Validation and Preliminary Results of the Parental Assessment of Children's External Genitalia Scale for Females (PACE-F) for Girls With Congenital Adrenal Hyperplasia

Konrad M. Szymanski, Benjamin Whittam, Patrick O. Monahan, Martin Kaefer, Heather Frady, Mark P. Cain, and Richard C. Rink

OBJECTIVE To validate a parental assessment of children's external genitalia scale for females (PACE-F) for girls with congenital adrenal hyperplasia (CAH) by adapting the validated adult female genital self-image scale.

METHODS PACE-F was administered to parents of girls (Tanner 1, 2 months-12 years) with and without CAH. Final questions were determined by clinical relevance and psychometric properties (scores: 0-100). A reference range was established using 95% confidence interval among controls. Age-matched controls were compared to girls with CAH (1) <4 years old before and after female genital reconstruction surgery (FGRS), and (2) 4-12-year olds after FGRS. Nonparametric statistics were used.

RESULTS Participants included 56 parents of 41 girls with CAH (median 3.9 years old, 97.6% FGRS) and 139 parents of 130 girls without CAH. Face and content validity was established by families, experts, and factor analysis. Internal consistency was high (Cronbach's alpha: 0.83). Population reference score range was 66.7-100. Ten consecutive girls had pre- and post-FGRS PACE-F scores. All scores improved at 4 months after surgery and all preoperative scores were below reference range and lower than controls ($P = .0001$). All postoperative scores were within reference range, no different from controls ($P = .18$). Scores for girls with CAH after FGRS aged 4-12 years were no different from controls (100.0 vs 88.9, $P = .77$) and 90.0% were in reference range, as expected ($P = .99$).

CONCLUSION We present a validated instrument for parental assessment of genital appearance in girls with CAH. We demonstrate improved parent-reported appearance after FGRS, with scores similar to age-matched controls. UROLOGY 130: 132–137, 2019. © 2019 Elsevier Inc.

Congenital adrenal hyperplasia (CAH) is the most common etiology of ambiguous genitalia in newborns.[1-3] The appropriateness and timing of female genital restoration surgery (FGRS) in females with CAH is an area of debate.[4,5] Adding to ethical considerations are concerns that outcomes may differ for surgeries performed in infancy vs puberty.[6,7] Limited data support either approach.[8,9] Since physician interpretations of patient and parental views are prone to bias, under-reporting and minimization,[10,11] validated methods of obtaining patient/parent-reported outcomes (PROs) are necessary.

While patient's perceptions are essential, parents currently make the early management decisions. Improved parental perception after FGRS has never been reliably documented in the literature. Moreover, without a validated tool for such an assessment in girls with and without genital ambiguity, reliable studies of relationships between anatomy, parental perceptions and, most importantly, perceptions by females themselves cannot be performed. The 4-item female genital self-image scale (FGSIS) was validated in a large national sample of adult women without known genital ambiguity, measuring women's perceptions about their genital appearance.[12] Lack of an age-appropriate parental FGSIS version precludes its use in children with genital ambiguity. We aimed (1) to validate a parental

Declaration of Interest: *None.*
Funding: *Departmental research fund.*
Internal review board approval: 1512039731.

From the Division of Pediatric Urology, Riley Hospital for Children at Indiana University Health, Indianapolis, IN; and the Department of Biostatistics, Indiana University School of Medicine and School of Public Health, Indianapolis, IN

Address correspondence to: Konrad M. Szymanski, M.D., M.P.H., Division of Pediatric Urology, Riley Hospital for Children at IU Health, 705 Riley Hospital Dr., Suite 4230, Indianapolis, IN 46202. E-mail: szymanko@iupui.edu

Submitted: January 31, 2019, accepted (with revisions): April 29, 2019

https://doi.org/10.1016/j.urology.2019.04.034
0090-4295

assessment of children's external genitalia scale for females (PACE-F) for girls with CAH by adapting the FGSIS, and (2) to report preliminary results of the PACE-F in girls with CAH. We hypothesized that PACE-F scores are lower for girls with CAH without FGRS than girls after FGRS and girls without CAH.

