SRIWIJAYA JOURNAL OF SURGERY
Vol. 10 No. 1 (2027) · pp. 24-44 · sriwijayasurgery.com
e-ISSN 2722-3558

ORIGINAL ARTICLE

Age at Surgery and Cosmetic Outcomes After Hypospadias Repair in Palembang, Indonesia: An Ambispective Cohort Study Using HOPE Scores

Yoga Trisendi[1*], Iqmal Perlianta[2], Debby Handayati Harahap[3]

[1]General Surgery Study Program, Department of Surgery, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia
[2]Department of Surgery, Dr. Mohammad Hoesin Central General Hospital, Palembang, Indonesia
[3]Department of Pharmacology, Faculty of Medicine, Universitas Sriwijaya, Palembang, Indonesia

ARTICLE INFO

Article history:
Received: October 30, 2026
Revised: November 5, 2026
Accepted: November 30, 2026
Available online: May 5, 2027
Published: June 1, 2027

Keywords:
Age at surgery; cosmetic outcome; HOPE score; hypospadias reconstruction; pediatric urology

Corresponding author:
Yoga Trisendi
Email:
trisendiyoga2013@gmail.com

DOI:
https://doi.org/10.37275/sjs.v10i1.183

Access licence:
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0).

Copyright:
© 2027 The Authors. Authors retain copyright and grant the journal the right of first publication. Published by HM Publisher.

ABSTRACT

Background: Delayed hypospadias repair remains common in tertiary referral settings. The Hypospadias Objective Penile Evaluation (HOPE) score standardizes postoperative cosmetic assessment.

Objective: To examine the association between age at surgery and postoperative HOPE scores at Dr. Mohammad Hoesin Central General Hospital, Palembang, Indonesia.

Methods: This ambispective cohort included 53 patients who underwent hypospadias reconstruction between July 2025 and July 2026. Two examiners assessed postoperative HOPE scores, categorized as optimal (≥57) or suboptimal (<57). Analyses used Mann–Whitney U, Kruskal–Wallis, Fisher exact, and receiver operating characteristic (ROC) tests.

Results: Mean age at surgery was 93.28 ± 53.09 months; 84.9% underwent surgery after 18 months, and 60.4% had proximal hypospadias. Mean HOPE score was 47.79 ± 6.33; 92.5% scored <57. Scores were higher after surgery at ≤18 months than at >18 months (54.38 ± 3.16 vs. 46.62 ± 6.04; $p < 0.001$), with suboptimal outcomes in 62.5% and 97.8%, respectively ($p = 0.009$; RR = 1.564; OR = 26.400). Scores did not differ significantly by subtype ($p = 0.315$). ROC analysis yielded AUC = 0.865 (95% CI: 0.681–1.000; $p = 0.016$) and an exploratory 13.5-month cutoff.

Conclusion: Earlier surgery was associated with higher postoperative cosmetic scores in this cohort. The exploratory age cutoff requires external validation; residual confounding and follow-up timing warrant consideration.

1. Introduction

Hypospadias is one of the most prevalent congenital malformations of the male external genitalia, characterized anatomically by an arrest in the ventral fusion of the urethral folds between the eighth and sixteenth weeks of embryogenesis.[1,2] This developmental disruption produces an ectopic urethral meatus positioned anywhere along the ventral aspect of the glans, penile shaft, penoscrotal junction, scrotum, or perineum, frequently accompanied by ventral penile curvature (chordee) and a dorsal hooded foreskin with ventral preputial deficiency.[1,3] Population-based birth defect registries across North America, Europe, and Asia estimate an international prevalence ranging from 11.3 to 23.8 cases per 10,000 live births, or approximately 1 in 200 to 1 in 250 male newborns, with contemporary active surveillance confirming sustained regional variation in severe phenotypes.[2,3] Etiologically, hypospadias arises from a multifactorial interplay of polygenic susceptibility, androgen receptor or 5α-reductase signaling perturbations, placental insufficiency associated with low birth weight, and maternal exposure to environmental endocrine-disrupting chemicals during the critical window of genital tubercle masculinization.[4–6] Beyond functional voiding disturbances such as urinary stream deflection or spraying, uncorrected or suboptimally reconstructed hypospadias imposes substantial long-term psychosexual morbidity, body-image dissatisfaction, and interpersonal anxiety as affected boys transition through adolescence and adulthood.[7–9]

The primary objectives of modern surgical reconstruction are threefold: to establish a straight penis through adequate orthoplasty, to construct a caliber-appropriate neourethra terminating in a slit-like meatus at the apex of the glans, and to achieve a symmetrical, aesthetically normal circumcised or foreskin-reconstructed penile appearance with minimal postoperative morbidity.[7,9] International clinical practice guidelines issued by the European Association of Urology (EAU) and the American Academy of Pediatrics (AAP) recommend performing primary hypospadias repair between 6 and 18 months of age.[7,8] This early infantile surgical window is grounded in both developmental psychology and microvascular tissue biology. Psychologically, completing genital reconstruction prior to 18 months precedes the consolidation of genital awareness and castration anxiety, thereby mitigating perioperative emotional trauma and behavioral maladjustment.[7,8] Biologically, infant penile tissues within the first 18 months of life exhibit favorable extracellular matrix compliance, a higher ratio of elastic type III to rigid type I collagen within the dartos fascia and corpus spongiosum, robust microvascular collateralization, and minimal nocturnal erectile tumescence tension along the neourethral suture line.[8,9] Conversely, delayed surgical presentation beyond 18 to 24 months has been associated in several clinical cohorts with progressive dartos fibrosis, reduced parameatal tissue elasticity, heightened mechanical stress from stronger erections, denser cutaneous bacterial colonization, and increased rates of urethrocutaneous fistula, glans dehiscence, or meatal stenosis.[8,9]

Historically, the evaluation of cosmetic outcomes following hypospadias surgery relied heavily on unstandardized subjective impressions by the operating surgeon, introducing substantial observer bias and precluding meaningful cross-institutional comparisons.[10] To overcome this methodological limitation, the Dutch Hypospadias Study Group developed and

validated the Hypospadias Objective Penile Evaluation (HOPE) score in 2013 as a rigorous, standardized, six-component psychometric instrument specifically designed to quantify postoperative penile appearance.[10] The HOPE scoring system evaluates six surgically modifiable anatomical domains against standardized reference photographs: position of the meatus, shape of the meatus, shape of the glans, shape of the penile skin, penile axis (torsion), and penile curvature during erection, generating a composite score ranging from 6 (most severe abnormality) to 60 (optimal normal appearance).[10,11] In the foundational validation study by van der Toorn et al., the HOPE instrument demonstrated high intra-observer reliability (0.817), inter-observer reliability (0.790), nonparametric internal validity (0.849), and parametric internal validity (0.842).[10] Subsequent prospective clinical investigations, including recent work by Neheman et al. and multicenter external validation by Kim et al., have established a composite HOPE threshold of ≥57 points to define an optimal cosmetic outcome and <57 points to indicate a suboptimal appearance potentially warranting further clinical scrutiny.[11,12] Furthermore, contemporary clinical trials and prospective registry protocols increasingly adopt the HOPE score alongside preoperative anatomical indices such as the Glans-Urethral Meatus-Shaft (GMS) score to standardize outcome reporting across diverse surgical techniques.[13,14]

Despite the biological plausibility linking early surgical timing to superior wound healing—supported by recent molecular evidence demonstrating that parameatal urethral plate expression of epidermal growth factor (EGF) and transforming growth factor-beta (TGF-β) as well as microvascular organization critically govern urethroplasty remodeling[15,16]—the empirical relationship between age at surgery and objective HOPE cosmetic scores remains incompletely characterized in low- and middle-income tertiary referral settings. In Indonesia, tertiary referral hospitals frequently encounter a distinct clinical epidemiology marked by delayed patient presentation during late childhood or adolescence and a high proportion of severe proximal hypospadias cases referred from peripheral districts.[17] While international cohorts dominated by distal subcoronal defects operated on in infancy report optimal HOPE scores (≥57) in up to 81% of patients,[11] real-world data quantifying how delayed surgical age and proximal anatomical predominance interact to shape objective HOPE outcomes in Indonesian tertiary centers are scarce, and no prior study has evaluated this relationship in South Sumatra. Therefore, this study aimed to analyze the association between age at surgery and three-month postoperative cosmetic outcomes assessed using the HOPE score, to evaluate HOPE score distributions across anatomical hypospadias subtypes, and to explore the discriminative performance and empirical age cutoff of surgical timing via receiver operating characteristic (ROC) curve analysis at Dr. Mohammad Hoesin Central General Hospital, Palembang.

From a health-systems and surgical governance perspective, investigating the relationship between surgical timing and objective cosmetic outcomes in South Sumatra addresses a critical gap in global pediatric surgery literature. Most validation cohorts and prognostic models for the HOPE score have originated from high-income healthcare systems characterized by universal neonatal screening, rapid primary-to-tertiary pediatric urology referral pathways, and early infantile operative correction predominantly for glanular or subcoronal defects.[10–13] In contrast, archipelagic and resource-variable settings across Southeast Asia face multifaceted structural barriers—including geographic distance from subspecialty surgical centers, delayed recognition of ventral chordee or ectopic meatal position by primary birth attendants, socioeconomic constraints, parental apprehension regarding infant anesthesia, and staged referral administration—that systematically shift primary hypospadias repair into late childhood and adolescence.[4,7,17] When surgery is deferred to a mean age of nearly eight years and more than three-fifths of presenting patients exhibit complex proximal defects, the reconstructive surgeon must contend simultaneously with altered tissue biology and heightened anatomical severity.[8,17] By pairing dual-examiner HOPE scoring with nonparametric distributional testing, exact small-sample contingency inference, Firth-penalized likelihood estimation, and full-spectrum ROC coordinate profiling, the present study establishes an empirical benchmark for surgical timing and cosmetic outcome evaluation in Indonesian tertiary reconstructive practice.

2. Methods

Study Design, Setting, and Reporting Standards

This analytical observational investigation employed an ambispective cohort design conducted at the Division of Plastic, Reconstructive, and Aesthetic Surgery and the Pediatric Urology Service within the Department of Surgery, Dr. Mohammad Hoesin Central General Hospital, Palembang, South Sumatra, Indonesia. As the principal tertiary referral and academic teaching hospital affiliated with the Faculty of Medicine, Universitas Sriwijaya, the institution receives complex congenital urogenital referrals from across South Sumatra and neighboring provinces. The ambispective framework integrated retrospective abstraction of baseline demographic, preoperative anatomical, and operative timing records for patients who underwent definitive hypospadias reconstruction between July 2025 and July 2026 with prospective, standardized clinical and photographic cosmetic evaluation conducted between September 2026 and October 2026 at a minimum follow-up interval of three months postoperatively. The study design, execution, and statistical reporting adhered strictly to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cohort studies.

To ensure methodological rigor and reproducibility in accordance with STROBE recommendations for ambispective cohort investigations, a standardized clinical data abstraction and photographic evaluation protocol was established prior to statistical analysis. Historical operative logs, pre-anesthetic physical examination records, and intraoperative anatomical diagrams from July 2025 to July 2026 were cross-checked against electronic hospital information system entries to verify exact dates of birth, dates of definitive reconstruction, native meatal positions after chordee release (orthoplasty), and absence of exclusion criteria. Subsequently, during the prospective evaluation window from August 2026 to September 2026, all eligible patients attending the outpatient reconstructive surgery clinic at a minimum postoperative interval of three months underwent uniform physical inspection and standardized three-view digital photography.

Study Population, Eligibility Criteria, and Sampling Strategy

The target population comprised all male patients diagnosed with congenital hypospadias who underwent surgical reconstruction at Dr. Mohammad Hoesin Central General Hospital during the twelve-month operative window from July 2025 to July 2026. Because the institutional caseload of completed reconstructions meeting the minimum three-month healing interval within the study period represented a finite,

well-defined clinical population, total consecutive sampling was implemented to enroll all eligible patients. Inclusion criteria required patients to have a confirmed diagnosis of distal, midshaft (medial), or proximal hypospadias, to have completed surgical reconstruction at least three months prior to the cosmetic evaluation without active urethrocutaneous fistula, and to have written informed consent provided by parents or legal guardians (with age-appropriate assent for older children and adolescents) for clinical examination and standardized medical photography. Exclusion criteria encompassed patients with concomitant disorders of sex development or severe non-hypospadias penile malformations that would confound anatomical scoring, patients experiencing severe acute postoperative complications (such as complete neourethral breakdown or active surgical site infection) precluding objective cosmetic grading, and patients with neurodevelopmental or behavioral conditions preventing standardized photographic documentation. Drop-out criteria applied to patients who failed to attend the scheduled three-month outpatient follow-up visit or whose photographic documentation was incomplete. Across the study period, 53 consecutive patients satisfied all eligibility criteria with zero exclusions or missing data (100.0% complete case retention).

The requirement that enrolled patients be at least three months post-reconstruction and free of active urethrocutaneous fistula was instituted for two specific methodological reasons. First, early postoperative complications such as acute wound dehiscence, active infection, or urine leakage through a fistula tract distort local cutaneous geometry and induce secondary inflammatory edema, which can confound objective grading of the six standardized cosmetic domains of the HOPE instrument.[10,11] Second, by restricting the cohort to structurally intact three-month repairs across all anatomical subtypes, the study isolated the intrinsic quality of anatomical and aesthetic restoration—meatal position, meatal slit configuration, glans symmetry, penile shaft skin uniformity, rotational axis, and longitudinal straightness—as a function of chronological age at surgery and baseline meatal location.

Variables, Operational Definitions, and Standardized HOPE Assessment

The primary independent variable was age at surgery, defined operationally as the patient's chronological age in completed months calculated from the verified date of birth to the date of definitive hypospadias reconstruction as recorded in the institutional surgical registry. Age at surgery was analyzed first as a continuous variable (months), second as a prespecified binary variable anchored to the international EAU/AAP guideline upper limit (≤18 months [6.00–18.00 months] vs. >18 months), and third as an exploratory binary variable defined by the optimal empirical cutoff derived from ROC curve analysis (≤13.5 months vs. >13.5 months). Preoperative anatomical hypospadias severity was recorded as a key clinical covariate and categorized according to the meatal location after chordee release into proximal (penoscrotal, scrotal, perineal, or proximal penile shaft), midshaft/medial (mid-penile shaft), and distal (glanular, coronal, subcoronal, or distal penile shaft) subtypes.[17,18]

The primary dependent variable was the postoperative cosmetic outcome at three months after reconstruction, quantified using the validated six-component Hypospadias Objective Penile Evaluation (HOPE) scoring system.[10] Standardized clinical photographs were acquired in three orthogonal projections (anterior, lateral, and ventral views) using a calibrated digital camera under uniform clinical lighting, supplemented by direct physical inspection of the glans, meatus, penile skin, shaft axis, and curvature. Each of the six HOPE domains was scored against the official Dutch Hypospadias Study Group reference atlas: first, position of the meatus (Position 1 [normal tip of glans] = 10 points, Position 2 = 8 points, Position 3 = 5 points, Position 4 = 3 points, Position 5 = 1 point); second, shape of the meatus (normal vertical slit = 10 points, slightly abnormal = 7 points, moderately abnormal = 4 points, severely abnormal = 1 point); third, shape of the glans (normal symmetrical acorn configuration = 10 points, slightly abnormal = 7 points, moderately abnormal = 4 points, severely abnormal = 1 point); fourth, penile skin shape (normal uniform coverage = 10 points, slightly abnormal = 7 points, moderately abnormal = 4 points, severely abnormal = 1 point); fifth, torsion of the penis (0–30° = 10 points, 30–50° = 7 points, 50–70° = 3 points, >70° = 1 point); and sixth, curvature during erection or simulated glans traction (0–30° = 10 points, 30–50° = 7 points, 50–70° = 3 points, >70° = 1 point).[10] The composite HOPE score was calculated as the arithmetic sum of all six domain scores (possible range: 6 to 60 points). In accordance with validated prognostic benchmarks established by Neheman et al.,[11] composite HOPE scores were dichotomized into optimal cosmetic outcome (HOPE ≥ 57 points) and suboptimal cosmetic outcome (HOPE < 57 points). To assess inter-rater reliability, two trained senior surgical residents independently scored all 53 standardized photographic sets after joint calibration with the HOPE reference atlas.

During photographic acquisition, patients were positioned supine on the examination table with the lower extremities abducted to expose the external genitalia and perineum clearly. Anterior, lateral, and ventral photographs were captured at a standardized focal distance under uniform diffuse clinical illumination without flash glare, utilizing a millimeter calibration ruler and goniometric reference alignment where indicated to assess residual penile torsion or ventral curvature during spontaneous tumescence or standardized gentle glans traction. Prior to scoring the study cohort, both evaluators completed a structured calibration session using the original visual reference sheets published by the Dutch Hypospadias Study Group,[10] reviewing the exact point weights assigned to each morphological grade (10, 8, 5, 3, or 1 point for meatal position; 10, 7, 4, or 1 point for meatal shape, glans shape, and penile skin shape; and 10, 7, 3, or 1 point for penile torsion and curvature).

