The External Genitalia Score (EGS): A European Multicenter Validation Study

Saskia van der Straaten,[1] Alexander Springer,[2] Aleksandra Zecic,[3] Doris Hebenstreit,[4] Ursula Tonnhofer,[2] Aneta Gawlik,[5] Malgorzata Baumert,[6] Kamila Szeliga,[5] Sara Debulpaep,[7] An Desloovere,[7] Lloyd Tack,[1] Koen Smets,[3] Malgorzata Wasniewska,[8] Domenico Corica,[8] Mariarosa Calafiore,[8] Marie Lindhardt Ljubicic,[9] Alexander Siegfried Busch,[9] Anders Juul,[9] Anna Nordenström,[10] Jon Sigurdsson,[10] Christa E. Flück,[11] Tanja Haamberg,[11] Stefanie Graf,[11] Sabine E. Hannema,[12,13] Katja P. Wolffenbuttel,[14] Olaf Hiort,[15] S. Faisal Ahmed,[16] and Martine Cools[1]

[1]Ghent University Hospital, Department of Pediatrics, Division of Pediatric Endocrinology and Ghent University, Department of Internal Medicine and Pediatrics, 9000 Ghent, Belgium; [2]Medical University of Vienna, Department of Pediatric Surgery, 1090 Vienna, Austria; [3]Ghent University Hospital, Department of Neonatology, 9000 Ghent, Belgium; [4]Medical University Innsbruck, Department of Urology; [5]Medical University of Silesia, Department of Pediatrics, Pediatric Endocrinology, 40-055, Katowice, Poland; [6]Medical University of Silesia, Department of Neonatology, 40-055, Katowice, Poland; [7]Ghent University Hospital, Department of Pediatrics, 9000 Ghent, Belgium; [8]University Hospital of Messina, Department of Human Pathology of Adulthood and Childhood, 98124 Messina, Italy; [9]Copenhagen University Hospital - Rigshospitalet, Department of Growth and Reproduction, 2100 Copenhagen, Denmark; [10]Karolinska University Hospital Stockholm, Department of Women's and Children's Health Karolinska Institutet, S-171 77 Stockholm, Sweden; [11]Bern University Children's Hospital Inselspital, Department of Pediatrics, Division of Pediatric Endocrinology, Diabetology and Metabolism and Department of BioMedical Research, University of Bern, 3010 Bern, Switzerland; [12]Erasmus MC - Sophia Children's Hospital, University Medical Center Rotterdam, Department of Pediatric Endocrinology, 3015 GD Rotterdam, the Netherlands; [13]Leiden University Medical Centre, Department of Paediatrics, 2300 RC Leiden, The Netherlands; [14]Erasmus MC - Sophia Children's Hospital, University Medical Center Rotterdam, Department of Urology and Paediatric Urology, 3015 GD Rotterdam, the Netherlands; [15]University of Lübeck, Division of Paediatric Endocrinology and Diabetes, Department of Paediatrics, 23538 Lübeck, Germany; and [16]Developmental Endocrinology Research Group, University of Glasgow, Glasgow, G12 8QQ, UK

Context: Standardized description of external genitalia is needed in the assessment of children with atypical genitalia.

Objectives: To validate the External Genitalia Score (EGS), to present reference values for preterm and term babies up to 24 months and correlate obtained scores with anogenital distances (AGDs).

Design, Setting: A European multicenter (n = 8) validation study was conducted from July 2016 to July 2018.

Patients and Methods: EGS is based on the external masculinization score but uses a gradual scale from female to male (range, 0–12) and terminology appropriate for both sexes. The

Abbreviations: AGD, anogenital distance;AGDac, ano-clitoral distance from the center of the anus to the anterior base of the clitoris; AGDaf, ano-fourchettal distance from the center of the anus to the fourchette;AGDap, ano-penile distance from the center of the anus to the anterior base of the penis; AGDas, ano-scrotal distance from the center of the anus to the posterior base of the scrotum; AGDlower (AGDl), measured from the center of the anus to the base of the labioscrotal border;AGDupper (AGDu), measured from the center of the anus to the anterior base of the genital tubercle;AGDl/u, lower/upper AGD ratio; CAH, congenital adrenal hyperplasia; CI, confidence interval; COST, European Cooperation in Science and Technology; DSD, difference of sex development; EGS, External Genitalia Score; EMS, External Masculinization Score; GTL, genital tubercle length; ICC, interclass correlation coefficient; PS, Prader score; TIDES, the Infant Development and the Environment Study.