METHODS

We performed an IRB-approved study of parents of girls (Tanner 1, 2 months-12 years old) with CAH at pediatric urology clinics and without CAH (controls) at pediatrician clinics and among hospital staff and colleagues (2016-2018). Girls with CAH were followed at our institution before and after FGRS, all performed by age 3 largely by a single surgeon (2005-2018). Parents of 105 children with CAH with contact information were eligible. Exclusion criteria for control participants were: girls having diagnoses suggestive of abnormal genitalia/perineum or conditions predisposing to frequent genital examinations (genital concerns, history of urinary tract infections, hematuria, clean intermittent catheterizations, neurogenic bladder, and imperforate anus). Controls presenting for routine checkup appointments received a $5 gift card.

Instrument Development

We used a stepwise approach to develop a brief, clinically useful instrument capturing parental perceptions of genital appearance of girls with CAH. First, the 4-item FGSIS was reworded to be age-appropriate and reflect parental, rather than patient, responses. It was revised after piloting with 4 parents and 6 clinicians. Responses were made on a 4-point Likert scale (strongly disagree, disagree, agree, and strongly agree), like the original FGSIS instrument. Readability was assessed by the Flesch-Kincaid Grade Level test.[13]

Item Reduction

Item reduction is commonly performed to arrive at shorter instruments without compromising utility. Each of the 4 reworded items was examined for medical plausibility, clinical relevance, and parental comfort ("I am comfortable answering this question": strongly disagree, disagree, agree, strongly agree; not validated). We performed a principal axes method of factor analysis and inspection of the scree plot inflexion point.[14,15] Since at least 3 items with loadings >0.4 would be required for a functional instrument,[14,15] 1 item could and was dropped. Factor analysis was used to assess whether the original 4 items, or a reduced set of 3, reasonably satisfied unidimensionality. Factor analysis of 4 items would require 20 participants (at least 5 participants/item) for adequate statistical estimation.[16,17]

Validation

Since PACE-F is a parent-reported measure, it was more appropriate to treat each parental response, rather than an individual child, as the unit of analysis during the validation. We assessed several types of validity to determine if PACE-F measures what it was intended to measure.[18] Face and content validity were established through a review by parents and experts. Construct validity was assessed through (1) factor analysis in the CAH group and (2) comparing PACE-F scores between children with and without CAH, regardless of history of FGRS. We calculated the domain means, standard deviations (SDs), medians, ranges, and percentages of subjects scoring the minimum (floor) and maximum (ceiling).

Reliability (reproducibility) was assessed by measuring internal consistency with Cronbach's alpha (0.7-0.9 signifies robust consistency without redundancy). Cronbach's alpha for the 4-item FGSIS is 0.86.[12]

Convergent validity evaluates the degree to which PACE-F scores converge with other instruments measuring similar outcomes or a gold standard. No instrument similar to PACE-F exists, thus we compared PACE-F scores between parents expressing different levels of comfort with changing diapers in a public bathroom for children <4 years old ("I am comfortable changing my child's diapers in a public bathroom or change room": strongly disagree, disagree, agree, strongly agree; not validated) and, when available, surgeon-assessed pre-FGRS level of virilization (Prader scale).

Clinical Utility

To determine a reference range of PACE-F values for future clinical use, we used the one-sided (lower) 95% confidence interval of scores from parents for girls without CAH, a method previously used by the World Health Organization.[19] Two distribution-based approaches were used to determine the minimal clinically important difference (MID) in PACE-F scores that could be appreciated by individuals. Differences were determined using 0.5 SD[20] and 1.0 standard error of measurement (SEM), where SEM was defined as $SD^*\sqrt{(1-\text{Cronbach's alpha})}$.[21,22] We selected the most conservative, largest point difference calculated by these methods as the MID.

Differences Between Parents

Maternal and paternal PACE-F scores were compared to determine if parents provided similar scores for their daughter.

Age-matched Comparisons

Since clinically meaningful analyses apply to individual patients, each child was the unit of analysis for each of the 3 comparisons between girls with CAH and age-matched controls. Because maternal and paternal PACE-F scores did not differ significantly (see Results section), a mean of the two scores was used when both were provided. Findings were similar when each analysis was repeated with parental scores as the unit of analysis (not shown).