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY, USA). Inter-rater agreement between Examiner 1 and Examiner 2 across the 53 patients was evaluated using Cronbach's alpha (α), inter-item Pearson correlation (r), and the two-way random-effects, absolute-agreement Intraclass Correlation Coefficient (Type A ICC for single and average measures) with 95% confidence intervals (CIs) and F-tests. Continuous variables (age at surgery and composite HOPE score) were summarized using both parametric metrics (mean ± standard deviation [SD], standard error [SE], 95% CI for the mean, and 5% trimmed mean) and nonparametric metrics (median, interquartile range [IQR], minimum, maximum, range, skewness, and kurtosis). Categorical variables were reported as exact counts (n) and percentages (%) summing to 100.0%.

Distributional normality of continuous HOPE scores within age and anatomical subgroups was formally evaluated using the Kolmogorov–Smirnov test with Lilliefors significance correction and the Shapiro–Wilk test. Because HOPE scores violated normality in the >18 months group (Kolmogorov–Smirnov $p < 0.001$; Shapiro–Wilk $p = 0.001$) and the proximal hypospadias

group (Kolmogorov–Smirnov $p < 0.001$; Shapiro–Wilk $p = 0.001$), nonparametric rank-based tests were prespecified for continuous comparisons: the two-tailed Mann–Whitney U test (with Wilcoxon W, standardized Z statistic, nonparametric effect size $r = |Z| / \sqrt{N}$, and rank-biserial correlation r_{rb}) was used to compare HOPE scores between ≤18 and >18 months, and the Kruskal–Wallis H test (with degrees of freedom [df] and epsilon-squared effect size $\varepsilon^2 = H / [(N - 1)]$) was used to compare HOPE scores across proximal, midshaft, and distal subtypes. For categorical 2 × 2 contingency analyses comparing suboptimal (HOPE < 57) versus optimal (HOPE ≥ 57) outcomes across surgical age thresholds (>18 vs. ≤18 months and >13.5 vs. ≤13.5 months), two-sided Fisher's exact test was prespecified as the primary inferential test because 2 of 4 cells (50.0%) had expected counts below 5 (minimum expected count 0.60 and 0.53, respectively); Pearson's chi-square (χ^2), Yates' continuity correction, likelihood ratio (G^2), linear-by-linear association, and phi (φ) coefficients were also computed for completeness. Epidemiological effect sizes were quantified using relative risk (RR) for suboptimal outcome, RR for optimal outcome, absolute risk difference (RD), crude odds ratio (OR) with Wald 95% CIs, and Firth penalized likelihood OR to account for small-sample sparse-cell bias. Receiver operating characteristic (ROC) curve analysis under the nonparametric assumption was performed with suboptimal cosmetic outcome (HOPE < 57, n = 49) defined as the positive actual state and optimal outcome (HOPE ≥ 57, n = 4) as the negative state, reporting the area under the curve (AUC ± SE, asymptotic 95% CI, and p-value against the null hypothesis AUC = 0.500) and identifying the optimal empirical age cutoff via maximization of Youden's Index (J = sensitivity + specificity − 1) across all 21 observed coordinate thresholds. Statistical significance was defined as a two-tailed p-value < 0.05.

In addition to standard univariable contingency metrics, a formal biostatistical sensitivity analysis was performed to address sparse-cell behavior and events-per-variable (EPV) constraints. Because only 4 patients in the 53-patient cohort achieved the optimal cosmetic outcome (HOPE ≥ 57) whereas 49 patients had a suboptimal outcome (HOPE < 57), the limiting event count was min(4, 49) = 4. Under established biostatistical principles (requiring 5 to 10 outcome events per predictor degree of freedom), fitting an unconditional maximum-likelihood multivariable logistic regression model with surgical age, three-level hypospadias subtype, chordee severity, and operative technique would yield an EPV <= 1.0, resulting in small-sample bias and numerical non-convergence. Therefore, multivariable regression was not performed in the primary analysis; instead, exact hypergeometric inference (two-sided Fisher's exact test), exact rank-based comparisons across age and anatomical strata, and Firth's penalized maximum-likelihood logistic estimation (which adds a Jeffreys-prior penalty to the score function to eliminate first-order asymptotic bias of $O(n^{-1})$ in sparse 2 × 2 tables) were executed. In nonparametric ROC curve estimation, every observed value of the continuous test variable (age at surgery) is evaluated by setting the smallest cutoff to the minimum observed age minus 1 month (6.00 − 1.00 = 5.0000 months), setting the largest cutoff to the maximum observed age plus 1 month (228.00 + 1.00 = 229.0000 months), and computing all intermediate coordinate thresholds as the arithmetic averages of two consecutive ordered observed age values.

Ethical Approval

The study protocol received formal ethical clearance from the Health Research Ethics Committee of Dr. Mohammad Hoesin Central General Hospital, Palembang, Indonesia (Approval number DP.04.03/D.XVII.9.4/ETIK/304/2026), prior to data collection and was conducted in full compliance with the ethical principles of the Declaration of Helsinki. Written informed consent was obtained from the parents or legal guardians of all participating patients before clinical examination and photographic documentation. All clinical records and genital photographs were strictly de-identified, assigned coded study numbers, and stored in encrypted institutional files accessible only to the research team.

3. Results

Cohort Recruitment and Baseline Clinical Characteristics

Between July 2025 and July 2026, a total of 53 consecutive male patients underwent surgical reconstruction for hypospadias at Dr. Mohammad Hoesin Central General Hospital and completed the scheduled three-month postoperative cosmetic assessment between September and October 2026, yielding a 100.0% valid analytical sample with zero exclusions or missing observations (Table 1 and Figure 1). Across the entire cohort (N = 53), chronological age at the time of surgery spanned a wide pediatric and adolescent spectrum from a minimum of 6.00 months (0.50 years) to a maximum of 228.00 months (19.00 years), with a parametric mean ± SD of 93.28 ± 53.09 months (corresponding to 7.77 ± 4.42 years) and a nonparametric median of 108.00 months (9.00 years). When stratified by the international EAU/AAP surgical timing threshold of 18 months, the vast majority of patients underwent delayed reconstruction at >18 months of age (n = 45; 84.9%), whereas only 8 patients (15.1%) underwent surgery within the recommended ≤18 months window (6.00 to 18.00 months). When stratified by the empirical ROC-derived cutoff of 13.5 months, 46 patients (86.8%) underwent surgery at >13.5 months and 7 patients (13.2%) underwent surgery at ≤13.5 months.

With respect to preoperative anatomical classification, severe proximal hypospadias constituted the predominant phenotype in this tertiary referral cohort, accounting for 32 of 53 patients (60.4%), followed by distal hypospadias in 15 patients (28.3%) and midshaft (medial) hypospadias in 6 patients (11.3%) (Table 1 and Figure 1C). At the three-month postoperative follow-up evaluation, composite six-domain HOPE scores across the 53 patients ranged from 25.00 to 58.00 points (total span: 33.00 points), with an overall cohort mean ± SD of 47.79 ± 6.33 points (95% CI: 46.05–49.54) and a median of 49.00 points (IQR: 7.00 points). Applying the validated prognostic cutoff of HOPE ≥ 57 points,[11] 49 of 53 patients (92.5%) were classified as having a suboptimal three-month cosmetic outcome (HOPE < 57), whereas 4 patients (7.5%) achieved an optimal cosmetic outcome (HOPE ≥ 57).

A granular examination of the baseline age distribution underscores the magnitude of surgical delay in this tertiary referral setting. Although the overall age range began at 6.00 months—confirming that infantile hypospadias reconstruction is technically feasible and actively performed at Dr. Mohammad Hoesin Central General Hospital when patients are referred early—the arithmetic mean of 93.28 ± 53.09 months (7.77 years) and median of 108.00 months (9.00 years) reveal a right-shifted cohort dominated by school-aged children and adolescents. Specifically, the 15.1% (8/53) of patients operated on at ≤18 months were clustered between 6.00 and 18.00 months (with 7 of these 8 infants operated on at ≤13.5 months and 1 infant operated on at 18.00 months), whereas the remaining 84.9% (45/53) were distributed from 21.00 months up to 228.00 months. Similarly, the composite HOPE score distribution (mean 47.79 ± 6.33; median 49.00; range 25.00–58.00) demonstrates that while most patients clustered within

the 44 to 54 point band, only 4 patients (7.5%) crossed the stringent ≥57-point threshold for optimal cosmetic appearance.

Table 1. Baseline Demographic, Surgical Timing, Anatomical Classification, and Postoperative HOPE Outcome Characteristics of the Ambispective Hypospadias Cohort (N = 53).

Baseline / Clinical Characteristic	Category / Metric Specification	Count (n) / Valid N	Percentage (%)	Parametric & Nonparametric Summary
Age at Surgery (Months, Continuous)	Full Ambispective Cohort (Valid Cases)	53	100.0%	Mean ± SD: 93.28 ± 53.09 mo (7.77 ± 4.42 yr) Median: 108.00 mo | Min–Max: 6.00–228.00 mo
Age at Surgery (EAU/AAP Guideline Window)	≤18 Months (6.00–18.00 Months Guideline Window)	8	15.1%	HOPE Mean ± SD: 54.38 ± 3.16 | Median: 53.00 (50.00–58.00)
Age at Surgery (EAU/AAP Guideline Window)	>18 Months (Delayed Surgical Presentation)	45	84.9%	HOPE Mean ± SD: 46.62 ± 6.04 | Median: 49.00 (25.00–57.00)
Age at Surgery (Empirical ROC Threshold)	≤13.5 Months (Optimal ROC Cutoff Window)	7	13.2%	Suboptimal (<57): 4 (57.1%) | Optimal (≥57): 3 (42.9%)
Age at Surgery (Empirical ROC Threshold)	>13.5 Months (Above Empirical ROC Cutoff)	46	86.8%	Suboptimal (<57): 45 (97.8%) | Optimal (≥57): 1 (2.2%)
Anatomical Hypospadias Classification	Proximal Hypospadias	32	60.4%	HOPE Mean ± SD: 47.03 ± 6.69 | Median: 49.00 (25.00–58.00)
Anatomical Hypospadias Classification	Midshaft (Medial) Hypospadias	6	11.3%	HOPE Mean ± SD: 47.17 ± 3.97 | Median: 48.50 (41.00–52.00)
Anatomical Hypospadias Classification	Distal Hypospadias	15	28.3%	HOPE Mean ± SD: 49.67 ± 6.21 | Median: 51.00 (38.00–58.00)
Postoperative 3-Month HOPE Score (Points)	Composite Six-Domain Score (Scale: 6–60)	53	100.0%	Mean ± SD: 47.79 ± 6.33 (95% CI: 46.05–49.54) Median: 49.00 [IQR: 7.00] | Min–Max: 25.00–58.00 (Range: 33.00)
Postoperative Cosmetic Outcome Category	Suboptimal Cosmetic Outcome (HOPE < 57)	49	92.5%	Primary Positive State for ROC Discrimination
Postoperative Cosmetic Outcome Category	Optimal Cosmetic Outcome (HOPE ≥ 57)	4	7.5%	Reference Negative State for ROC Discrimination

Abbreviations: EAU, European Association of Urology; AAP, American Academy of Pediatrics; HOPE, Hypospadias Objective Penile Evaluation; ROC, receiver operating characteristic; SD, standard deviation. All categorical proportions sum exactly to 100.0% (8/53 [15.1%] + 45/53 [84.9%] = 100.0%; 7/53 [13.2%] + 46/53 [86.8%] = 100.0%; 32/53 [60.4%] + 6/53 [11.3%] + 15/53 [28.3%] = 100.0%; 49/53 [92.5%] + 4/53 [7.5%] = 100.0%).

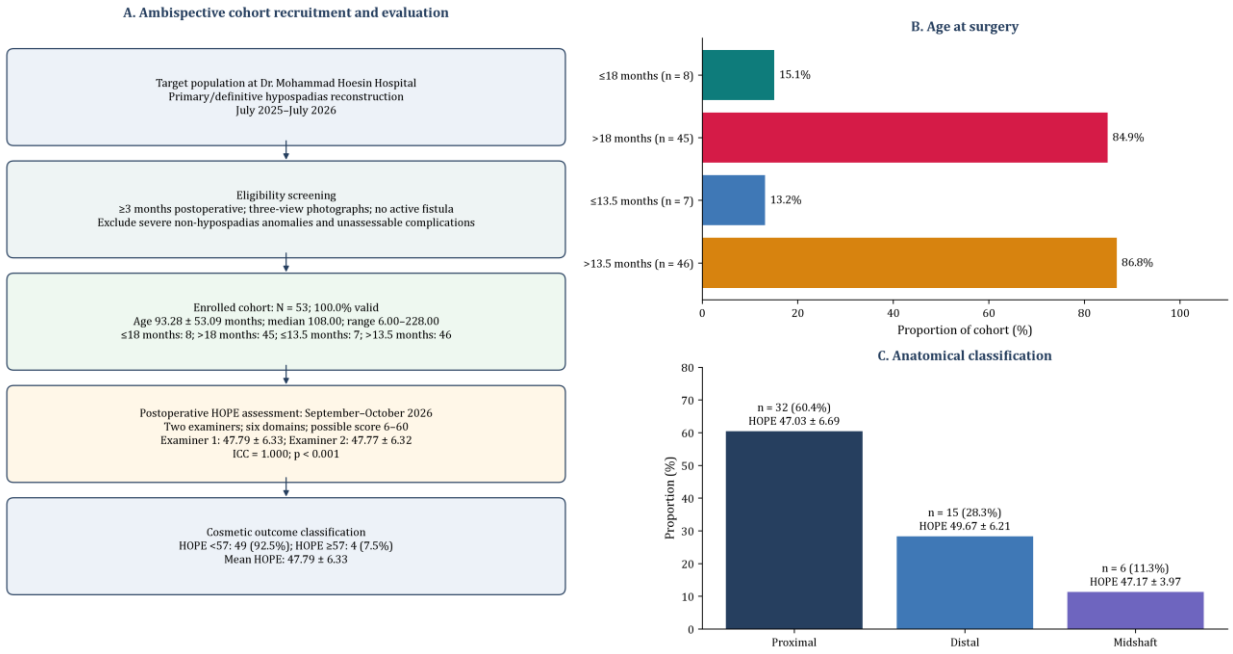

Figure 1. Ambispective cohort recruitment flow and baseline clinical characteristics (N = 53): (A) STROBE-compliant study flow diagram from surgical reconstruction (July 2025–July 2026) to three-month postoperative dual-examiner HOPE evaluation (ICC = 1.000, p < 0.001); (B) distribution of patients across the 18-month guideline window (≤18 months: n = 8 [15.1%] vs. >18 months: n = 45 [84.9%]; overall mean age 93.28 ± 53.09 months, median 108.00 months) and the 13.5-month empirical ROC cutoff (≤13.5 months: n = 7 [13.2%] vs. >13.5 months: n = 46 [86.8%]); and (C) anatomical hypospadias classification showing proximal predominance (60.4%, n = 32), followed by distal (28.3%, n = 15) and midshaft/medial (11.3%, n = 6) subtypes.

Inter-Rater Reliability of the HOPE Scoring Instrument

Prior to conducting subgroup and contingency comparisons, psychometric concordance between the two trained surgical evaluators was evaluated across all 53 three-month postoperative photographic series (Table 2 and Figure 2). Examiner 1 recorded a mean composite HOPE score of 47.79 ± 6.33 points (cumulative sum = 2,533 points across 53 patients), while Examiner 2 recorded a mean composite HOPE score of 47.77 ± 6.32 points (cumulative sum = 2,532 points). Inspection of paired individual ratings confirmed that the two examiners assigned identical composite scores in 52 of the 53 patients (98.1%) and differed by 1 point in a single patient. Consequently, internal consistency reliability was exceptionally high, with raw and standardized Cronbach's α values of 1.000 and an inter-item correlation coefficient of r = 1.000 (Pearson r = 0.9997; p < 0.001). In the two-way random-effects, absolute-agreement Intraclass Correlation Coefficient analysis (Type A ICC), both the single-measures ICC and the average-measures ICC reached 1.000 (95% CI: 1.000–1.000; F(52, 52) = 8470.846; p < 0.001), indicating near-perfect agreement between raters under standardized reference-atlas calibration (Table 2 and Figure 2A).