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received 25 June 2019. Accepted 25 October 2019.
First Published Online 29 October 2019.
Corrected and Typeset 19 February 2020.

reliability of EGS and AGDs was determined by the interclass correlation coefficient (ICC). Cross-sectional data were obtained in 686 term babies (0–24 months) and 181 preterm babies, and 111 babies with atypical genitalia.

Results: The ICC of EGS in typical and atypical genitalia is excellent and good, respectively. Median EGS (10th to 90th centile) in males < 28 weeks gestation is 10 (8.6–11.5); in males 28–32 weeks 11.5 (9.2–12); in males 33–36 weeks 11.5 (10.5–12) and in full-term males 12 (10.5–12). In all female babies, EGS is 0 (0-0). The mean (SD) lower/upper AGD ratio (AGDl/u) is 0.45 (0.1), with significant difference between AGDl/u in males 0.49 (0.1) and females 0.39 (0.1) and intermediate values in differences of sex development (DSDs) 0.43 (0.1). The AGDl/u correlates with EGS in males with typical genitalia and in atypical genitalia.

Conclusions: EGS is a reliable and valid tool to describe external genitalia in premature and term babies up to 24 months. EGS correlates with AGDl/u in males. It facilitates standardized assessment, clinical decision-making and multicenter research. **(*J Clin Endocrinol Metab* 105: 1–9, 2020)**

Key Words: external genitalia score, external masculinization score, anogenital distances, atypical genitalia

Differences (or disorders) of sex development (DSDs) are heterogeneous congenital conditions that affect the development of the urogenital tract and reproductive system and result in atypical sex differentiation (1–4). The incidence of DSDs where sex assignment may be unclear at birth is estimated at 1/5500 births (5). For milder variations, such as hypospadias, prevalence rates vary from 13.8/10 000 to 40/10 000 (6, 7). The clinical management of these conditions is complex and requires specialized care by a multidisciplinary team (1, 8). A precise understanding of the underlying cause, preferably up to the molecular genetic level, is crucial to allow individualized management as well as for research purposes. Detailed evaluation of the genital phenotype will inform clinicians about the need for further referral to an expert center, and guide them to specific diagnostic tests such as hormonal, imaging, and genetic investigations (9). The genital phenotype at birth has also been related to long-term outcomes, such as with regard to genital (dis)satisfaction (10), the prevalence of cardiac (11) or other comorbidities (12) or the risk for the development of gonadal germ cell tumors (13, 14). The relevance of a precise description of the genital phenotype has even increased in recent years as genital surgery in childhood has become controversial, and currently many children who have a DSD grow up with a genital difference. The long-term outcome of this approach will need to be determined. Lastly, given that the individual DSD conditions are (very) rare, meaningful research requires a multicenter approach and thus a standardized battery of tools across centers to assess and document this phenotypic variability.

A comprehensive genital exam contains the following landmarks: the presence and location of the gonads, genital tubercle development, degree of fusion of the labioscrotal folds and location of the urethral meatus. A micropenis is defined as a short penis, ≤ 24–25 mm (ie, ≤ 2.5 SD below the mean) with a normal configuration (15). Minor racial differences for stretched penile length have been published (16). The distance between the anus and various landmarks of the external genitalia has been shown to be a sensitive index of androgen activity during fetal development and is sexually dimorphic (17, 18). Various anogenital distances (AGDs) have been proposed. In male term newborns, the mean (SD) anoscrotal anogenital distance (AGDas), measured from the center of the anus to the posterior scrotal wall is 24.7 (4.5) mm. In female term newborns the mean (SD) ano-fourchette AGD (AGDaf), measured from the center of the anus to the fourchette is 16.0 (3.2) mm (17). AGDas and AGDaf are represented in Fig. 1 as lower AGD (AGDl), while anopenile AGD (AGDap) and anoclitoral AGD (AGDac) are represented as upper AGD (AGDu). A shorter AGDas and penile length have been found in infants with hypospadias and cryptorchidism, a longer AGDaf has been described in female infants with androgen excess, for example, in congenital adrenal hyperplasia (CAH). In typical female infants, it was shown that calculating the anogenital-ratio (AGDaf:AGDac) offers advantages as it follows a normal distribution and does not correlate with anthropometric variables or gestational age (19, 20). The Prader score (PS) was developed by Andrea Prader in 1954 to capture genital variation in children who have CAH. Apart from the typical female and male phenotypes, it categorizes external genitalia in children with CAH in 5 additional stages with progressive virilization from a phenotypic female with mild clitoromegaly (stage 1) to a phenotypic male with glandular hypospadias (stage 5) (21). In 2000, the External Masculinization Score (EMS) was introduced to improve the initial assessment of boys with a genital difference. The EMS (range 0–12) allocates points to 5 different characteristics of the external genitalia (scrotal fusion 3/0, micropenis < 25 mm 3/0, urethral meatus 3/2/1/0, right and left gonad 1,5/1/0) (22). The EMS allows standardization of genital assessment, but a refinement of the score is needed to capture the appearance of the genitalia more comprehensively across the phenotypic spectrum in both sexes. We here present the External Genitalia Score (EGS) (Table 1) (23) as a modified, nonbinary version of EMS. EGS was developed by Working Group 1 of the European Cooperation in Science and Technology (COST) Action BM1303. The EGS uses a gradual scale from female to male (range, 0–12) of the same anatomical landmarks as the EMS. To provide a full description of the external genitalia, the various AGDs were measured and, in line with EGS, a gender-neutral lower/upper AGD ratio (AGDl/u) was calculated as a marker of genital virilization independent of body (Fig. 1).