1) The first comparison included only girls with prospectively collected pre- and post-FGRS PACE-F scores. In order to capture all post-FGRS PACE-F scores for patients whose surgery occurred by the age of 3, this analysis was restricted to girls younger than 4. An association between post-FGRS changes in PACE-F scores and changes in levels of comfort changing diapers in a public bathroom was tested to assess sensitivity, or responsiveness, of PACE-F, another form of validity.
2) The second analysis consisted of single-time point scores for girls <4 years old with only a pre- or post-FGRS assessment.
3) The third comparison included single-time point PACE-F scores of 4-12 year olds with CAH (all after FGRS) and controls.

Sensitivity to Change Validity

To quantify effect size observed in the prospective analysis, we calculated a standardized response mean by dividing the mean change in scores by the SD of the change in scores.[23,24] Effect

size can be categorized as small (0.2-0.5), moderate (0.5-0.8), and large (>0.8). Standardized response mean can be helpful in future power calculations.

Statistics

Study data were managed using REDCap.[25] Due to skewed score distributions, nonparametric statistics were used (Fisher's exact test for categorical variables; Wilcoxon rank sum and Kruskal-Wallis rank tests for cross-sectional comparisons of continuous data; Wilcoxon matched-pairs signed-rank test to compare scores before/after surgery). Factor analysis was used, as described above. Linear regression was used to test if PACE-F scores varied with age among controls. A critical $P = .05$ was used (software: Stata, StataCorp, College Station, TX).

RESULTS

Overall, 56 parental questionnaires were completed for 41 girls with CAH (38.7% response rate). Nineteen questionnaires were for 14 girls <4 years old without FGRS and 37 were after FGRS (17 for girls <4 years old, 20 for girls 4-12 year). Ten girls <4 years old had pre- and post-FGRS scores collected prospectively (23 parental responses, Analysis 1). Eleven girls <4 years old (13 parental responses) had either a pre-FGRS (4) or post-FGRS scores (7) (Analysis 2). Twenty girls 4-12 years had post-FGRS scores from 20 parents (Analysis 3). Fourteen of the last 15 (93.3%) consecutive families assessed at our institution participated.

Median age of 41 girls with CAH was 3.9 years old at the end of the study (95.1% salt-wasting 21-hydroxylase deficiency). Forty girls underwent FGRS at median age of 8 months (range: 2-37 months; 1 did not have surgery) with a median follow-up 2.9 years. Only 1 girl underwent surgery at another center. Overall, 87.5% of girls underwent vaginoplasty (71.4% posterior skin flap, 20.0% posterior sinus flap, 8.6% pull-through), 92.5% partial urogenital sinus mobilization (PUM)[26] and 95.0% clitoroplasty. Preoperative Prader scale was documented for 21 girls: Prader 3 in 47.6%, 4 in 47.6% and 5 in 4.8%.

Regarding controls, 139 parents of 130 girls without CAH participated. Parents in the CAH group were undistinguishable from parent controls in terms of age (36 vs 36 years, $P = .59$) and proportion being fathers (16.1% vs 23.9%, $P = .26$).

Construct Validity

Like the original 4-item FGSIS, all PACE-F items loaded significantly on one factor, accounting for 87.4% of item variance. Factor loadings were high for item 1 (satisfaction: 0.94), item 2 (comfort: 0.97) and item 4 (embarrassment: 0.49). Item 3 ("I think my child's genitals smell fine"), was not retained because it had a loading <0.4 (0.11), was least clinically relevant and least comfortable for parents to answer. The scree plot indicated a single dominant dimension, especially after deleting Item 3.

The final 3-item PACE-F has grade 7 readability, takes 1 minute to complete and is scored 0-100, where higher values signify higher satisfaction (Table 1). Median PACE-F score for the whole CAH groups was 66.7 (range 11.1-100.0), mean: 67.3 (SD 28.1; Table 2). A ceiling effect was observed, with 28.6% of responses being the maximal possible score of 100. Importantly, all scores for girls without FGRS were less than 100, indicating the ceiling effect applies only after FGRS (and to controls). Without considering FGRS history, median scores were lower in CAH group than controls ($P = .0001$).