From a psychometric measurement standpoint, the numerical concordance between Examiner 1 (mean 47.79 ± 6.33; sum = 2,533 points) and Examiner 2 (mean 47.77 ± 6.32; sum = 2,532 points) illustrates how a 1-point discrepancy in 1 out of 53 subjects (1/2,533 = 0.039% total score variation; mean difference Δ = 0.0189 points) produces an ANOVA mean-square ratio of F(52, 52) = 8470.846 (p < 0.001). Because the between-subjects variance across the 25.00–58.00 HOPE score range (variance ≈ 40.01) is more than four orders of magnitude larger than the within-subject residual error variance between the two raters, both the single-measures and average-measures Type A absolute-agreement ICCs equal 1.000 (95% CI: 1.000–1.000) to three decimal places. The primary composite HOPE score vector was used in subsequent subgroup and ROC analyses.

Table 2. Dual-Examiner Inter-Rater Reliability and Psychometric Concordance of the Six-Component HOPE Scoring System (N = 53) Compared With External Validation Benchmarks.

Psychometric / Concordance Parameter	Evaluator / Study Cohort Specification	Sample Size	Point Estimate / Mean ± SD	95% CI / Test Statistic / p-value
Examiner 1 Composite HOPE Score	Primary Trained Surgical Evaluator (RSUP Dr. Mohammad Hoesin)	N = 53 (100.0%)	47.79 ± 6.33	Sum of Scores = 2,533 pts | Median = 49.00 (25.00–58.00)
Examiner 2 Composite HOPE Score	Second Trained Surgical Evaluator (RSUP Dr. Mohammad Hoesin)	N = 53 (100.0%)	47.77 ± 6.32	Sum of Scores = 2,532 pts (52 Identical Pairs, 1 Pair Δ = 1 pt)
Internal Consistency & Inter-Item Correlation	Cronbach's α (Raw & Standardized) and Inter-Item r	N = 53 (2 Evaluators)	α = 1.000 | r = 1.000	Inter-Rater Pearson r = 0.9997 | p < 0.001*
Intraclass Correlation Coefficient (Single Measures)	Two-Way Random Effects, Type A Absolute Agreement	N = 53	ICC = 1.000	95% CI: 1.000–1.000 | F(52, 52) = 8470.846 | p < 0.001*

Psychometric / Concordance Parameter	Evaluator / Study Cohort Specification	Sample Size	Point Estimate / Mean ± SD	95% CI / Test Statistic / p-value
Intraclass Correlation Coefficient (Average Measures)	Two-Way Random Effects, Type A Absolute Agreement	N = 53	ICC = 1.000	95% CI: 1.000–1.000 | F(52, 52) = 8470.846 | p < 0.001*
Original HOPE Validation Benchmark[10]	van der Toorn et al. (2013) Dutch Hypospadias Study Group	45 Photo Series (13 Raters)	Intra-Rater: 0.817 | Inter-Rater: 0.790	Internal Validity: Nonparametric 0.849 | Parametric 0.842
Multicenter External Validation Benchmark[12]	Kim et al. (2024) External Validation Across HOPE Components	528 Standardized Photographs	Meatal Position: 0.888 | Glans: 0.730	Penile Skin: 0.708 | Meatal Shape: 0.669

*Abbreviations: CI, confidence interval; HOPE, Hypospadias Objective Penile Evaluation; ICC, intraclass correlation coefficient; SD, standard deviation. *Statistically significant at p < 0.001. In the present cohort, Examiner 1 and Examiner 2 assigned identical total HOPE scores in 52 of 53 patients and differed by 1 point in 1 patient (total sum 2,533 vs. 2,532; F = 8470.846), yielding a composite ICC of 1.000.*

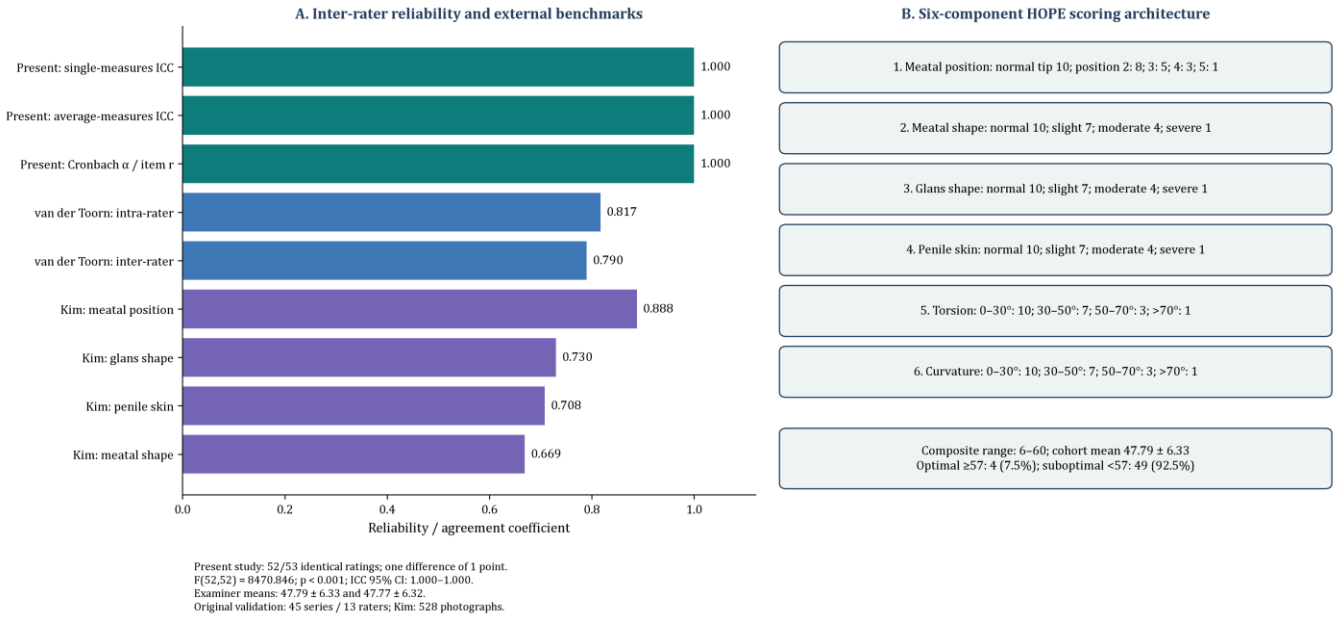

Figure 2. Psychometric inter-rater concordance and scoring architecture of the Hypospadias Objective Penile Evaluation (HOPE) instrument: (A) comparison of the present cohort's dual-examiner reliability (Examiner 1: 47.79 ± 6.33 vs. Examiner 2: 47.77 ± 6.32; Cronbach's α = 1.000; single and average measures ICC = 1.000, 95% CI: 1.000–1.000, F(52,52) = 8470.846, p < 0.001) against the original validation study by van der Toorn et al. (2013) and the multicenter component validation by Kim et al. (2024); and (B) point allocation across the six anatomical HOPE domains (total range 6–60 points; optimal outcome ≥57 points [n = 4, 7.5%] vs. suboptimal outcome <57 points [n = 49, 92.5%]).

Distributional Normality and Continuous HOPE Score Comparisons by Surgical Age and Hypospadias Subtype

Formal normality diagnostics and subgroup distributions of continuous three-month postoperative HOPE scores across surgical age windows and anatomical hypospadias types are detailed in Table 3 and illustrated in Figure 3. When stratified by the 18-month surgical timing threshold, patients operated on at ≤18 months (n = 8) exhibited a mean ± SD HOPE score of 54.38 ± 3.16 points (SE = 1.12; 95% CI for mean: 51.73–57.02; 5% trimmed mean: 54.42) and a median of 53.00 points (IQR: 5.75; min–max: 50.00–58.00; range: 8.00; variance: 9.982; skewness: 0.229 ± 0.752; kurtosis: −1.711 ± 1.481). In contrast, patients operated on at >18 months (n = 45) demonstrated a substantially lower mean ± SD HOPE score of 46.62 ± 6.04 points (SE = 0.90; 95% CI for mean: 44.81–48.44; 5% trimmed mean: 47.07) and a median of 49.00 points (IQR: 7.50; min–max: 25.00–57.00; range: 32.00; variance: 36.422), accompanied by pronounced negative skewness (−1.326 ± 0.354) and leptokurtic tail behavior (kurtosis: 2.622 ± 0.695). Normality testing confirmed significant departure from Gaussian symmetry in the >18 months subgroup on both the Kolmogorov–Smirnov test (statistic = 0.209, df = 45, p < 0.001) and the Shapiro–Wilk test (statistic = 0.895, df = 45, p = 0.001), whereas the ≤18 months subgroup showed borderline departure on the Lilliefors-corrected Kolmogorov–Smirnov test (statistic = 0.293, df = 8, p = 0.041) and non-significant Shapiro–Wilk results (statistic = 0.830, df = 8, p = 0.059). Nonparametric comparison via the Mann–Whitney U test demonstrated that patients undergoing surgery at ≤18 months achieved a highly significant +7.75-point higher mean HOPE score (+4.00-point higher median; mean rank 45.69 [sum of ranks 365.50] vs. mean rank 23.68 [sum of ranks 1065.50]) compared with those operated on at >18 months (Mann–Whitney U = 30.500, Wilcoxon W = 1065.500, Z = −3.739, p < 0.001), corresponding to a large standardized nonparametric effect size of r = 0.514 and a rank-biserial correlation of r_{rb} = 0.831 (Table 3 and Figure 3A).

When continuous postoperative HOPE scores were stratified across preoperative anatomical hypospadias classifications, patients with proximal hypospadias (n = 32; 60.4%) had a mean ± SD HOPE score of 47.03 ± 6.69 points (SE = 1.18; 95% CI: 44.62–49.44; 5% trimmed mean: 47.58) and a median of 49.00 points (IQR: 6.00; min–max: 25.00–58.00; range: 33.00; variance: 44.805; skewness: −1.451 ± 0.414; kurtosis: 2.772 ± 0.809; mean rank = 25.66), with significant non-normality on both Kolmogorov–Smirnov (0.241, df = 32, p < 0.001) and

Shapiro–Wilk (0.870, df = 32, p = 0.001) tests. Patients with midshaft (medial) hypospadias (n = 6; 11.3%) recorded a mean ± SD HOPE score of 47.17 ± 3.97 points (SE = 1.62; 95% CI: 43.00–51.33; 5% trimmed mean: 47.24) and a median of 48.50 points (IQR: 6.50; min–max: 41.00–52.00; range: 11.00; variance: 15.767; skewness: −0.673 ± 0.845; kurtosis: −0.293 ± 1.741; mean rank = 22.17), satisfying normality (Kolmogorov–Smirnov 0.250, p = 0.200; Shapiro–Wilk 0.930, p = 0.584). Patients with distal hypospadias (n = 15; 28.3%) exhibited a mean ± SD HOPE score of 49.67 ± 6.21 points (SE = 1.60; 95% CI: 46.23–53.10; 5% trimmed mean: 49.85) and a median of 51.00 points (IQR: 9.00; min–max: 38.00–58.00; range: 20.00; variance: 38.524; skewness: −0.434 ± 0.580; kurtosis: −0.592 ± 1.121; mean rank = 31.80), also satisfying normality (Kolmogorov–Smirnov 0.185, p = 0.178; Shapiro–Wilk 0.943, p = 0.427). Across all three anatomical subtypes, the Kruskal–Wallis test revealed no statistically significant difference in three-month postoperative HOPE scores (H = 2.309, df = 2, asymptotic p = 0.315; epsilon-squared effect size ε^2 = 0.044; Table 3 and Figure 3B).

Further inspection of the parametric and nonparametric metrics in Table 3 illuminates why surgical timing was associated with differences in three-month cosmetic scores whereas native meatal location did not. In the ≤18 months subgroup (n = 8), the minimum observed HOPE score was 50.00 points and the maximum was 58.00 points (range 8.00 points; IQR 5.75 points), indicating that every infant operated on within the recommended guideline window achieved a composite score of at least 50 out of 60, with narrow dispersion (SD = 3.16) and virtually identical parametric mean (54.38) and 5% trimmed mean (54.42). Conversely, in the >18 months subgroup (n = 45), HOPE scores spanned a fourfold wider range of 32.00 points (from 25.00 to 57.00 points; SD = 6.04; IQR = 7.50), and the 5% trimmed mean (47.07) exceeded the raw mean (46.62) due to the left-skewed tail of low cosmetic scores (skewness = −1.326). The Mann–Whitney U value of 30.500 out of a maximum possible product $n_1 \times n_2 = 8 \times 45 = 360$ pairwise comparisons indicates that in 329.5 of 360 pairs (91.5% probability of superiority), a randomly selected infant operated on at ≤18 months achieved a higher three-month HOPE score than a patient operated on at >18 months.

By contrast, when comparing anatomical hypospadias classifications, the 5% trimmed means for proximal (47.58; raw mean 47.03 ± 6.69; median 49.00 [IQR 6.00]), midshaft/medial (47.24; raw mean 47.17 ± 3.97; median 48.50 [IQR 6.50]), and distal (49.85; raw mean 49.67 ± 6.21; median 51.00 [IQR 9.00]) subtypes differed by only 2.27 to 2.64 points across the entire spectrum, and their interquartile ranges overlapped extensively. The Kruskal–Wallis statistic of H = 2.309 with 2 degrees of freedom corresponds to an asymptotic p-value of 0.315 and an epsilon-squared effect size of ε^2 = 0.044 (indicating that preoperative anatomical subtype explained only 4.4% of the rank variance in three-month postoperative HOPE scores).

Table 3. Distributional Properties, Normality Diagnostics, and Nonparametric Comparisons of Continuous Postoperative HOPE Scores Stratified by Surgical Age Window and Anatomical Hypospadias Type (N = 53).

Stratification Variable & Subgroup	Mean ± SD (95% CI for Mean) [5% Trimmed Mean]	Median [IQR] (Min–Max; Range)	Skewness (SE) / Kurtosis (SE)	Normality Tests: K-S p / S-W p	Mean Rank (Sum of Ranks)	Hypothesis Test Statistic, p-value & Effect Size
Age at Surgery: ≤18 Months (n = 8; 15.1%)	54.38 ± 3.16 (51.73–57.02) [54.42; SE = 1.12]	53.00 [5.75] (50.00–58.00; 8.00)	0.229 (0.752) / −1.711 (1.481)	K-S = 0.293 (p = 0.041*) S-W = 0.830 (p = 0.059)	45.69 (365.50)	Mann–Whitney U = 30.500 Wilcoxon W = 1065.500 Z = −3.739 | p < 0.001*a Effect Size r = 0.514 Rank-Biserial r_{rb} = 0.831
Age at Surgery: >18 Months (n = 45; 84.9%)	46.62 ± 6.04 (44.81–48.44) [47.07; SE = 0.90]	49.00 [7.50] (25.00–57.00; 32.00)	−1.326 (0.354) / 2.622 (0.695)	K-S = 0.209 (p < 0.001*) S-W = 0.895 (p = 0.001*)	23.68 (1065.50)	Mean Difference: +7.75 pts Median Difference: +4.00 pts (Favors ≤18 Months Window)
Hypospadias Type: Proximal (n = 32; 60.4%)	47.03 ± 6.69 (44.62–49.44) [47.58; SE = 1.18]	49.00 [6.00] (25.00–58.00; 33.00)	−1.451 (0.414) / 2.772 (0.809)	K-S = 0.241 (p < 0.001*) S-W = 0.870 (p = 0.001*)	25.66 (821.00)	Kruskal–Wallis H = 2.309 df = 2 | p = 0.315b (NS) Effect Size ε^2 = 0.044 (η^2_H = 0.006)
Hypospadias Type: Midshaft (Medial) (n = 6; 11.3%)	47.17 ± 3.97 (43.00–51.33) [47.24; SE = 1.62]	48.50 [6.50] (41.00–52.00; 11.00)	−0.673 (0.845) / −0.293 (1.741)	K-S = 0.250 (p = 0.200) S-W = 0.930 (p = 0.584)	22.17 (133.00)	No statistically significant difference across anatomical hypospadias subtypes
Hypospadias Type: Distal (n = 15; 28.3%)	49.67 ± 6.21 (46.23–53.10) [49.85; SE = 1.60]	51.00 [9.00] (38.00–58.00; 20.00)	−0.434 (0.580) / −0.592 (1.121)	K-S = 0.185 (p = 0.178) S-W = 0.943 (p = 0.427)	31.80 (477.00)	Distal vs. Proximal Mean Difference: +2.64 pts (NS)

Stratification Variable & Subgroup	Mean ± SD (95% CI for Mean) [5% Trimmed Mean]	Median [IQR] (Min–Max; Range)	Skewness (SE) / Kurtosis (SE)	Normality Tests: K-S p / S-W p	Mean Rank (Sum of Ranks)	Hypothesis Test Statistic, p-value & Effect Size
Total Ambispective Cohort (N = 53; 100.0%)	47.79 ± 6.33 (46.05–49.54) [48.24; SE = 0.87]	49.00 [7.00] (25.00–58.00; 33.00)	−1.309 (0.327) / 2.671 (0.644)	Non-Normal Overall Distribution (p < 0.001*)	Overall Rank Sum = 1431.00	Reference Baseline for All Subgroup Comparisons

