, C. S. Total hip replacement in congenital dislocation and dysplasia of the hip. *J. Bone Joint Surg. Am.* **61,** 15–23 (1979).

14. Nie, Y. et al. Radiographic underestimation of in vivo cup coverage provided by total hip arthroplasty for dysplasia. *Orthopedics* **41,** e46–e51 (2018).

15. Yon, C.-J., Lee, K.-J., Choi, B.-C., Suh, H.-S. & Min, B.-W. The validation of two-dimensional and three-dimensional radiographic measurements of host bone coverage in total hip arthroplasty for hip dysplasia: a comparison with intra-operative measurements. *J. Clin. Med.* **12,** 6227 (2023).

16. Ueno, T. et al. Three-dimensional host bone coverage required in total hip arthroplasty for developmental dysplasia of the hip and its relationship with 2-dimensional coverage. *J. Arthroplasty* **34,** 93–101 (2019).

17. Stotter, C. et al. Deep learning for fully automated radiographic measurements of the pelvis and hip. *Diagnostics* **13**, 497 (2023).

18. Kim, M., Oh, I.-S. & Yoon, S.-J. Deep learning and computer vision techniques for automated total hip arthroplasty planning on 2-D radiographs. *IEEE Access* **10**, 94145–94157 (2022).

19. Taghavi, S. P. et al. Three-dimensional and artificial intelligence-assisted preoperative templating in total hip arthroplasty: a systematic review and meta-analysis. *Arthroplasty* **8,** 56 (2026).

20. Xie, H. et al. Application and evaluation of artificial intelligence 3D preoperative planning software in developmental dysplasia of the hip. *J. Orthop. Surg. Res.* **19,** 176 (2024).

21. Huo, J. et al. Value of 3D preoperative planning for primary total hip arthroplasty based on artificial intelligence technology. *J. Orthop. Surg. Res.* **16**, 156 (2021).

22. Mulford, K. L. et al. A deep learning tool for automated landmark annotation on hip and pelvis radiographs. *J. Arthroplasty* **38,** 2024–2031.e1 (2023).

23. Rouzrokh, P. et al. A deep learning tool for automated radiographic measurement of acetabular component inclination and version after total hip arthroplasty. *J. Arthroplasty* **36,** 2510–2517.e6 (2021).

24. Archer, H. et al. Artificial intelligence-generated hip radiological measurements are fast and adequate for reliable assessment of hip dysplasia: an external validation study. *Bone Jt Open* **3**, 877–884 (2022).

25. Rouzrokh, P. et al. THA-Net: a deep learning solution for next-generation templating and patient-specific surgical execution. *J. Arthroplasty* **39**, 727–733.e4 (2024).

26. Yao, J. et al. Development and Application of Automated Preoperative Planning System for Developmental Dysplasia of the Hip via Keypoint Selective Integration. J. Imaging Inform. Med. https://doi.org/10.1007/s10278-026-02149-x (2026).

27. Momin, S. et al. Knowledge-based radiation treatment planning: a data-driven method survey. *J. Appl. Clin. Med. Phys.* **22,** 16–44 (2021).

28. Huang, C. et al. Learning image representations for content-based image retrieval of radiotherapy treatment plans. *Phys. Med. Biol.* **68,** 095025 (2023).

29. Schmidt, R., Montani, S., Bellazzi, R., Portinale, L. & Gierl, L. Cased-based reasoning for medical knowledge-based systems. *Int. J. Med. Inform.* **64,** 355–367 (2001).

30. Maier-Hein, L. et al. Metrics reloaded: recommendations for image analysis validation. *Nat. Methods* **21**, 195–212 (2024).

31. Reinke, A. et al. Understanding metric-related pitfalls in image analysis validation. *Nat. Methods* **21**, 182–194 (2024).

32. Rouzrokh, P. et al. Applying deep learning to establish a total hip arthroplasty radiography registry: a stepwise approach. *J. Bone Joint Surg. Am.* **104**, 1649–1658 (2022).

33. Nepple, J. J. et al. Three patterns of acetabular deficiency are common in young adult patients with acetabular dysplasia. *Clin. Orthop. Relat. Res.* **475**, 1037–1044 (2017).

34. Zeng, Y. et al. Three-dimensional computerized preoperative planning of total hip arthroplasty with high-riding dislocation developmental dysplasia of the hip. *Orthop. Surg.* **6**, 95–102 (2014).

35. Ding, X. et al. Value of preoperative three-dimensional planning software (AI-HIP) in primary total hip arthroplasty: a retrospective study. *J. Int. Med. Res.* **49**, 03000605211058874 (2021).

36. Stirling, P. et al. Does cup position at the high hip center or anatomic hip center in THA for developmental dysplasia of the hip result in better Harris hip scores and revision incidence? A systematic review. *Clin. Orthop. Relat. Res.* **479,** 1119–1130 (2021).