Internal Reliability

The initial 4-item PACE-F had a Cronbach's alpha of 0.73 among parents of girls with CAH. After excluding the "smell item," the Cronbach's alpha for the 3-item PACE-F was 0.83, suggesting improved internal consistency without redundancy.

Table 1. Parental assessment of children's external genitalia scale for females (PACE-F)

Question	Strongly Disagree	Disagree	Agree	Strongly Agree
1. I am satisfied with the way my child's genitals look.				
2. I would feel comfortable letting my child's future sexual partner look at her genitals when she is older.				
3. I am not embarrassed about my child's genitals.				

Answers options are: strongly disagree (0.0), disagree (11.1), agree (22.2), strongly agree (33.3). Final score range is: 0-100.

Table 2. Domain characteristics of the parent-reported PACE-F scores in girls with and without CAH

Parents of Girls With CAH	*N*	Median Age (Years, Range)	Mean (SD)	Median (Range)	% Scoring Minimum	% Scoring Maximum
Overall	56	3.5 (6 months-12.0 years)	67.3 (28.1)	66.7 (11.1-100.0)	0.0	28.6
Without FGRS	19	7 months (6 months-3.1 years)	37.4 (114.9)	33.3 (11.1-77.8)	0.0	0.0
After FGRS (Less than 4 years)	17	14 months (11 months-3.9 years)	81.1 (13.5)	77.8 (66.7-100.0)	0.0	23.5
After FGRS (4-12 years)	20	8.5 (4.3-12.0)	83.9 (23.8)	100 (22.2-100.0)	0.0	60.0
Parents of Girls Without CAH	***N***	**Median Age (Range)**	**Mean (SD)**	**Median (Range)**	**% Scoring Minimum**	**% Scoring Maximum**
Overall[a]	139	3.5 (2 months-12.0)	89.1 (13.8)	100.0 (55.6-100.0)	0.0	53.2
Less than 4 years	68	1.3 (2 months-3.9)	91.2 (12.8)	100 (55.6-100.0)	0.0	58.8
4-12 years	71	6.1 (4.0-12.0)	86.8 (14.8)	88.9 (66.7-100.0)	0.0	48.4

CAH, Congenital Adrenal Hyperplasia; SD, standard deviation.
[a] No statistically significant differences in scores were noted between different age groups.

Not surprisingly, the 3-item PACE-F Cronbach's alpha for controls was 0.90, indicating that PACE-F items are redundant among parents of healthy girls.

Convergent Validity

Stronger parental comfort changing diapers in a public bathroom was associated with higher PACE-F scores of girls with CAH <4 years old ($P = 0.03$, Fig. 1). Among 13 girls with a documented pre-FGRS Prader grade and PACE-F scores, greater virilization was associated with slightly lower PACE-F scores, although this did not reach statistical significance. This was true for individual parental scores (8 parents Prader 3: 44.4 vs 11 Prader 4-5: 33.3, $P = .40$) and average scores per child if both parents participated (6 children Prader 3: 38.5 vs 7 Prader 4-5: 33.3, $P = .36$).

Population Reference Range of PACE-F Scores

Median score for controls was 100, although 46.8% of parents reported lower scores (Table 2). The 95% confidence interval of scores was 66.7-100.0, meaning that scores below 66.7 would be considered outside of the population range. PACE-F scores did not vary by child's age in the first 12 years of life on linear regression ($P = .15$) or between children younger than 4 and 4-12 years old ($P = .11$).

Minimally Important Difference

The minimally important difference in PACE-F scores that can be appreciated clinically was 14.1 (14.1 for 0.5 SD, 11.6 for SEM using Cronbach's alpha). Since there are 10 possible scores for the 3-item PACE-F, each being 11.1 points apart, a difference of at least 22.2, but not 11.1 would be perceivable to an individual.

PACE-F Scores Reported by Mothers and Fathers

Only 5 mother-father pairs provided individual pre-FGRS PACE-F scores, limiting comparison of the two parents. Pairing these scores, mothers and fathers reported undistinguishable scores (median: 33.3 vs 33.3, $P = .99$).