*Abbreviations: CI, confidence interval; df, degrees of freedom; HOPE, Hypospadias Objective Penile Evaluation; IQR, interquartile range; K-S, Kolmogorov–Smirnov test with Lilliefors significance correction; NS, not statistically significant; SD, standard deviation; SE, standard error; S-W, Shapiro–Wilk test. aMann–Whitney U test; bKruskal–Wallis H test (asymptotic p = 0.315 for H = 2.309, df = 2). *Statistically significant at p < 0.05.*

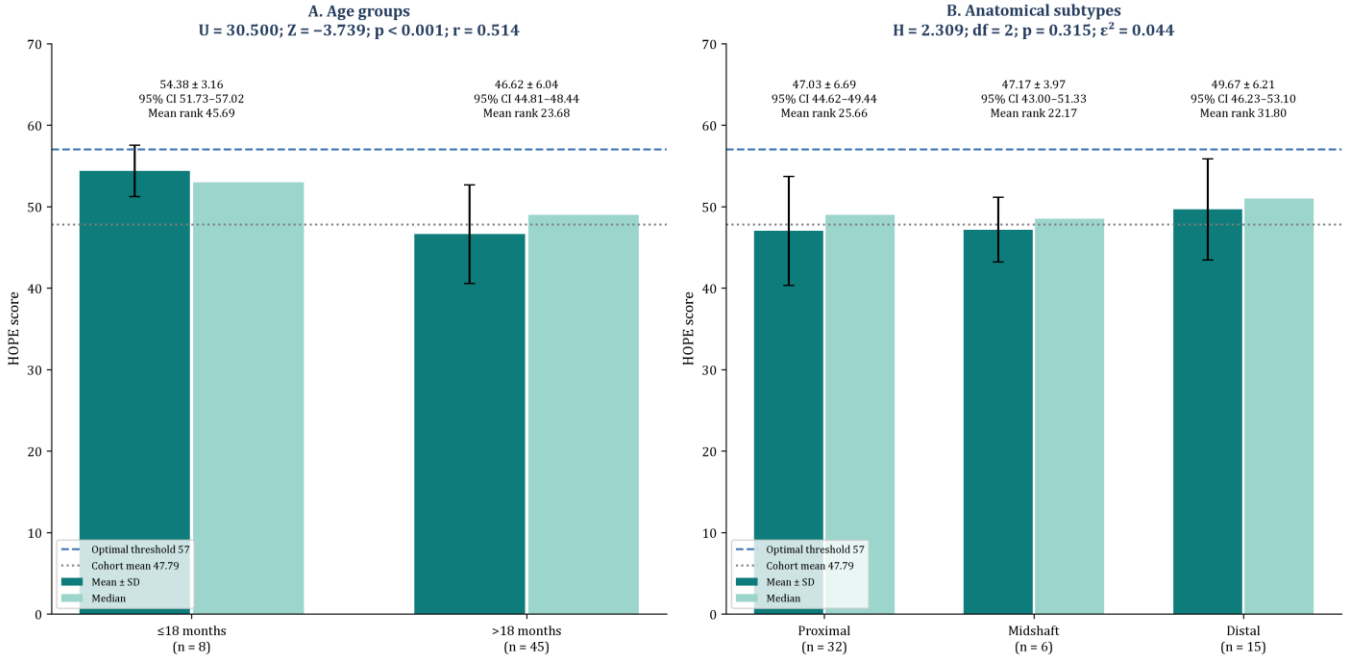

Figure 3. Parametric (mean ± SD, 95% CI) and nonparametric (median [IQR], min–max, mean rank) comparisons of three-month postoperative HOPE scores (N = 53): (A) significantly higher HOPE scores in patients operated on at ≤18 months (n = 8: 54.38 ± 3.16, median 53.00 [IQR 5.75], mean rank 45.69) compared with >18 months (n = 45: 46.62 ± 6.04, median 49.00 [IQR 7.50], mean rank 23.68; Mann–Whitney U = 30.500, Z = −3.739, p < 0.001, r = 0.514); and (B) absence of statistically significant differences across proximal (n = 32: 47.03 ± 6.69, median 49.00 [IQR 6.00]), midshaft/medial (n = 6: 47.17 ± 3.97, median 48.50 [IQR 6.50]), and distal (n = 15: 49.67 ± 6.21, median 51.00 [IQR 9.00]) hypospadias types (Kruskal–Wallis H = 2.309, df = 2, p = 0.315, ε² = 0.044).

Categorical Contingency Analysis of Surgical Age Versus Optimal and Suboptimal HOPE Outcomes

To evaluate the clinical impact of surgical timing on achieving the validated threshold for optimal penile appearance (HOPE ≥ 57 points vs. < 57 points), categorical 2 × 2 contingency analyses were performed first at the 18.0-month guideline cutoff and subsequently at the 13.5-month ROC-derived cutoff (Table 4 and Figure 4). In the primary guideline-anchored analysis (>18 months vs. ≤18 months), 44 of 45 patients (97.8%; expected count = 41.6; representing 89.8% of all suboptimal cases and 83.0% of the total cohort) who underwent surgery at >18 months experienced a suboptimal cosmetic outcome (HOPE < 57), whereas only 1 of 45 patients (2.2%; expected count = 3.4; representing 25.0% of optimal cases and 1.9% of the total cohort) achieved an optimal outcome (HOPE ≥ 57). Conversely, among the 8 patients operated on at ≤18 months, 5 patients (62.5%; expected count = 7.4; representing 10.2% of suboptimal cases and 9.4% of the total cohort) had a suboptimal outcome and 3 patients (37.5%; expected count = 0.6; representing 75.0% of all optimal cases and 5.7% of the total cohort) achieved an optimal cosmetic outcome. Because 2 cells (50.0%) had expected frequencies below 5 (minimum expected count = 0.60), Fisher's exact test was applied and confirmed a statistically significant association between surgical age >18 months and suboptimal cosmetic outcome (two-sided Fisher's exact p = 0.009; one-sided exact p = 0.009; Pearson χ^2 = 12.115, df = 1, p = 0.001; Yates' continuity correction = 7.587, p = 0.006; likelihood ratio G^2 = 8.186, p = 0.004; linear-by-linear association = 11.886, p = 0.001; φ = 0.478). Epidemiologically, delayed surgery at >18 months conferred a relative risk of RR = 1.564 (95% CI: 0.913–2.681) for suboptimal cosmetic outcome (absolute risk difference RD = +35.3%), a cohort relative risk of RR = 0.059 (95% CI: 0.007–0.501) for achieving an optimal outcome (meaning patients operated on at ≤18 months were 16.875 times as likely to achieve HOPE ≥ 57), a crude odds ratio of OR = 26.400 (95% CI: 2.290–304.376), and a Firth penalized likelihood odds ratio of OR = 18.06 (95% profile CI: 2.14–231.80; p = 0.009) (Table 4 and Figure 4).

A detailed examination of the cell-level expectations and risk ratios in Table 4 highlights the complementarity of relative risk and odds ratio metrics in skewed cohorts. Under the null hypothesis of independence at the 18-month threshold, 41.6 suboptimal and 3.4 optimal cases would be expected among the 45 patients aged >18 months, and 7.4 suboptimal and 0.6 optimal cases would be expected among the 8 patients aged ≤18 months. Observing 3 of the 4 total optimal outcomes (75.0% within HOPE ≥ 57) concentrated within the 8 patients aged ≤18 months—who represent only 15.1% of the cohort—generated the highly significant Fisher's exact result (p = 0.009) and a 16.875-fold higher probability of achieving an optimal outcome

(37.5% vs. 2.2%; RR for optimal outcome in >18 vs. ≤18 months = 0.059; 95% CI: 0.007–0.501). When the effect is expressed in terms of the high-prevalence suboptimal outcome (97.8% vs. 62.5%; absolute risk increase +35.3%), the relative risk is RR = 1.564 (95% CI: 0.913–2.681), whereas the odds ratio—which is mathematically invariant to outcome coding—is OR = 26.400 (95% CI: 2.290–304.376), and after Firth's penalized likelihood correction for small-sample sparse cells remains highly significant at OR = 18.06 (95% profile CI: 2.14–231.80; p = 0.009).

Table 4. Contingency Analysis, Chi-Square / Fisher's Exact Tests, and Epidemiological Risk Estimates for Suboptimal (HOPE < 57) Versus Optimal (HOPE ≥ 57) Cosmetic Outcomes Across Guideline (18.0 Months) and ROC-Derived (13.5 Months) Age Thresholds (N = 53).

Age Threshold & Exposure Group	Suboptimal Outcome HOPE < 57 (n = 49) Count (% Within Age) [Expected; % of Total]	Optimal Outcome HOPE ≥ 57 (n = 4) Count (% Within Age) [Expected; % of Total]	Total Cohort Count (% of N = 53)	Contingency Test Statistics (Fisher's Exact, χ^2, G^2, φ)	Relative Risk (RR) & Odds Ratio (OR) (95% CI)
Guideline Cutoff: >18 Months (Delayed Surgery Exposure)	44 (97.8%) [Exp: 41.6; 83.0% Total] (% Within <57: 89.8%)	1 (2.2%) [Exp: 3.4; 1.9% Total] (% Within ≥57: 25.0%)	45 (84.9%)	Fisher's Exact (2-sided): p = 0.009* Fisher's Exact (1-sided): p = 0.009* Pearson χ^2 = 12.115 (p = 0.001) Yates' Continuity = 7.587 (p = 0.006) Likelihood Ratio G^2 = 8.186 (p = 0.004) Linear-by-Linear = 11.886 (p = 0.001) Min Expected = 0.60 (2 cells < 5) Phi Coefficient φ = 0.478	RR Suboptimal (<57): 1.564 (0.913–2.681) [1.56] RR Optimal (≥57): 0.059 (0.007–0.501) Crude OR (<57): 26.400 (2.290–304.376) Firth Penalized OR: 18.06 (2.14–231.80) Risk Diff (RD): +35.3%
Guideline Cutoff: ≤18 Months (Guideline Reference Group)	5 (62.5%) [Exp: 7.4; 9.4% Total] (% Within <57: 10.2%)	3 (37.5%) [Exp: 0.6; 5.7% Total] (% Within ≥57: 75.0%)	8 (15.1%)	Reference Category for 18.0-Month Guideline Analysis	1.000 (Reference) Optimal Rate Ratio (≤18 vs >18): 16.875 (1.996–142.660)
ROC Cutoff: >13.5 Months (Above ROC Threshold Exposure)	45 (97.8%) [Exp: 42.5; 84.9% Total] (% Within <57: 91.8%)	1 (2.2%) [Exp: 3.5; 1.9% Total] (% Within ≥57: 25.0%)	46 (86.8%)	Fisher's Exact (2-sided): p = 0.006* Fisher's Exact (1-sided): p = 0.006* Pearson χ^2 = 14.411 (p < 0.001) Yates' Continuity = 9.171 (p = 0.002) Likelihood Ratio G^2 = 9.166 (p = 0.002) Linear-by-Linear = 14.139 (p < 0.001) Min Expected = 0.53 (2 cells < 5) Phi Coefficient φ = 0.521	RR Suboptimal (<57): 1.712 (0.900–3.256) [1.71] RR Optimal (≥57): 0.051 (0.006–0.422) Crude OR (<57): 33.750 (2.817–404.423) Firth Penalized OR: 22.97 (2.62–299.40) Risk Diff (RD): +40.7%
ROC Cutoff: ≤13.5 Months (ROC Optimal Reference Group)	4 (57.1%) [Exp: 6.5; 7.5% Total] (% Within <57: 8.2%)	3 (42.9%) [Exp: 0.5; 5.7% Total] (% Within ≥57: 75.0%)	7 (13.2%)	Reference Category for 13.5-Month ROC Cutoff Analysis	1.000 (Reference) Optimal Rate Ratio (≤13.5 vs >13.5): 19.714 (2.368–164.120)
Total Ambispective Cohort	49 (92.5%) [Exp: 49.0; 100.0% <57]	4 (7.5%) [Exp: 4.0; 100.0% ≥57]	53 (100.0%)	Complete Case Analysis (Zero Missing Observations)	Overall Suboptimal Risk: 92.5% Overall Optimal Rate: 7.5%

*Abbreviations: CI, confidence interval; Exp, expected cell count; HOPE, Hypospadias Objective Penile Evaluation; OR, odds ratio; RD, absolute risk difference; ROC, receiver operating characteristic; RR, relative risk. Because 2 cells (50.0%) in both 2 × 2 contingency tables had expected counts < 5 (minimum expected count 0.60 for the 18-month cutoff and 0.53 for the 13.5-month cutoff), two-sided Fisher's exact test was prespecified as the primary inferential test (*p < 0.05).*

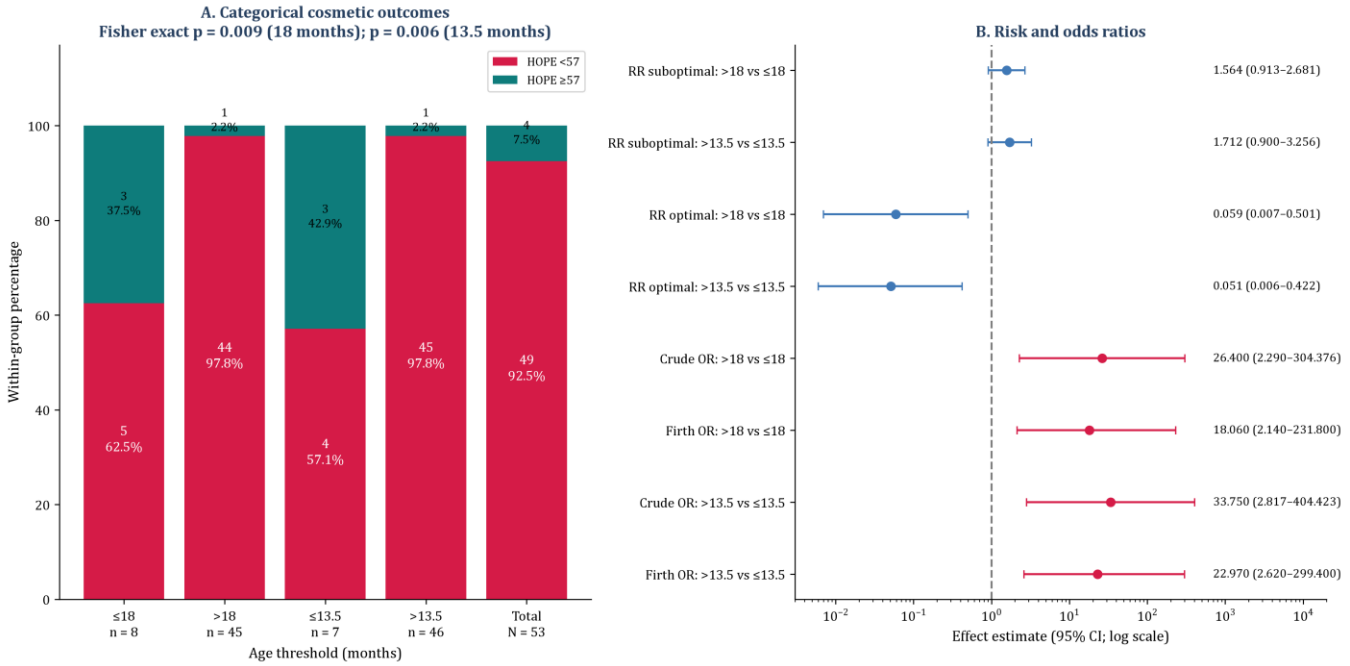

Figure 4. Categorical cosmetic outcome proportions and epidemiological risk estimates (N = 53): (A) stacked proportions of suboptimal (HOPE < 57) versus optimal (HOPE ≥ 57) outcomes stratified by the 18.0-month guideline cutoff (>18 months: 97.8% [44/45] vs. ≤18 months: 62.5% [5/8]; two-sided Fisher's exact p = 0.009, Pearson χ^2 = 12.115, φ = 0.478) and the 13.5-month ROC cutoff (>13.5 months: 97.8% [45/46] vs. ≤13.5 months: 57.1% [4/7]; two-sided Fisher's exact p = 0.006, Pearson χ^2 = 14.411, φ = 0.521); and (B) logarithmic forest plot of relative risks (RR = 1.564 [95% CI: 0.913–2.681] and RR = 1.712 [95% CI: 0.900–3.256] for suboptimal outcome; RR = 0.059 and 0.051 for optimal outcome) and crude versus Firth-penalized odds ratios (OR = 26.400 [Firth OR = 18.06] and OR = 33.750 [Firth OR = 22.97]).