37. Kannan, A., Madurawe, C., Pierrepont, J. & McMahon, S. Does total hip arthroplasty with high hip centre in dysplasia compromise acetabular bone stock? *Hip Int.* **33,** 518–524 (2023).

38. Nawabi, D. H., Meftah, M., Nam, D., Ranawat, A. S. & Ranawat, C. S. Durable fixation achieved with medialized, high hip center cementless THAs for Crowe II and III dysplasia. *Clin. Orthop. Relat. Res.* **472**, 630–636 (2014).

39. Feng, T. et al. An algorithm for personalized optimal acetabular cup implant positioning based on bony coverage rate and impingement-free range of motion. *Med. Biol. Eng. Comput.* **64,** 739–753 (2026).

40. Dennis, D. A. et al. Does individualization of cup position affect prosthetic or bone impingement following total hip arthroplasty? *J. Arthroplasty* **38,** S257–S264 (2023).

41. LaValva, S. M. et al. A multicentre prospective assessment of the utility of robotic assisted total hip arthroplasty with virtual range of motion on intraoperative implant positioning. *Hip Int.* **34,** 717–723 (2024).

42. Tachibana, T., Fujii, M., Kitamura, K., Nakamura, T. & Nakashima, Y. Does acetabular coverage vary between the supine and standing positions in patients with hip dysplasia? *Clin. Orthop. Relat. Res.* **477**, 2455–2466 (2019).

43. Vorimore, C. et al. Change in cup orientation from supine to standing posture: a prospective cohort study of 419 total hip arthroplasties. *Acta Orthop.* **95**, 425–432 (2024).

44. Esmaeili, S. et al. Total hip arthroplasty in patients who have Crowe type IV developmental dysplasia of the hip: a systematic review. *J. Arthroplasty* **39,** 2645–2660.e19 (2024).

45. Tateda, K., Nagoya, S., Suzuki, D., Kosukegawa, I. & Yamashita, T. Acetabular morphology in patients with developmental dysplasia of the hip with high dislocation. *Hip Pelvis* **33**, 25–32 (2021).

46. The, B., Diercks, R. L., Stewart, R. E., van Ooijen, P. M. A. & van Horn, J. R. Digital correction of magnification in pelvic x rays for preoperative planning of hip joint replacements: theoretical development and clinical results of a new protocol. *Med. Phys.* **32**, 2580–2589 (2005).

47. Eggli, S., Pisan, M. & Müller, M. E. The value of preoperative planning for total hip arthroplasty. *J. Bone Joint Surg. Br.* **80**, 382–390 (1998).

48. Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S. & Scheinost, D. Data leakage inflates prediction performance in connectome-based machine learning models. *Nat. Commun.* **15**, 1829 (2024).

49. Gaube, S. et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *npj Digit. Med.* **4**, 31 (2021).

50. Tschandl, P. et al. Human–computer collaboration for skin cancer recognition. *Nat. Med.* **26**, 1229–1234 (2020).

51. Vasey, B. et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat. Med.* **28**, 924–933 (2022).

52. Lekadir, K. et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ* **388,** e081554 (2025).

53. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18,** 203–211 (2021).

54. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer, 2015).

55. Vaswani, A. et al. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 5998–6008 (2017).

56. Brody, S., Alon, U. & Yahav, E. How attentive are graph attention networks? In International Conference on Learning Representations https://openreview.net/forum?id=F72ximsx7C1 (2022).

57. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. *Proc. Mach. Learn. Res.* **119**, 1597–1607 (2020).

58. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *International Conference on Learning Representations* https://arxiv.org/abs/1412.6980 (2015).

59. Bland, J. M. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327,** 307–310 (1986).

60. Lin, L. I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45,** 255–268 (1989).

61. McGraw, K. O. & Wong, S. P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1,** 30–46 (1996).

62. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

63. Field, C. A. & Welsh, A. H. Bootstrapping clustered data. *J. R. Stat. Soc. B* **69**, 369–390 (2007).

64. Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E. & Smith, S. M. Multi-level block permutation. *Neuroimage* **123**, 253–268 (2015).

65. Phipson, B. & Smyth, G. K. Permutation P-values should never be zero: calculating exact P-values when permutations are randomly drawn. *Stat. Appl. Genet. Mol. Biol.* **9**, Article 39 (2010).

66. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8026–8037 (2019).

67. Fey, M. & Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds* https://arxiv.org/abs/1903.02428 (2019).

68. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

69. Chen, S. KAP-DDH Planning v1.0.0. Zenodo https://doi.org/10.5281/zenodo.23163959 (2026).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf
- SupplementaryVideo1.mp4