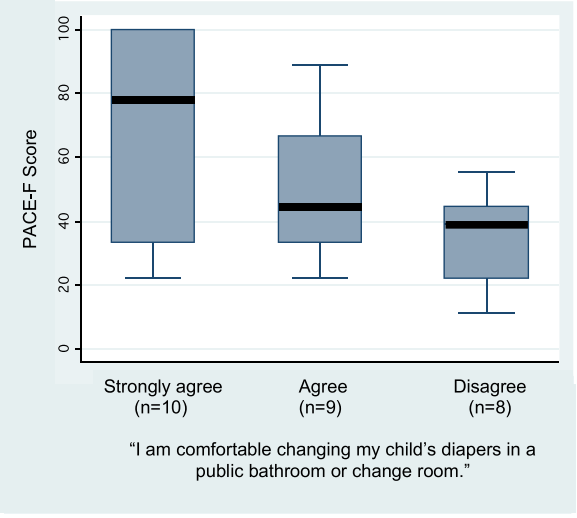

Figure 1. PACE-F scores stratified by comfort with changing diapers in a public bathroom for parents of girls with CAH younger than 4 years old ($P = .03$ test for trend, $n = 27$). (Color version available online.)

Prospective Analysis: Girls <4 Years Old With CAH Undergoing FGRS

Baseline and postoperative PACE-F scores were available for 10 girls undergoing FGRS (23 parental responses). FGRS occurred at median 8 months old (80.0% vaginoplasty, 100.0% PUM, 100.0% clitoroplasty). All scores improved at median 4 months after surgery (median: 38.9 vs 88.9, $P = .01$, Fig. 2). Pre-FGRS scores were lower than age-matched controls ($P = .0001$) and postoperative scores were undistinguishable from age-matched controls ($P = .18$). All scores improved from below to within the population reference range.

Preoperatively, parents of girls with CAH reported greater discomfort changing diapers in public compared to age-matched controls ($P = .0002$). Level of comfort improved after surgery ($P = .02$) and was undistinguishable from age-matched controls ($P = .43$). After surgery, all parents of girls with CAH agreed/strongly agreed they were comfortable changing diapers. These findings indicate PACE-F is a sensitive instrument, responsive to change.

Sensitivity to Change Validity ($n = 10$)

After surgery, mean PACE-F scores increased by 47.2 (SD 16.8). This signifies a large effect size (47.2/16.8 = 2.8).

Cross-sectional Analysis: Girls <4 Years Old With Either a Pre- or Post-FGRS PACE-F Score

Eleven girls younger than 4 years old were included in this analysis (4 pre-, 7 post-FGRS, 13 parental scores). The 7 parents provided PACE-F scores at a median of 19 months after FGRS (85.7% vaginoplasty, 71.4% PUM, 85.7% clitoroplasty). Median pre-FGRS scores were lower than post-FGRS scores (36.1 vs 77.8, $P = .046$). Three of four pre-FGRS scores were below the population reference range and lower than the age-matched controls (median: 100.0, $P = .001$). All post-FGRS scores were in the reference range, but lower than controls ($P = .01$).

Cross-sectional Analysis: Girls 4-12 Years Old With CAH After FGRS

Twenty PACE-F scores were available for 20 girls who underwent FGRS by the age of 3 (90.0% vaginoplasty, 95.0% PUM, 95.0% clitoroplasty) with a median follow-up of 7.8 years (median age 8.5 years). Median scores were undistinguishable between girls with CAH and controls (100.0 vs 88.9, respectively, $P = .77$).

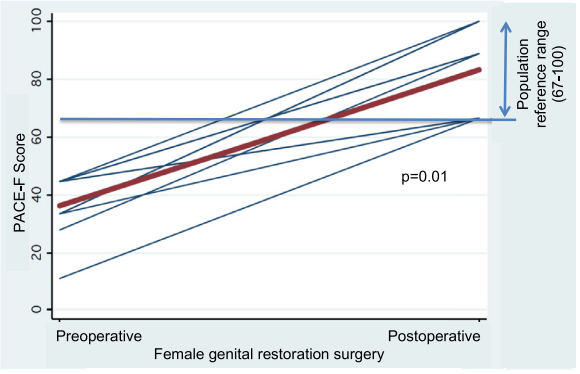

Figure 2. Parental assessment of children's external genitalia scale for females (PACE-F) for 10 girls with CAH prospectively administered before and after female genital restoration surgery. (Color version available online.)