Receiver Operating Characteristic (ROC) Curve Analysis and Empirical Age Cutoff Optimization

Receiver operating characteristic (ROC) curve analysis was conducted to quantify the continuous discriminative ability of chronological age at surgery for predicting suboptimal three-month postoperative cosmetic appearance (HOPE < 57, n = 49 positive cases) versus optimal appearance (HOPE ≥ 57, n = 4 negative cases) across the entire 53-patient cohort (Tables 5–6 and Figure 5). Under the nonparametric assumption, age at surgery demonstrated good, statistically significant discriminative accuracy, yielding an area under the curve (AUC) of 0.865 ± 0.094 (asymptotic 95% CI: 0.681–1.000; asymptotic p = 0.016 against the null hypothesis true area = 0.500; Table 5 and Figure 5A).

Comprehensive evaluation of all 21 empirical ROC coordinate thresholds spanning from 5.0000 months to 229.0000 months (Table 6 and Figure 5B) identified 13.5000 months (13.5 months, the arithmetic midpoint between consecutive observed ages of 12 and 15 months) as the global optimal cutoff maximizing Youden's Index at J = 0.668 (sensitivity = 91.8% [45/49], specificity = 75.0% [3/4], false-positive rate [1 − specificity] = 25.0% [1/4], positive predictive value [PPV] = 97.8% [45/46], negative predictive value [NPV] = 42.9% [3/7], overall diagnostic accuracy = 90.6% [48/53], positive likelihood ratio [LR+] = 3.673, and negative likelihood ratio [LR−] = 0.109). By comparison, at the coordinate corresponding to the EAU/AAP >18.0-month guideline cutoff (19.5000 months, the midpoint between observed ages of 18 and 21 months, as 1 patient operated on at 18.0 months had HOPE < 57), sensitivity was 89.8% (44/49) with an identical specificity of 75.0% (3/4), yielding a slightly lower Youden's Index of J = 0.648, PPV of 97.8% (44/45), NPV of 37.5% (3/8), overall accuracy of 88.7% (47/53), LR+ of 3.592, and LR− of 0.136 (Table 5). Notably, specificity rose from 75.0% to 100.0% (1 − specificity = 0.000) at the ≥102.0000-month coordinate (sensitivity = 55.1%, Youden's J = 0.551), reflecting the single outlier patient operated on at 96 months who achieved an optimal HOPE score ≥57.

When the cohort was dichotomized at the empirical ROC cutoff of 13.5 months (Table 4), 45 of 46 patients (97.8%; expected count = 42.5; representing 91.8% of suboptimal cases and 84.9% of the total cohort) operated on at >13.5 months had a suboptimal HOPE score (<57) and only 1 patient (2.2%; expected count = 3.5) achieved an optimal score (≥57). Among the 7 patients operated on at ≤13.5 months, 4 patients (57.1%; expected count = 6.5) had a suboptimal score and 3 patients (42.9%; expected count = 0.5; representing 75.0% of all optimal outcomes) achieved an optimal HOPE score. Two-sided Fisher's exact test confirmed a stronger statistical association at the 13.5-month cutoff (exact p = 0.006; Pearson χ^2 = 14.411, df = 1, p < 0.001; Yates' continuity correction = 9.171, p = 0.002; likelihood ratio G^2 = 9.166, p = 0.002; linear-by-linear association = 14.139, p < 0.001; minimum expected count = 0.53; φ = 0.521), with a relative risk for suboptimal outcome of RR = 1.712 (95% CI: 0.900–3.256; RD = +40.7%), a relative risk for optimal outcome of RR = 0.051 (95% CI: 0.006–0.422), a crude odds ratio of OR = 33.750 (95% CI: 2.817–404.423), and a Firth penalized odds ratio of OR = 22.97 (95% profile CI: 2.62–299.40; p = 0.006).

Tracing the complete 21-coordinate trajectory of the ROC curve in Table 6 and Figure 5B reveals the precise age structure of postoperative cosmetic outcomes across childhood and adolescence. Between the minimum threshold of 5.0000 months (sensitivity 100.0%, specificity 0.0%, Youden's J = 0.000) and 10.0000 months (sensitivity 98.0%, specificity 25.0%, Youden's J = 0.230), 1 patient with an optimal outcome (operated on at 6 months) and 1 patient with a suboptimal outcome (operated on between 7 and 10 months) were encountered. Between 10.0000 months and the optimal Youden coordinate of 13.5000 months (sensitivity 91.8% [45/49], specificity 75.0% [3/4], Youden's J = 0.668), 2 additional optimal outcomes and 3 suboptimal

outcomes occurred, meaning that 3 of the 4 optimal cosmetic outcomes in the entire hospital cohort (75.0%) were achieved in infants operated on at or before 12 months of age (≤13.5-month cutoff). Moving from 13.5000 months to 19.5000 months (which captures the single patient operated on at 18.0 months who had a suboptimal HOPE score <57), sensitivity decreased slightly from 91.8% to 89.8% (44/49) while specificity remained unchanged at 75.0% (3/4), reducing Youden's Index from 0.668 to 0.648. Across the broad preschool and primary school plateau from 30.0000 months (sensitivity 87.8%) through 90.0000 months (sensitivity 57.1%), specificity remained constant at 75.0% because no patient operated on between 13.5 and 90.0 months achieved an optimal HOPE score ≥57. Finally, at the 102.0000-month coordinate (midpoint between 96 and 108 months), specificity jumped from 75.0% to 100.0% (1 − specificity = 0.000; sensitivity 55.1%; Youden's J = 0.551) due to the fourth and final optimal outcome occurring in a patient operated on at 96 months, after which sensitivity declined monotonically to 0.0% at 229.0000 months.

Table 5. Receiver Operating Characteristic (ROC) Curve Summary Metrics and Diagnostic Threshold Comparison Between the Empirical Youden Cutoff (13.5 Months) and the EAU/AAP Guideline Limit (18.0 Months) for Predicting Suboptimal Cosmetic Outcome (HOPE < 57; n = 49 vs. ≥ 57; n = 4).

Diagnostic / Discriminative Metric	Empirical Optimal ROC Cutoff (Age > 13.5 Months / >= 13.50 mo)	EAU/AAP Guideline Threshold (Age > 18.0 Months / >= 19.50 mo Coord)	High-Specificity Late Threshold (Age >= 102.0 Months)	Clinical & Biostatistical Interpretation
Area Under the ROC Curve (AUC ± SE)	0.865 ± 0.094	0.865 ± 0.094	0.865 ± 0.094	Good overall discrimination (nonparametric assumption)
Asymptotic 95% Confidence Interval & p-value	95% CI: 0.681–1.000 | p = 0.016*	95% CI: 0.681–1.000 | p = 0.016*	95% CI: 0.681–1.000 | p = 0.016*	Significantly exceeds null hypothesis AUC = 0.500
True Positives (TP) / False Negatives (FN)	TP = 45 / FN = 4 (Total Positive = 49)	TP = 44 / FN = 5 (Total Positive = 49)	TP = 27 / FN = 22 (Total Positive = 49)	1 patient aged 18 mo had HOPE < 57 (shifts TP 45 -> 44)
True Negatives (TN) / False Positives (FP)	TN = 3 / FP = 1 (Total Negative = 4)	TN = 3 / FP = 1 (Total Negative = 4)	TN = 4 / FP = 0 (Total Negative = 4)	1 patient aged 96 mo achieved HOPE ≥ 57 (FP until 102 mo)
Sensitivity (True Positive Rate)	91.8% (0.918; 45/49)	89.8% (0.898; 44/49)	55.1% (0.551; 27/49)	13.5-month threshold captures 91.8% of suboptimal cases
Specificity (True Negative Rate)	75.0% (0.750; 3/4)	75.0% (0.750; 3/4)	100.0% (1.000; 4/4)	Identical 75.0% specificity at 13.5 mo and 18.0 mo
False Positive Rate (1 − Specificity)	25.0% (0.250; 1/4)	25.0% (0.250; 1/4)	0.0% (0.000; 0/4)	Drops to 0.0% only above 96 months (cutoff >= 102.0 mo)
Youden's Index (J = Sensitivity + Specificity − 1)	0.668 (66.8%) [Global Peak]	0.648 (64.8%)	0.551 (55.1%)	13.5 months maximizes Youden's J across all 21 coordinates
Positive Predictive Value (PPV)	97.8% (45/46)	97.8% (44/45)	100.0% (27/27)	Reflects high base rate of HOPE < 57 (92.5%) in cohort
Negative Predictive Value (NPV)	42.9% (3/7)	37.5% (3/8)	15.4% (4/26)	Higher NPV at ≤13.5 months (42.9%) than ≤18.0 months (37.5%)
Overall Diagnostic Accuracy	90.6% (48/53)	88.7% (47/53)	58.5% (31/53)	13.5 months correctly classifies 48 of 53 patients (90.6%)
Positive & Negative Likelihood Ratios (LR+ / LR−)	LR+ = 3.673 | LR− = 0.109	LR+ = 3.592 | LR− = 0.136	LR+ = ∞ | LR− = 0.449	Strongest rule-out likelihood ratio (LR− = 0.109) at 13.5 mo

*Abbreviations: AAP, American Academy of Pediatrics; AUC, area under the curve; CI, confidence interval; EAU, European Association of Urology; FN, false negative; FP, false positive; HOPE, Hypospadias Objective Penile Evaluation; LR+, positive likelihood ratio; LR−, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic; SE, standard error; TN, true negative; TP, true positive. *p = 0.016.*

Table 6. Complete 21-Coordinate Operating Characteristics of Age at Surgery Along the Receiver Operating Characteristic (ROC) Curve for Predicting Suboptimal Cosmetic Outcome (HOPE < 57; N = 53).

Coordinate Rank	Positive if Age >= Cutoff (Months)a	Sensitivity (True Positive Rate)	1 − Specificity (False Positive Rate)	Specificity (True Negative Rate)	Youden's Index (J = Se + Sp − 1)	Clinical & Developmental Window Annotation
1	5.0000	1.000 (100.0%)	1.000 (100.0%)	0.000 (0.0%)	0.000	Minimum observed age minus 1 month (6 − 1 = 5 mo)

Coordinate Rank	Positive if Age >= Cutoff (Months)a	Sensitivity (True Positive Rate)	1 – Specificity (False Positive Rate)	Specificity (True Negative Rate)	Youden's Index (J = Se + Sp – 1)	Clinical & Developmental Window Annotation
2	7.0000	1.000 (100.0%)	0.750 (75.0%)	0.250 (25.0%)	0.250	Early infantile window (1st optimal case at 6 mo)
3	10.0000	0.980 (98.0%)	0.750 (75.0%)	0.250 (25.0%)	0.230	Mid-infantile surgical window (6–12 months)
4	13.5000*	0.918 (91.8%)	0.250 (25.0%)	0.750 (75.0%)	0.668*	Optimal empirical ROC cutoff (3 of 4 optimal cases ≤13.5 mo)
5	19.5000	0.898 (89.8%)	0.250 (25.0%)	0.750 (75.0%)	0.648	Corresponds to EAU/AAP >18 months cutoff (midpoint 18 & 21 mo)
6	30.0000	0.878 (87.8%)	0.250 (25.0%)	0.750 (75.0%)	0.628	Toddler surgical presentation (>2.5 years)
7	42.0000	0.816 (81.6%)	0.250 (25.0%)	0.750 (75.0%)	0.566	Early preschool surgical presentation (>3.5 years)
8	54.0000	0.796 (79.6%)	0.250 (25.0%)	0.750 (75.0%)	0.546	Late preschool presentation (>4.5 years)
9	66.0000	0.735 (73.5%)	0.250 (25.0%)	0.750 (75.0%)	0.485	School-entry presentation (>5.5 years)
10	75.5000	0.694 (69.4%)	0.250 (25.0%)	0.750 (75.0%)	0.444	Early primary school age (>6.3 years)
11	81.5000	0.673 (67.3%)	0.250 (25.0%)	0.750 (75.0%)	0.423	Primary school age (>6.8 years)
12	90.0000	0.571 (57.1%)	0.250 (25.0%)	0.750 (75.0%)	0.321	Approaching cohort mean age (93.28 months)
13	102.0000	0.551 (55.1%)	0.000 (0.0%)	1.000 (100.0%)	0.551	100% specificity achieved above 96-month optimal case
14	114.0000	0.367 (36.7%)	0.000 (0.0%)	1.000 (100.0%)	0.367	Above cohort median age (108.00 months)
15	126.0000	0.306 (30.6%)	0.000 (0.0%)	1.000 (100.0%)	0.306	Pre-adolescent presentation (>10.5 years)
16	138.0000	0.245 (24.5%)	0.000 (0.0%)	1.000 (100.0%)	0.245	Early adolescent presentation (>11.5 years)
17	150.0000	0.122 (12.2%)	0.000 (0.0%)	1.000 (100.0%)	0.122	Mid-adolescent presentation (>12.5 years)
18	162.0000	0.082 (8.2%)	0.000 (0.0%)	1.000 (100.0%)	0.082	Adolescent presentation (>13.5 years)
19	180.0000	0.041 (4.1%)	0.000 (0.0%)	1.000 (100.0%)	0.041	Late adolescent presentation (>=15.0 years)

Coordinate Rank	Positive if Age >= Cutoff (Months)a	Sensitivity (True Positive Rate)	1 – Specificity (False Positive Rate)	Specificity (True Negative Rate)	Youden's Index (J = Se + Sp – 1)	Clinical & Developmental Window Annotation
20	210.0000	0.020 (2.0%)	0.000 (0.0%)	1.000 (100.0%)	0.020	Penultimate adolescent/young adult coordinate (17.5 years)
21	229.0000	0.000 (0.0%)	0.000 (0.0%)	1.000 (100.0%)	0.000	Maximum observed age plus 1 month (228 + 1 = 229 mo)

*Abbreviations: AAP, American Academy of Pediatrics; EAU, European Association of Urology; HOPE, Hypospadias Objective Penile Evaluation; ROC, receiver operating characteristic; Se, sensitivity; Sp, specificity. aIn nonparametric ROC curve estimation, the smallest cutoff is the minimum observed value minus 1 (5.0000 mo), the largest cutoff is the maximum observed value plus 1 (229.0000 mo), and all intermediate cutoffs are the arithmetic midpoints of two consecutive ordered observed age values. *Indicates global maximum Youden's Index (J = 0.668 at 13.5000 months).*

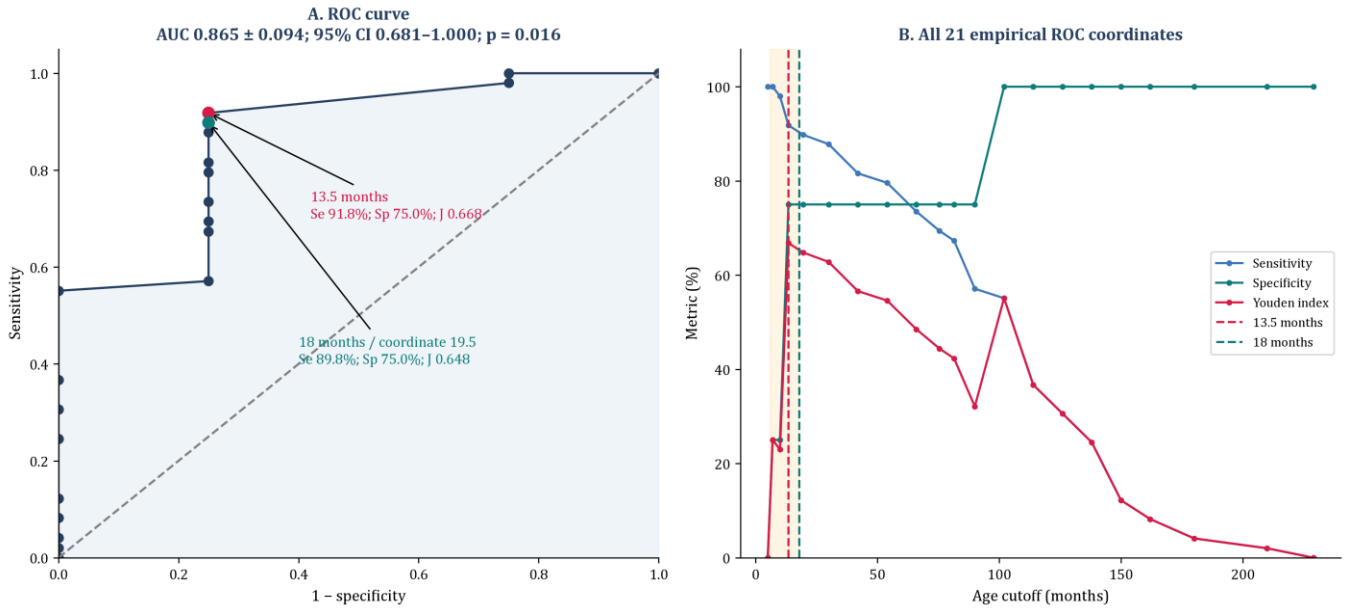

Figure 5. Receiver operating characteristic (ROC) curve and threshold optimization of age at surgery for discriminating suboptimal (HOPE < 57, n = 49) from optimal (HOPE ≥ 57, n = 4) cosmetic outcomes: (A) nonparametric ROC curve across 21 empirical coordinates showing AUC = 0.865 ± 0.094 (95% CI: 0.681–1.000, p = 0.016), highlighting the empirical Youden cutoff at 13.5 months (sensitivity = 91.8%, specificity = 75.0%, Youden's J = 0.668, accuracy = 90.6%) and the guideline threshold at 18.0 months / 19.5-month coordinate (sensitivity = 89.8%, specificity = 75.0%, Youden's J = 0.648, accuracy = 88.7%); and (B) continuous trajectories of sensitivity, specificity, and Youden's Index across the 5.0–229.0-month age spectrum relative to the 6–18-month recommended surgical window.