Two of 20 girls (10.0%) had PACE-F scores below the population reference range (22.2 and 33.3). This was not statistically distinct from the expected 1/20 (5%) of responses being below the 95% confidence interval-derived population reference range ($P = .99$). Both patients underwent a vaginoplasty and PUM without surgical complications in the first year of life. One patient did not undergo a clitoroplasty at initial surgery and parents had concerns about clitoromegaly at age 6 (awaiting patient to make decision about further surgery). Second patient underwent FGRS and was 5 years old at follow-up.

DISCUSSION

We present a validation of parent-reported genital appearance scale for use in prepubertal girls with CAH. The genital appearance of infants and toddlers with CAH, per parents, is outside the population reference range. Short-term prospective evidence indicates that genital appearance improves after FGRS. Cross-sectional data suggest these results are maintained until puberty.

Few PROs explicitly address issues of appearance and function associated with ambiguous genitalia in a way that is both sensitive and meaningful to patients, families, and clinicians. The validation of the parent-reported PACE-F is a necessary first step toward giving those living with CAH and their families a voice in an effort replace anecdote-based reporting with robust data.

While following the statistical rigor of PRO validation, we wished to ensure that the resulting instrument was clinically meaningful. For that reason, we determined the population reference range ("What is a reference to compare to?") as well as the minimally important difference between scores ("What is a significant difference?"). These characteristics are rarely established for PROs, restricting their real-life implementation. It should be noted that scores derived using PACE-F, or other questionnaires, are intended to be only one of the tools to provide quality care.

A single-item question was not used as a final PRO. First, a single item tends to have a limited scale, curbing its discrimination ability, sensitivity to change, and usefulness in both clinical and research settings. More importantly, a single item ignores potential multiple dimensions of a single concept. In this instance, parents of girls with CAH (and women who endorsed the original FGSIS items) submit that genital appearance encompasses aspects of not only satisfaction and embarrassment in general, but also comfort during future sexual contact in particular. We believe a multi-item questionnaire better reflects this multifaceted concept of female genital appearance.

The mean PACE-F score for girls with CAH, including those without surgery, was 67.3 (SD 28.1), similar to FGSIS scores (linearly transformed to 0-100) for the general population of adult women without known ambiguity (mean: 67.0, SD 21.7). The PACE-F MID was 14, also similar to the FGSIS of 11 points. While the MID for FGSIS is not published, it was derived using the approach outlined in the Methods section (10.8 for 0.5 SD, 8.1 for SEM using Cronbach's alpha). Undoubtedly, the two measures differ in who reports them and how many items they contain, but the similarity of scores informs us about women's perceptions of their own genitalia. It appears that healthy women experience (dis)satisfaction with their genital appearance comparable to parents of girls with CAH, begging the question of what women in the general population consider "normal" or "satisfactory."

Our study has several limitations. Given its small size, further work is required to externally validate PACE-F in larger, particularly prospective, cohorts. We did not assess test-retest reliability of PACE-F scores, although FGSIS has excellent test-retest reliability.[12] This will be the focus of future work. Despite noting a trend, we did not note a statistically significant correlation between the level of virilization and PACE-F scores. This was most likely due to low statistical power. Alternatively, it is plausible that parental perceptions of genital appearance may be more of a more binary phenomenon, rather than a gradient associated with levels of virilization.

Correlations between parental and patient assessments of genital appearance remain unknown. How PACE-F scores relate to anatomical variables, including clitoral appearance, and functional outcomes requires a multi-institutional study. Finally, since all older girls in our study underwent FGRS, it is crucial to note that no scientific data exist on how parents, girls, and women with CAH who have not undergone FGRS perceive their genitalia.

CONCLUSION

We present a new, validated instrument for a parental assessment of genital appearance in girls with CAH. We also demonstrate improvement in parent-reported appearance after FGRS, which appears to be similar to age-matched controls. PACE-F may prove useful in clinical practice and research.