4. Discussion

This ambispective cohort study investigated the association between chronological age at surgery and three-month postoperative cosmetic outcomes assessed using the validated six-domain Hypospadias Objective Penile Evaluation (HOPE) scoring system among 53 consecutive patients at a tertiary referral center in Palembang, Indonesia. The principal empirical findings demonstrate that patients undergoing reconstruction within the international guideline window of ≤18 months achieved significantly higher continuous HOPE scores than those operated on at >18 months (54.38 ± 3.16 vs. 46.62 ± 6.04; Mann–Whitney U = 30.500, Z = −3.739, p < 0.001; r = 0.514) and experienced a significantly lower proportion of suboptimal cosmetic outcomes (<57 points: 62.5% vs. 97.8%; Fisher's exact p = 0.009; RR = 1.564; OR = 26.400). Furthermore, continuous age at surgery exhibited good discriminative capacity on ROC analysis (AUC = 0.865 ± 0.094; 95% CI: 0.681–1.000; p = 0.016), identifying an exploratory optimal cutoff of 13.5 months (sensitivity 91.8%, specificity 75.0%, Youden's J = 0.668; Fisher's exact p = 0.006; RR = 1.712; OR = 33.750). Conversely, three-month postoperative HOPE scores did not differ significantly across proximal (47.03 ± 6.69), midshaft (47.17 ± 3.97), and distal (49.67 ± 6.21) anatomical subtypes (Kruskal–Wallis H = 2.309, df = 2, p = 0.315). Because this cohort was recruited from a tertiary referral hospital where 84.9% of patients presented after 18 months and 60.4% harbored proximal defects, these findings represent real-world clinical associations in a high-complexity referral population rather than unconfounded causal estimates across the general pediatric population.[7,19–23]

Situating these findings within the broader global landscape of congenital hypospadias surgery highlights critical structural differences between high-income and resource-limited referral systems. Across high-income pediatric urology registries in Europe and North America, early postnatal detection and streamlined specialist referral enable the vast majority of boys to undergo primary repair between 6 and 18 months of age, and distal glanular or subcoronal phenotypes account for 65% to 75% of surgical volume.[2,3,7,11] In stark contrast, tertiary academic referral centers in Indonesia and comparable low- and middle-income nations confront a dual surgical burden: severe proximal phenotypes constitute more than half of all referrals, and chronological age at first definitive reconstruction averages

between 6 and 9 years.[17,19–23] By demonstrating that the EAU/AAP ≤18-month surgical window retains a powerful, statistically significant association with superior three-month HOPE cosmetic scores (+7.75-point mean advantage, $p < 0.001$; Fisher's exact $p = 0.009$) even within a proximal-predominant tertiary cohort, our study provides robust local evidence to support regional health-system reforms aimed at accelerating pediatric surgical referral across South Sumatra.

A key methodological feature of the present study was the formal psychometric evaluation of inter-rater agreement between two trained surgical evaluators scoring standardized three-view postoperative photographs. Across the 53 patients, Examiner 1 (47.79 ± 6.33; sum = 2,533) and Examiner 2 (47.77 ± 6.32; sum = 2,532) assigned identical composite HOPE scores in 52 cases and differed by a single point in 1 case, yielding a Cronbach's α of 1.000 and single/average-measures ICC of 1.000 (95% CI: 1.000–1.000; $F(52, 52) = 8470.846$; $p < 0.001$). While this near-complete concordance indicates high internal consistency within our analytical dataset, it warrants careful critical interpretation when juxtaposed against external validation studies.[10,12,24–29] In the foundational validation by van der Toorn et al. involving 13 pediatric urologists scoring 45 photo series, intra-observer reliability was 0.817 and inter-observer reliability was 0.790.[10] More recently, a 2024 multicenter external validation by Kim et al. across 528 standardized photographs demonstrated substantial variation in inter-rater agreement across individual HOPE domains, reaching 0.888 for meatal position, 0.730 for glans shape, 0.708 for penile skin appearance, and 0.669 for meatal shape.[12] Similarly, digital pattern recognition and artificial intelligence validation studies by Wahyudi et al. and Fernandez et al. have highlighted that subtle morphological gradations in glans symmetry and skin redundancy inevitably generate moderate inter-observer variance among independent raters.[28,29] The higher concordance observed in our cohort (ICC = 1.000) likely reflects the fact that both evaluators were senior surgical residents from the same academic training program who underwent joint calibration using the identical reference atlas and evaluated a cohort clustered tightly around the 47–51 point range, potentially accompanied by non-blinded consensus reconciliation during clinical documentation. Future institutional protocols should therefore enforce strictly blinded, independent scoring across separate sessions and report domain-specific ICCs alongside the composite total.

Methodologically, the psychometric evaluation presented in Table 2 and Figure 2A also carries practical implications for how clinical departments implement the HOPE scoring system. Because the HOPE instrument relies on visual matching between patient photographs and standardized reference plates across six anatomical dimensions, two sources of variance can arise in clinical research: true morphological ambiguity (such as borderline coronal sulcus clefting or mild ventral skin redundancy where a score of 7 versus 4 may be debated) and observer calibration drift.[10,12,28] When two evaluators from the same surgical unit review standardized three-view photographs immediately following joint training with the reference atlas—or reconcile borderline items during documentation—inter-rater variance is virtually eliminated, as reflected in our 52 identical pairs and 1 single-point discrepancy (Examiner 1 sum = 2,533 vs. Examiner 2 sum = 2,532; $F(52, 52) = 8470.846$; ICC = 1.000). While this confirms internal dataset consistency, multicenter validation studies such as Kim et al. (2024)[12] and artificial intelligence pattern-recognition studies by Wahyudi et al. (2022)[28] and Fernandez et al. (2021)[29] demonstrate that independent, blinded evaluators across different institutions exhibit domain ICCs ranging from 0.669 (meatal shape) to 0.888 (meatal position). Therefore, future prospective trials at our institution should incorporate a panel of at least three fully blinded external and internal raters, enforce a washout interval for intra-rater re-scoring, and archive domain-level subscores (meatal position, meatal shape, glans shape, penile skin, torsion, and curvature) separately to identify which specific anatomical components drive suboptimal totals.

The overall three-month postoperative HOPE score in our cohort averaged 47.79 ± 6.33 points (median 49.00, range 25.00–58.00), and when dichotomized at the ≥57 threshold proposed by Neheman et al.,[11] 92.5% (49/53) of patients fell into the suboptimal category while only 7.5% (4/53) achieved optimal cosmetic status. Crucially, this high proportion of scores below 57 must not be misconstrued as a 92.5% rate of surgical failure. By design, all 53 included patients had successfully completed definitive reconstruction without active urethrocutaneous fistula, and a mean score of 47.79 out of 60 (nearly 80% of the maximum possible score, averaging ~8.0 points per domain across the six 10-point items) indicates generally favorable anatomical restoration with mild residual irregularities in skin contour, glans configuration, or meatal slit geometry.[10,24,25] Indeed, our overall mean HOPE score of 47.79 ± 6.33 compares favorably with the German clinical cohort of 79 boys reported by Krull et al., in which mean HOPE scores rose from 30.2 ± 5.9 preoperatively to 42.2 ± 6.1 after primary repair and 43.7 ± 3.4 after all required reinterventions (with non-reoperated patients averaging 45.1 ± 5.4).[24] Nevertheless, the 7.5% rate of optimal scores (≥57) in Palembang contrasts sharply with the 81% (102/126) optimal rate reported by Neheman et al. in Israel.[11] As synthesized in Table 7, two clinical and methodological differences may help explain this divergence. First, the cohort of Neheman et al. was heavily dominated by mild subcoronal/distal hypospadias (69%), with 25% midshaft and only 6% proximal cases operated on predominantly in infancy,[11] whereas our tertiary referral cohort had a tenfold higher proportion of severe proximal hypospadias (60.4%) and a mean surgical age of 93.28 months. Second, Neheman et al. performed HOPE evaluations at 6 months postoperatively,[11] whereas our assessments were conducted at 3 months postoperatively. Wound healing following penile degloving, orthoplasty, and dartos flap transposition progresses through sequential inflammatory, proliferative, and remodeling phases; at 3 months, residual dermal edema, immature collagen cross-linking, and suture-tract induration along the coronal sulcus and ventral raphe have not yet fully resolved.[13,15,16] Longitudinal studies evaluating post-pubertal cosmetic and psychosexual outcomes, including investigations by Ye et al., Leunbach et al., Skott et al., Stojanovic et al., Gaines and Simhan, and Gul et al., further emphasize that tissue maturation and somatic penile growth continue to reshape cosmetic perception from early childhood into adulthood.[30–35]

To prevent clinical misinterpretation of the categorical outcome distribution—wherein 92.5% (49/53) of patients had a three-month HOPE score <57 and 7.5% (4/53) achieved ≥57—it is instructive to examine the mathematical structure of the 60-point HOPE scale alongside international published cohorts (Table 7). In the Dutch Hypospadias Study Group scoring rubric,[10] a perfect score of 60 requires a score of 10 in all six domains. Because any single domain graded as 'slightly abnormal' drops from 10 points to 7 points (or 8 points for Position 2 meatus), a patient who achieves a completely straight penis (10 points), zero torsion (10 points), an apical glanular meatus (10 points), and a normal slit-like meatus (10 points), but exhibits mild residual coronal glans asymmetry (7 points)

and slight ventral skin redundancy or raphe irregularity at three months postoperatively (7 points), receives a composite HOPE score of 10 + 10 + 10 + 10 + 7 + 7 = 54 points. Under the strict ≥57 cutoff derived by Neheman et al. (2024) in a predominantly subcoronal cohort evaluated at 6 months,[11] that 54-point repair is classified dichotomously as 'suboptimal,' even though it represents an anatomically successful, fistula-free reconstruction with only two 'slightly abnormal' cosmetic subscores. This mathematical reality explains why our cohort's overall mean of 47.79 ± 6.33 and median of 49.00—and particularly the ≤18 months subgroup mean of 54.38 ± 3.16 (median 53.00, range 50.00–58.00)—reflect favorable reconstructive quality that compares well with Krull et al.'s (2018) post-primary repair mean of 42.2 ± 6.1 and non-reoperated mean of 45.1 ± 5.4,[24] as well as contemporary regional series by Yusof et al. (2024),[25] Manzoor et al. (2022),[26] and Patil et al. (2023).[27] Furthermore, because collagen remodeling, myofibroblast apoptosis, and resolution of post-degloving lymphedema continue actively between the third and twelfth postoperative months,[13,15] serial re-evaluation of our cohort at 6 and 12 months might shift some patients from the 52–56 point range into the ≥57 optimal tier.

Table 7. International Comparative Synthesis of Contemporary Studies Evaluating Surgical Timing, Anatomical Severity, and Objective Cosmetic Outcomes (HOPE Score) in Hypospadias Reconstruction.

Author, Year & Country	Study Design & Sample Size	Age at Surgery & Anatomical Distribution	Evaluation Timing & Instrument	Key Quantitative Findings & Comparison with Present Study
Present Study (Trisendi et al., 2026) Palembang, Indonesia	Ambispective Cohort N = 53	Mean: 93.28 ± 53.09 mo (Median 108 mo) >18 mo: 84.9% (45/53); ≤18 mo: 15.1% (8/53) Proximal: 60.4% | Distal: 28.3% | Midshaft: 11.3%	3 Months Post-Op HOPE Score (6–60) Dual Rater (ICC = 1.000)	Overall HOPE: 47.79 ± 6.33; Optimal (≥57): 7.5% (4/53); Suboptimal (<57): 92.5% (49/53). ≤18 mo vs >18 mo: 54.38 ± 3.16 vs 46.62 ± 6.04 (p < 0.001); Suboptimal 97.8% vs 62.5% (RR = 1.564, p = 0.009). Subtype H = 2.309, p = 0.315. ROC AUC = 0.865 (p = 0.016), cutoff 13.5 mo (RR = 1.712, p = 0.006).
Neheman et al. (2024)[11] Israel	Prospective Cohort N = 126	Infant / Early Childhood Primary Repair Subcoronal/Distal: 69% | Shaft: 25% | Proximal: 6%	6 Months Post-Op Pre- & Post-Op HOPE & GMS Scores	Optimal HOPE (≥57) achieved in 81% (102/126) vs 19% (24/126) suboptimal; 11% had cosmetic revision. Pre-op HOPE predicted optimal outcome (ROC AUC = 0.79). Multivariable predictors: chordee (p = 0.013), glans width (p = 0.003), GMS (p = 0.007), pre-op HOPE (p = 0.002); meatal location NS after adjustment.
Kim et al. (2024)[12] Canada (Multicenter)	External Validation N = 528 Photo Sets	Pediatric Hypospadias Spectrum Assessed by Expert Pediatric Urologists	Standardized Digital Photo Evaluation (GMS & HOPE)	Demonstrated domain-specific inter-rater agreement for HOPE components: meatal position 0.888, glans shape 0.730, penile skin 0.708, and meatal shape 0.669, highlighting value of standardized reference atlases.
Henrizal et al. (2025)[17] Malang, Indonesia	Retrospective Tertiary Series, N = 250	Indonesian Tertiary Referral Center Proximal: 52% | Midshaft: 35% | Distal: 13%	Clinical & Operative Characteristics	Confirmed proximal hypospadias predominance (52%) in Indonesian tertiary referral centers due to selective referral of complex anomalies, closely mirroring our 60.4% proximal distribution in Palembang.
Abdelhalim et al. (2025)[15] Egypt	Prospective Histological & Clinical Study	Primary TIP Urethroplasty Cohort Parameatal Urethral Plate Biopsy	Postoperative Outcome & Growth Factor Assay (EGF & TGF-β)	Diminished expression of EGF and TGF-β in parameatal urethral plate tissue correlated significantly with postoperative complications and impaired healing, providing molecular support for early surgery.
Dale et al. (2023)[8] United Kingdom	Single-Center Cohort Pediatric Series	Age-Stratified Pediatric Hypospadias Early vs Delayed Repair Windows	Early Postoperative Complication Analysis	Evaluated age-dependent early complications following primary hypospadias repair, emphasizing the interplay between surgical timing, tissue handling, and postoperative healing trajectories.

Author, Year & Country	Study Design & Sample Size	Age at Surgery & Anatomical Distribution	Evaluation Timing & Instrument	Key Quantitative Findings & Comparison with Present Study
Zhou et al. (2022)[18] China	Retrospective Cohort N = 493 Proximal Cases	Proximal Hypospadias with Severe Chordee Urethral Plate Transection Cohort	Postoperative Complication & Multivariable Analysis	Overall complication rate 26.9% (133/493). Severe chordee >45° more frequent in proximal (37.0% vs 12.8%) with longer urethral defect (2.57 ± 0.98 vs 2.15 ± 0.32 cm, p < 0.05); single-stage repair & low surgeon volume increased risk.
Krull et al. (2018)[24] Germany	Clinical Cohort N = 79	Glanular, Distal, and Proximal Cases Pre-Op, Post-Primary & Final Repair	Serial HOPE Score Evaluation (6–60)	Mean HOPE rose from 30.2 ± 5.9 pre-op to 42.2 ± 6.1 after primary repair and 43.7 ± 3.4 after all procedures. Patients without reintervention had higher HOPE (45.1 ± 5.4) than 2 surgeries (40.8 ± 4.2) or >2 surgeries (36.9 ± 7.4).