References

1. Hughes IA. Management of congenital adrenal hyperplasia. *Arch Dis Child*. 1988;63:1399–1404.
2. Yau M, Khattab A, Pina C, et al. Defects of adrenal steroidogenesis. In: DeGroot LJ, Jameson JL eds. *Endocrinology: Adult and Pediatric*. Philadelphia: Saunders/Elsevier; 2010:1810–1832. e1816.
3. Rink RC, Kaefer M: Surgical management of disorders of sexual differentiation, cloacal malformation, and other abnormalities of the genitalia in girls. In: Wein AJ, Kavoussi LR, Campbell MF eds. *Campbell-Walsh Urology*, Editor-in-Chief, Alan J. Wein; [Editors, Louis R. Kavoussi ... et al.]. Philadelphia, PA: Elsevier Saunders; 2012, vol. 4, 3629−−3666.
4. Diamond M, Garland J. Evidence regarding cosmetic and medically unnecessary surgery on infants. *J Pediatr Urol*. 2014;10:2–6.
5. Lee PA, Houk CP. Commentary to 'Evidence regarding cosmetic and medically unnecessary surgery on infants. *J Pediatr Urol*. 2014;10:7.
6. Krege S, Walz KH, Hauffa BP, et al. Long-term follow-up of female patients with congenital adrenal hyperplasia from 21-hydroxylase deficiency, with special emphasis on the results of vaginoplasty. *BJU Int*. 2000;86:253–258; discussion 258-259.
7. Creighton S, Chernausek SD, Romao R, et al. Timing and nature of reconstructive surgery for disorders of sex development—introduction. *J Pediatr Urol*. 2012;8:602–610.
8. Braga LH, Pippi Salle JL. Congenital adrenal hyperplasia: a critical appraisal of the evolution of feminizing genitoplasty and the

controversies surrounding gender reassignment. *Eur J Pediatr Surg*. 2009;19:203–210.
9. Speiser PW, Azziz R, Baskin LS, et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95:4133–4160.
10. Pasch L, He SY, Huddleston H, et al. Clinician vs self-ratings of hirsutism in patients with polycystic ovarian syndrome: associations with quality of life and depression. *JAMA Dermatol*. 2016;152:783–788.
11. Hinami K, Alkhalil A, Chouksey S, et al. Clinical significance of physical symptom severity in standardized assessments of patient reported outcomes. *Qual Life Res*. 2016; 25:2239-2243.
12. Herbenick D, Schick V, Reece M, et al. The female genital self-image scale (FGSIS): results from a nationally representative probability sample of women in the United States. *J Sex Med*. 2011;8:158–166.
13. Kincaid JP, Fishburne Jr. LRP, Rogers RL, et al. Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. *Res Branch Rep*. 1975;8-75:1–48.
14. Joliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. *Stat Methods Med Res*. 1992;1:69–95.
15. Streiner DL. Figuring out factors: the use and misuse of factor analysis. *Can J Psychiatry*. 1994;39:135–140.
16. Bryant FB, Yarnold PR. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm LG, Yarnold RR, eds. Reading and Understanding Multivariable Statistics. Washington, DC: American Psychological Association; 1995:99–136.
17. Gorsuch RL. *Factor Analysis*. Hillsdale, NJ: L. Erlbaum Associates, 1983, 425 p.
18. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford; New York: Oxford University Press, 2003, 283 p.
19. World Health Organization. *WHO Laboratory Manual for the Examination and Processing of Human Semen*. Geneva: World Health Organization, 2010, 271 p.
20. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. *Expert Rev Pharmacoecon Outcomes Res*. 2004;4:581–585.
21. Chassany O, Sagnier P, Marquis P, et al. Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. *Ther Innov Regul Sci*. 2002;36:209–238.
22. Wyrwich KW, Bullinger M, Aaronson N, et al. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res*. 2005;14:285–295.
23. Cohen J. A power primer. *Psychol Bull*. 1992;112:155–159.
24. Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2:63–67.
25. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377–381.
26. Rink RC, Metcalfe PD, Kaefer MA, et al. Partial urogenital mobilization: a limited proximal dissection. *J Pediatr Urol*. 2006;2:351–356.