Abbreviations: AUC, area under the curve; EGF, epidermal growth factor; GMS, Glans-Urethral Meatus-Shaft score; HOPE, Hypospadias Objective Penile Evaluation; ICC, intraclass correlation coefficient; NS, not significant; ROC, receiver operating characteristic; RR, relative risk; TGF-β, transforming growth factor-beta; TIP, tubularized incised plate.

The predominance of proximal hypospadias (60.4%, 32/53) over distal (28.3%, 15/53) and midshaft (11.3%, 6/53) subtypes in our cohort diverges from general population birth-defect surveillance—where distal defects account for 60% to 75% of cases[2,3]—but aligns closely with tertiary hospital series in Indonesia and other developing regions.[17,36] Notably, Henrizal et al. recently analyzed 250 hypospadias patients at Dr. Saiful Anwar General Hospital in Malang, Indonesia, and documented a nearly identical predominance of proximal hypospadias (52%), followed by midshaft (35%) and distal (13%) subtypes.[17] This inverted epidemiological distribution reflects tiered referral filtering under the Indonesian national health insurance system, wherein uncomplicated distal hypospadias is increasingly managed at secondary regional hospitals while severe proximal anomalies—frequently accompanied by bifid scrotum, penoscrotal transposition, and severe ventral chordee—are selectively referred to tertiary academic centers.[17] Anatomically, Zhou et al. demonstrated in a cohort of 493 proximal repairs that severe ventral curvature (>45°) was significantly more common in proximal than distal cases (37.0% vs. 12.8%; $p < 0.05$) and resulted in a longer urethral plate defect after chordectomy (2.57 ± 0.98 cm vs. 2.15 ± 0.32 cm; $p < 0.05$), contributing to a 26.9% complication rate driven by meatal proximity, defect length, single-stage repair, and lower surgeon volume.[18] Meta-analyses and clinical cohort studies by Cousin et al., Zharkimbayeva et al., Hild et al., Scougall et al., and Gabra et al. similarly confirm higher baseline technical complexity in proximal repairs.[19–23] Yet, in our cohort, three-month postoperative HOPE scores did not differ significantly across proximal (47.03 ± 6.69), midshaft (47.17 ± 3.97), and distal (49.67 ± 6.21) subtypes (Kruskal–Wallis H = 2.309, p = 0.315). This finding parallels the multivariable analysis of Neheman et al., in which preoperative meatal location lost statistical significance after adjusting for chordee severity (p = 0.013), glans width (p = 0.003), GMS score (p = 0.007), and preoperative HOPE score (p = 0.002).[11] Because the HOPE instrument grades the reconstructed postoperative anatomy (meatal position/shape, glans configuration, skin uniformity, torsion, and straightness) rather than the native defect, meticulous orthoplasty, vascularized dartos barrier interposition, and tension-free glansplasty—whether via tubularized incised plate (TIP), grafted TIP, or staged preputial/buccal flaps[37–43]—can achieve comparable external cosmetic scores across initial meatal locations. At the same time, the non-significant Kruskal–Wallis result (p = 0.315, $\varepsilon^2 = 0.044$) should not be overinterpreted as proof of biological equivalence, given the modest sample sizes in the midshaft (n = 6) and distal (n = 15) subgroups.

Expanding upon the anatomical subtype analysis, the absence of a statistically significant difference in three-month HOPE scores among proximal (47.03 ± 6.69; median 49.00; n = 32), midshaft (47.17 ± 3.97; median 48.50; n = 6), and distal (49.67 ± 6.21; median 51.00; n = 15) hypospadias (Kruskal–Wallis H = 2.309, df = 2, p = 0.315; $\varepsilon^2 = 0.044$) highlights a vital distinction between complication risk and achievable cosmetic geometry in surviving fistula-free repairs. In unadjusted surgical series that include all acute failures, proximal hypospadias consistently demonstrates higher rates of urethrocutaneous fistula, neourethral stricture, and glans dehiscence—ranging from 26.9% in the 493-patient cohort of Zhou et al. (2022)[18] to 30%–50% across the meta-analyses and cohort studies of Cousin et al. (2022),[19] Zharkimbayeva et al. (2024),[20] Hild et al. (2024),[21] Scougall et al. (2023),[22] and Gabra et al. (2024).[23] However, our inclusion criteria specifically evaluated intact reconstructions without active fistula at three months. When a proximal or midshaft defect undergoes complete chordee release (orthoplasty), tailored urethral plate tubularization or onlay/staged grafting, well-vascularized dartos barrier coverage, and anatomical glans wing approximation,[37–41] the resulting external configuration of the glans, meatus, shaft skin, and penile axis can closely approach that of a distal repair. This interpretation is strongly corroborated by Neheman et al. (2024), who found that while meatal location correlated with HOPE outcome on univariable screening, it ceased to be an independent predictor in multivariable analysis once chordee severity (p = 0.013), glans width (p = 0.003), GMS score (p = 0.007), and preoperative HOPE score (p = 0.002) were accounted for.[11] At the same time, we emphasize that our non-significant p-value of 0.315 reflects an observational comparison with limited statistical power in the midshaft (n = 6) and distal (n = 15) groups, where distal repairs still showed a +2.64-point higher mean (49.67 vs. 47.03) and a +2.00-point higher median (51.00 vs. 49.00) than proximal repairs.

The central clinical finding of this study—that delayed surgery beyond 18 months (mean cohort age 93.28 ± 53.09

months; 84.9% >18 months) was associated with a 7.75-point reduction in mean HOPE score (46.62 ± 6.04 vs. 54.38 ± 3.16; p < 0.001) and a significantly higher proportion of suboptimal cosmetic outcomes (97.8% vs. 62.5%; Fisher's exact p = 0.009; RR = 1.564; OR = 26.400; Firth OR = 18.06), reinforced by an exploratory ROC cutoff at 13.5 months (AUC = 0.865, p = 0.016; 97.8% vs. 57.1%, Fisher's exact p = 0.006; RR = 1.712; OR = 33.750; Firth OR = 22.97)—is supported by converging cellular, microvascular, and biomechanical mechanisms (Figure 6). First, at the molecular level, Abdelhalim et al. (2025) demonstrated in a prospective histological study of primary TIP urethroplasty that diminished expression of epidermal growth factor (EGF) and transforming growth factor-beta (TGF-β) within the parameatal urethral plate correlates significantly with impaired epithelialization, defective extracellular matrix remodeling, and postoperative complications.[15] Complementing this, Tsachouridis et al. (2025) revealed through microscopic and molecular profiling that the urethral plate in hypospadias possesses structurally distinct microvascular networks compared with normal urethral spongiosum.[16] During infancy (≤13.5 to ≤18 months), higher capillary density, preserved angiogenic signaling, and compliant dartos connective tissue facilitate tension-free tubularization and rapid dermal remodeling with minimal hypertrophic scarring.[8,15,16] In older children and adolescents (>18 months), progressive perimeatal fibrosis, stiffer tunica albuginea and dartos collagen matrices, more frequent and forceful nocturnal erections generating cyclic mechanical shear across the glansplasty and skin closure, and higher perineal bacterial colonization collectively impair cosmetic healing.[8,30,44,45] Second, from a biostatistical standpoint, while Fisher's exact tests confirmed highly significant associations at both 18 months (p = 0.009) and 13.5 months (p = 0.006) and the OR confidence intervals excluded unity (2.290–304.376 and 2.817–404.423), the asymptotic Wald 95% CIs for the suboptimal relative risks (0.913–2.681 for RR = 1.564 and 0.900–3.256 for RR = 1.712) slightly crossed 1.0 due to the small denominators in the early-surgery reference groups (n = 8 and n = 7) and the high baseline incidence of HOPE < 57. Similarly, because the 13.5-month threshold was derived and tested within the same 53-patient sample containing only 4 optimal events (HOPE ≥ 57), it is subject to in-sample optimism and must be viewed strictly as a hypothesis-generating exploratory finding that complements—rather than replaces—the established 6–18-month EAU/AAP guideline window.[7,8]

A deeper examination of the pathobiological and biomechanical cascade illustrated in Figure 6 clarifies why surgical timing exerted such a pronounced effect on three-month HOPE scores across our cohort. Cutaneous and urethral wound healing after hypospadias reconstruction depends on a tightly regulated sequence of hemostasis, neutrophilic and macrophagic inflammation, angiogenic and fibroproliferative tissue formation, and long-term extracellular matrix remodeling.[13,15,16] In their landmark 2025 prospective clinico-histological study, Abdelhalim et al. quantified growth factor expression within parameatal urethral plate biopsies obtained during primary TIP repair and demonstrated that reduced immunohistochemical expression of epidermal growth factor (EGF) and transforming growth factor-beta (TGF-β) was strongly associated with impaired neourethral healing and postoperative complications.[15] EGF stimulates mitogenic migration and stratification of urothelial and keratinocyte progenitor cells across the incised urethral plate and ventral glans closure, while TGF-β orchestrates fibroblast chemotaxis, angiogenesis, and balanced deposition of type III and type I collagen fibers.[15] Concurrently, Tsachouridis et al. (2025) demonstrated via high-resolution microscopic and molecular profiling that the urethral plate and abortive corpus spongiosum in hypospadias exhibit distinct microvascular architecture and extracellular matrix organization compared with normal urethral tissues.[16]

During the infantile surgical window (≤13.5 to ≤18 months), penile tissues benefit from three synergistic biological advantages: first, higher microvascular density and preserved local growth factor responsiveness within the urethral plate and dartos fascia, supporting rapid microvascular inosculation of transposed flaps; second, a compliant extracellular matrix enriched in elastic type III collagen with minimal pre-existing perimeatal fibrosis, allowing glans wings and shaft skin to be approximated without mechanical tension; and third, low-amplitude, infrequent erectile tumescence and lower colonization by mature perineal skin flora.[8,15,16] Conversely, when surgery is delayed into late childhood or adolescence (mean age 93.28 months in our cohort), chronic exposure of the dysplastic ventral urethral plate and para-meatal mucosa leads to progressive subepithelial fibrosis and diminished vascular plasticity.[16,30] Furthermore, somatic growth of the corpora cavernosa increases penile diameter and erectile rigidity; according to LaPlace's law, larger corporeal circumference and forceful nocturnal erections in older boys generate substantially greater circumferential tensile stress across the ventral glansplasty and skin suture lines, predisposing to micro-dehiscence, widened coronal scars, asymmetric skin dog-ears, and lower HOPE subscores.[8,30,44]

From a biostatistical perspective, our dual reporting of relative risks (RR) and odds ratios (OR)—including Firth's penalized likelihood estimates—and our full 21-coordinate ROC analysis provide transparent insight into the precision limits of this 53-patient dataset. In 2 × 2 contingency tables where the outcome of interest (suboptimal cosmetic score, HOPE < 57) occurs in 62.5% (5/8) of the ≤18 months reference group and 97.8% (44/45) of the >18 months group, the ratio of proportions is mathematically constrained by the ceiling of 100% (maximum possible RR = 100% / 62.5% = 1.600). Because the observed RR of 1.564 (and RR = 1.712 at the 13.5-month cutoff) lies very close to this mathematical ceiling, yet the reference denominator contains only n = 8 (or n = 7) patients, the symmetric Wald log-binomial 95% confidence interval (0.913–2.681 for >18 months; 0.900–3.256 for >13.5 months) slightly overlaps 1.000 even though exact hypergeometric inference confirms a statistically significant association (two-sided Fisher's exact p = 0.009 and p = 0.006, respectively). When the same contingency table is parameterized using the rare optimal outcome (HOPE ≥ 57: 2.2% vs. 37.5% at 18 months; 2.2% vs. 42.9% at 13.5 months), the relative risk (RR = 0.059, 95% CI: 0.007–0.501; and RR = 0.051, 95% CI: 0.006–0.422), the crude odds ratio (OR = 26.400, 95% CI: 2.290–304.376; and OR = 33.750, 95% CI: 2.817–404.423), and the Firth bias-corrected odds ratio (OR = 18.06, 95% CI: 2.14–231.80; and OR = 22.97, 95% CI: 2.62–299.40) all exclude the null value of 1.000. Likewise, in the ROC coordinate analysis (AUC = 0.865 ± 0.094, p = 0.016), the identification of 13.5 months as the Youden peak (J = 0.668 vs. J = 0.648 at the 18.0-month / 19.5-month coordinate) reflects the empirical clustering of 3 of the 4 optimal cases at ≤12 months of age, whereas the jump to 100.0% specificity at ≥102.0 months reflects the single optimal case operated on at 96 months. Consequently, the 13.5-month cutoff should be interpreted as an exploratory internal benchmark that reinforces the clinical urgency of early infantile referral within the established 6–18-month EAU/AAP window.

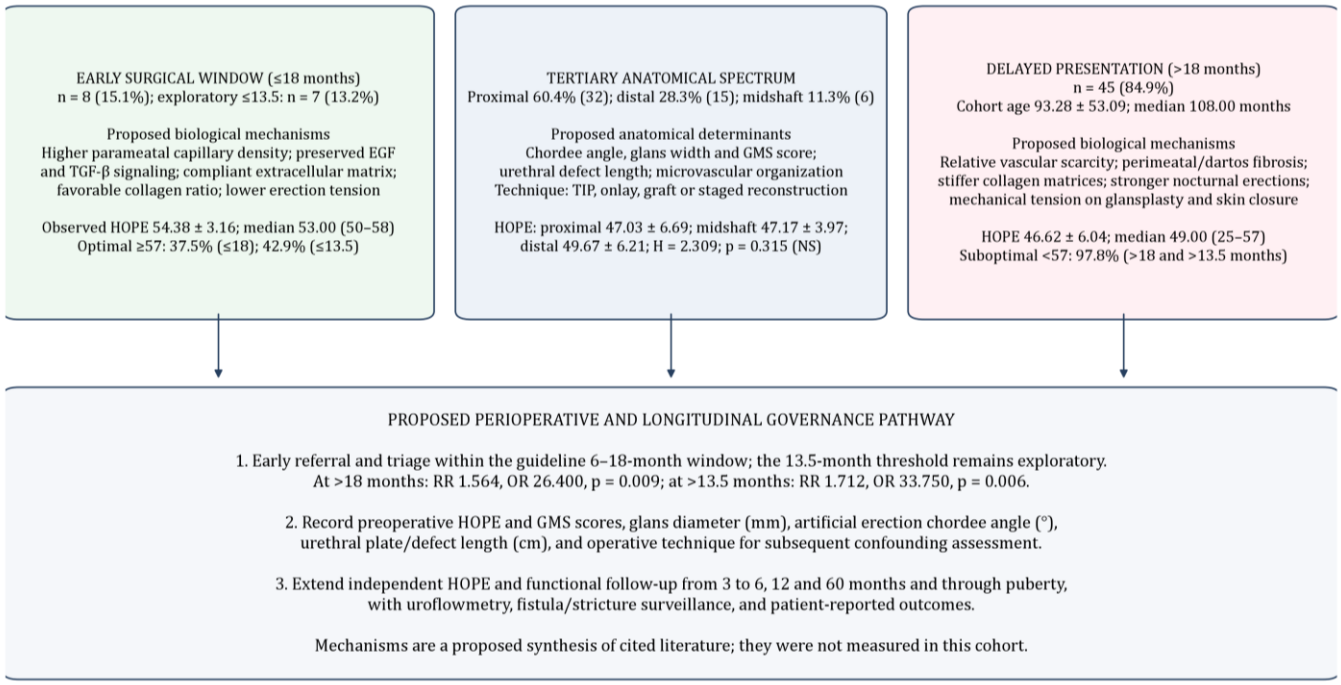

Figure 6. Integrated pathobiological wound-remodeling cascade and clinical governance pathway for pediatric hypospadias reconstruction: synthesizing age-dependent parameatal urethral plate growth factor expression (EGF and TGF-β), microvascular density, extracellular matrix compliance, and anatomical chordee/glans determinants with three-month and longitudinal HOPE surveillance protocols at RSUP Dr. Mohammad Hoesin Palembang.

Several limitations must be considered when interpreting these findings. First, the observational ambispective design at a single tertiary referral center and the absence of multivariable regression adjustment (precluded mathematically by the presence of only 4 optimal outcome events, yielding an events-per-variable ratio below the minimum threshold required for stable logistic modeling) mean that residual confounding by preoperative chordee angle, glans diameter, GMS score, urethral plate width, specific urethroplasty technique (single-stage vs. two-stage), prior operative history, and surgeon learning curve cannot be excluded.[11,14,18,42] Second, the pronounced subgroup imbalance between patients operated on at >18 months (n = 45) and ≤18 months (n = 8) resulted in wide confidence intervals around the odds ratios and ROC AUC (0.681–1.000). Third, cosmetic evaluation was performed at a single three-month postoperative time point without concomitant objective functional metrics such as uroflowmetry, post-void residual volume, or long-term surveillance for late urethral stricture, recurrent curvature, and adolescent psychosexual satisfaction.[30–35] Future research should therefore employ prospective multicenter cohorts with balanced age strata, strictly blinded multi-rater photographic panels, standardized preoperative GMS and HOPE baseline scoring, serial follow-up at 3, 6, 12, and 60 months, and multivariable modeling to externally validate the 13.5-month and 18.0-month thresholds.

To translate these findings into actionable clinical governance at Dr. Mohammad Hoesin Central General Hospital and comparable regional referral centers across Indonesia, a three-pillar clinical pathway is proposed (Figure 6). First, pediatricians, general practitioners, and midwives across South Sumatra's primary and secondary healthcare network should be trained to screen male newborns for ventral preputial hooding, chordee, and ectopic meatal location at birth and initiate specialist referral within the first 3 to 6 months of life, enabling definitive surgical planning within the 6–18-month window (with the 13.5-month cutoff treated as exploratory). Second, institutional hypospadias registries should mandate standardized preoperative phenotyping—recording baseline HOPE score, GMS score, glans width in millimeters, artificial erection chordee angle in degrees, urethral plate width and post-chordectomy defect length, preputial/dartos vascular status, and exact suture/flap technique—so that future multicenter studies achieve sufficient events-per-variable to execute multivariable logistic and linear regression modeling.[11,14,18,41–43] Third, postoperative follow-up should be extended longitudinally from 3 months to 6, 12, and 60 months and into post-pubertal transition, combining blinded multi-rater photographic HOPE scoring with objective uroflowmetry, fistula/stricture surveillance, and patient-reported psychosexual outcome instruments.[30–35]

5. Conclusion

In this ambispective cohort of 53 patients undergoing hypospadias reconstruction at Dr. Mohammad Hoesin Central General Hospital, Palembang, proximal hypospadias was the predominant anatomical subtype (60.4%), followed by distal (28.3%) and midshaft (11.3%) defects, and 84.9% of patients underwent delayed surgery after 18 months of age (mean 93.28 ± 53.09 months; median 108.00 months). At three months postoperatively, the overall mean HOPE score was 47.79 ± 6.33 points (median 49.00, range 25.00–58.00), with 7.5% (4/53) achieving an optimal cosmetic score (HOPE ≥ 57) and 92.5% (49/53) classified as suboptimal (HOPE < 57); postoperative HOPE scores did not differ significantly across anatomical hypospadias subtypes (Kruskal–Wallis H = 2.309, p = 0.315). Patients operated on within the guideline window of ≤18 months achieved significantly higher continuous HOPE scores than those operated on at >18 months (54.38 ± 3.16 vs. 46.62 ± 6.04; p < 0.001) and had a significantly lower proportion of suboptimal cosmetic outcomes (62.5% vs. 97.8%; Fisher's exact p = 0.009; RR = 1.564, 95% CI: 0.913–2.681; OR = 26.400). Exploratory ROC analysis demonstrated good discrimination for surgical age (AUC = 0.865, 95% CI: 0.681–1.000; p = 0.016) with

an optimal empirical cutoff of 13.5 months (suboptimal rate 97.8% vs. 57.1%; Fisher's exact p = 0.006; RR = 1.712, 95% CI: 0.900–3.256; OR = 33.750). Strengthening regional pediatric referral networks to facilitate timely hypospadias repair within the recommended 6–18-month developmental window, alongside routine pre- and postoperative HOPE score documentation and long-term longitudinal follow-up, is recommended to optimize reconstructive outcomes. By integrating early regional referral within the 6–18-month window, standardized preoperative anatomical phenotyping (GMS and baseline HOPE scoring), meticulous tissue-preserving reconstructive technique, and blinded longitudinal cosmetic and functional surveillance, tertiary surgical centers can substantially improve both short-term and lifelong reconstructive outcomes for children born with hypospadias.

6. Declarations

Ethics approval and consent to participate

The study protocol received formal ethical clearance from the Health Research Ethics Committee of Dr. Mohammad Hoesin Central General Hospital, Palembang, Indonesia (Approval number DP.04.03/D.XVII.9.4/ETIK/304/2026), prior to data collection and was conducted in full compliance with the ethical principles of the Declaration of Helsinki. Written informed consent was obtained from the parents or legal guardians of all participating patients prior to clinical examination and standardized photographic documentation.

Consent for publication

Written informed consent was obtained from parents or legal guardians; no identifiable individual patient photographs or personal identifiers are published in this article.

Availability of data and materials

All statistical summaries, contingency distributions, and receiver operating characteristic coordinates generated or analyzed during this study are included in the published tables and figures; de-identified analytical records are available from the corresponding author upon reasonable request.

Artificial intelligence use statement

The authors declare no generative AI were used in preparation of this manuscript.

Competing interests

The authors declare that they have no competing financial or non-financial interests.

Funding

This research received no external grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions (CRediT)

YT: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft, and Writing – review & editing. IP: Conceptualization, Methodology, Supervision, Validation, and Writing – review & editing. DHH: Methodology, Supervision, Validation, and Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge the clinical and administrative staff of the Division of Plastic, Reconstructive, and Aesthetic Surgery, the Pediatric Urology Service, the Department of Surgery, and the Medical Records Installation of Dr. Mohammad Hoesin Central General Hospital, Palembang, for their support during patient follow-up and documentation.

7. References

1. Halaseh SA, Halaseh S, Ashour M. Hypospadias: A Comprehensive Review Including Its Embryology, Etiology and Surgical Techniques. Cureus. 2022;14(7):e27544. doi: 10.7759/cureus.27544
2. Shan X, Aguilar S, Canon S. Hypospadias prevalence by severity and associated risk factors: A population-based active surveillance analysis. J Pediatr Urol. 2023;19(6):720.e1-720.e9. doi: 10.1016/j.jpurol.2023.08.010
3. Lavoie C, Chun B, Au M, et al. Comparing the incidence of hypospadias across the United States: A contemporary analysis. J Pediatr Urol. 2025;21(3):627-632. doi: 10.1016/j.jpurol.2025.01.002
4. Indriasari V, Diposarosa R, Syukriani YF, et al. Nongenetic Risk Factors of Severe Hypospadias: A Case–Control Study. J Indian Assoc Pediatr Surg. 2024;29(5):488-491. doi: 10.4103/jiaps.jiaps_31_24
5. Ajayi AF, Oyowvi MO, Akanbi GB, et al. Genetic Susceptibility and Pathogenesis of Hypospadias. Biochem Genet. 2026;64(3):3047-3075. doi: 10.1007/s10528-025-11199-9
6. Peng X, Zu J, Wang S, et al. Novel NR5A1 variants associated with hypospadias and disorders of sex development: A series case report of 4 patients. Medicine (Baltimore). 2025;104(50):e46626. doi: 10.1097/MD.0000000000046626
7. Ceccarelli PL, Lucaccioni L, Poluzzi F, et al. Hypospadias: clinical approach, surgical technique and long-term outcome. BMC Pediatr. 2021;21(1):523. doi: 10.1186/s12887-021-02941-4
8. Dale J, Woodward B, Elagami H. Age-dependent early complications of hypospadias repair: a single institutional experience. Pediatr Surg Int. 2023;39(1):115. doi: 10.1007/s00383-023-05388-z
9. Bhatia V, Fernandez N, Long C, et al. Advancements in hypospadias management: trends, techniques, training, and patient-centric outcomes. Urol Res Pract. 2024;50(2):94-101. doi: 10.5152/tud.2024.23219
10. van der Toorn F, de Jong TPVM, de Gier RPE, et al. Introducing the HOPE (Hypospadias Objective Penile Evaluation)-score: A validation study of an objective scoring system for evaluating cosmetic appearance in hypospadias patients. J Pediatr Urol. 2013;9(6):1006-1016. doi: 10.1016/j.jpurol.2013.01.015
11. Neheman A, Schwarztuch Gildor O, Shumaker A, et al. Use of Validated Questionnaires to Predict Cosmetic Outcomes of Hypospadias Repair. Children (Basel). 2024;11(2):189. doi: 10.3390/children11020189
12. Kim KJ, Kim JK, Chua ME, et al. Expert classification of hypospadias: an external validation and evaluation of agreement for Glans–Urethral Meatus–Shaft (GMS) and Hypospadias Objective Penile Evaluation (HOPE) scores. Pediatr Surg Int. 2024;40(1):233. doi: 10.1007/s00383-024-05816-8
13. Lawo K, Feller D, Klein BS, et al. A prospective randomized clinical trial to evaluate wound healing, cosmetic and functional results as well as postoperative adverse events comparing two types of dressing technique after hypospadias repair. Pediatr Surg Int. 2025;41(1):176. doi: 10.1007/s00383-025-06063-1
14. Abbas T. PREDICT-H Protocol: A Multicenter Prospective Cohort Study on Preoperative Anatomical Determinants and Postoperative Complications in Primary Hypospadias Repair. Diagnostics (Basel). 2025;15(16):2087. doi: 10.3390/diagnostics15162087
15. Abdelhalim K, Wasfy I, Zalata K, et al. Relation between outcome of primary hypospadias tubularized incised plate repair and penile growth factors expression. Afr J Urol. 2025;31(1):56. doi: 10.1186/s12301-025-00535-5
16. Tsachouridis G, Liu X, de Bakker B, et al. Microscopic and Molecular Analysis of the Urethral Plate: A Step Towards Improved Hypospadias Surgery Outcomes. Urology. 2025;200:142-150. doi: 10.1016/j.urology.2025.01.062
17. Henrizal F, Daryanto B, Nurhadi P. Clinical profile of patients with hypospadias in Saiful Anwar General Hospital, Malang. Indones J Urol. 2025;32(2):112-118. doi: 10.32421/juri.v32i2.1051
18. Zhou G, Wang R, Zhu W, et al. Risk factors for postoperative complications in children with proximal hypospadias with severe

chordee who underwent urethral plate transection. Int J Urol. 2022;29(11):1310-1314. doi: 10.1111/iju.14986

19. Cousin I, Basmaison C, Cousin E, et al. Complication rates of proximal hypospadias: meta-analyses of four surgical repairs. J Pediatr Urol. 2022;18(5):587-597. doi: 10.1016/j.jpurol.2022.08.005
20. Zharkimbayeva A, Aubakirov M, Lozovoy V, et al. Comparative Analysis of Surgical Methods for Distal, Mid-, and Proximal Shaft Hypospadias in Young Males: A Prospective Study on Postoperative Outcomes. Medicina (Kaunas). 2024;60(11):1903. doi: 10.3390/medicina60111903
21. Hild O, Fotso Kamdem A, Boulard N, et al. Primary hypospadias repair outcomes: results from a retrospective cohort of 292 children. World J Urol. 2024;42(1):137. doi: 10.1007/s00345-024-04811-x
22. Scougall K, Bryce J, Baronio F, et al. Predictors of surgical complications in boys with hypospadias: data from an international registry. World J Pediatr Surg. 2023;6(4):e000599. doi: 10.1136/wjps-2023-000599
23. Gabra A, Beyari BM, AlNuwaiser SJ, et al. Outcomes of Hypospadias Repair Based on Surgical Techniques: A 4-Year Retrospective Study. Res Rep Urol. 2024;16:79-87. doi: 10.2147/RRU.S451552
24. Krull S, Krause H, Mohnike K, et al. Outcome after Hypospadias Repair: Evaluation Using the Hypospadias Objective Penile Evaluation Score. Eur J Pediatr Surg. 2018;28(3):268-272. doi: 10.1055/s-0037-1602252
25. Yusof MR, Khairul-Asri MG, Ahmad Faiz NMG, et al. Hypospadias Repair: Outcome and Objective Evaluation with Hypospadias Objective Scoring Evaluation (HOSE), Hypospadias Objective Penile Evaluation (HOPE) and Pediatric Penile Perception Score (PPPS) from Parents Perspective and Preference in COVID-19 Era. Androl Genit Surg. 2024;25(1):65-72. doi: 10.62968/2070-9781-2024-25-1-65-72
26. Manzoor A, Talat N, Adnan HM, et al. Post-Hypospadias Repair Penile Score in Follow-Up Patients of Urethroplasty. Cureus. 2022;14(4):e23816. doi: 10.7759/cureus.23816
27. Patil JD, Mohamed YM, Farhan A, et al. Outcomes Assessment of Hypospadias Repair. Cureus. 2023;15(11):e48808. doi: 10.7759/cureus.48808
28. Wahyudi I, Utomo CP, Djauzi S, et al. Digital Pattern Recognition for the Identification of Various Hypospadias Parameters via an Artificial Neural Network: Protocol for the Development and Validation of a System and Mobile App. JMIR Res Protoc. 2022;11(11):e42853. doi: 10.2196/42853
29. Fernandez N, Lorenzo AJ, Rickard M, et al. Digital Pattern Recognition for the Identification and Classification of Hypospadias Using Artificial Intelligence vs Experienced Pediatric Urologist. Urology. 2021;147:264-269. doi: 10.1016/j.urology.2020.09.019
30. Ye ZH, Wang C, Zhang ZC, et al. Risk factors and timing of complication presentation following primary hypospadias repair in adolescents. Int J Urol. 2024;31(3):245-251. doi: 10.1111/iju.15342
31. Leunbach TL, Skott M, Ernst A, et al. Referral patterns, clinical features and management of uncorrected hypospadias in a series of adult men. J Pediatr Urol. 2022;18(4):480.e1-480.e7. doi: 10.1016/j.jpurol.2022.06.008
32. Skott M, Leunbach TL, Hvistendahl GM, et al. Management of complications after hypospadias repair during childhood. Dan Med J. 2025;72(4):A08240506. doi: 10.61409/a08240506
33. Stojanovic B, Bizic M, Bencic M, et al. Adult non-urethral complications after hypospadias repair in childhood: presentation, treatment and outcomes. Front Endocrinol (Lausanne). 2023;14:1184948. doi: 10.3389/fendo.2023.1184948
34. Gaines T, Simhan J. Adult Hypospadias Outcomes for the Pediatric Urologist. Curr Urol Rep. 2024;25(4):63-70. doi: 10.1007/s11934-024-01196-7
35. Gul M, Hildorf S, Silay MS. Sexual functions and fertility outcomes after hypospadias repair. Int J Impot Res. 2021;33(2):149-163. doi: 10.1038/s41443-020-00377-5
36. Gama M, Abitew B, Abebe K. Clinical profiles and surgical outcome of hypospadias repair at a teaching hospital in Ethiopia. Ethiop J Health Sci. 2022;32(3):613-622. doi: 10.4314/ejhs.v32i3.18
37. Herzberg H, Dubi-Sobol A, Mendelson T, et al. Operative techniques and long-term outcomes of hypospadias repair in the absence of preputial skin after neonatal circumcision. J Pediatr Surg. 2022;57(11):676-680. doi: 10.1016/j.jpedsurg.2022.06.021
38. Nerli RB, Chandra S, Rai S, et al. Grafted Tubularised Incised Plate: A Right Option in the Management of Failed Mid-Penile and Distal Hypospadias. Afr J Paediatr Surg. 2023;20(3):197-201. doi: 10.4103/ajps.ajps_19_22
39. Kılıç S, Kaba M, Gecit İ. Comparative study in distal hypospadias repair: a meatal-based flaps technic vs. tubularized incised plate urethroplasty. BMC Urol. 2024;24(1):249. doi: 10.1186/s12894-024-01603-8
40. El-Helaly HAA, Youssof HA, Ibrahim HM, et al. Distal hypospadias repair: Comparative study between snodgrass and transverse preputial onlay flap. J Pediatr Urol. 2022;18(5):610.e1-610.e6. doi: 10.1016/j.jpurol.2022.08.016
41. Subihardi LME, Yatindra IBGTY, Rahman LA, et al. A thorough analysis of the effects and complications of two different suturing techniques in hypospadias repair using tubularized-incised plate urethroplasty: a meta-analysis. Med Arch. 2023;77(3):194-201. doi: 10.5455/medarh.2023.77.194-201
42. Hisamatsu E, Sugita Y, Haruna A, et al. The learning curve in proximal hypospadias repair. J Pediatr Urol. 2021;17(3):330.e1-330.e6. doi: 10.1016/j.jpurol.2021.01.005
43. Abbas T, Tiryaki S, Tekin A, et al. Hypospadias Reconstruction Training: Development of an Ex-Vivo Model for Objective Evaluation of Surgical Skills. Urology. 2025;195:112-119. doi: 10.1016/j.urology.2024.09.013
44. Nguyen S, Durbin-Johnson B, Kurzrock EA. Reoperation after Hypospadias Repair: Long-Term Analysis. J Urol. 2021;205(6):1778-1784. doi: 10.1097/JU.0000000000001569
45. Gozar H, Bara Z, Dicu E, et al. Current perspectives in hypospadias research: A scoping review of articles published in 2021 (Review). Exp Ther Med. 2023;25(5):211. doi: 10.3892/etm.2023.11910