Materials and Methods

Based on expert opinion and group discussions, members of the DSDnet COST Action (www.dsdnet.eu) Working Group 1 modified the existing EMS to describe the same anatomical features with a refined categorical scale for the items labioscrotal fusion, urethral meatus, and the position of the gonads and a continuous scale for the size of the genital tubercle, ranging from typical female to typical male (Table 1) (23). In addition, the vocabulary was adjusted in a way that suits both sexes.

Measurements

Genital assessment and measurements included EMS and EGS, PS, and AGDs. The same digital caliper (Carbon Fiber Digital Caliper, resolution: 0.1 mm, QST-Express, type QST008, China) was used for all measurements across centers. Length of the genital tubercle (GTL) was measured along its dorsal aspect in a nonerect state, gently stretching it between two fingers until the point of increased resistance, from the base of the genital tubercle (as close to the pubic bone as possible) to the tip of the glans and excluding the foreskin (15). The measurement was performed twice, and the mean was calculated. Location of the gonads was determined by palpation, as described by Ogilvy-Stuart (3). The position of the meatus and degree of labioscrotal fusion were determined by visual inspection. AGD measurements were standardized according to the Infant Development and the Environment Study (TIDES) (17), with some modifications, and the accompanying training video (kindly provided by the TIDES research group) was distributed among participating centers. Modifications to the

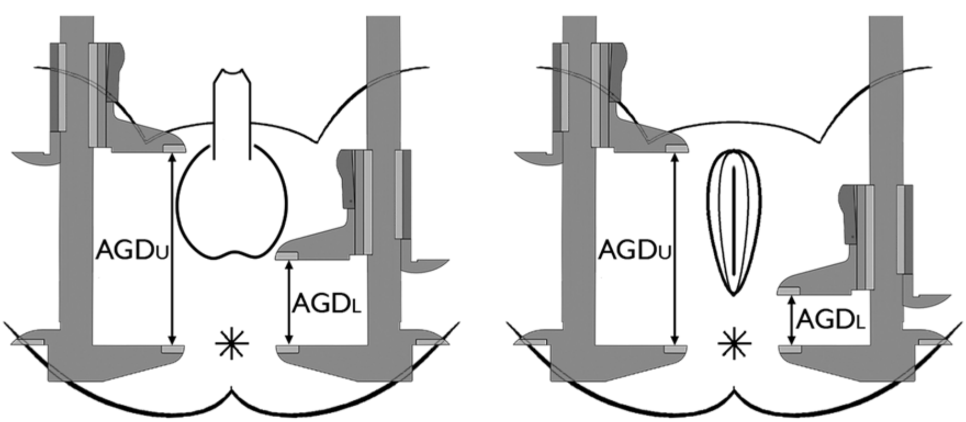

Figure 1. Measurement of anogenital distances. In order to obtain a single measure that is suitable for all babies, AGDap and AGDac were defined as AGDu, and AGDas and AGDaf as AGDl. Abbreviations: AGDl (AGDlower), measured from the center of the anus to the base of the labioscrotal border; AGDu (AGDupper), measured from the center of the anus to the anterior base of the genital tubercle

Table 1. ***External Genitalia Score* Describes Phenotypic Features at 5 Anatomical Landmarks of the Genitalia: Degree of Labioscrotal Fusion, Length of the Genital Tubercle, Position of the Urethral Meatus, and Locations of the Right and Left Gonads. The Final Score is the Sum of Points Allocated to Features 1–5.**

EGS	Labioscrotal Fusion	Genital Tubercle Length (mm)	Urethral Meatus	Right Gonad	Left Gonad
3	Fused	>31	Top of the GT		
2.5		26–30	Coronal Glandular		
2			Along the GT		
1.5	Posterior fusion	21–25	At the GT base	Labioscrotal	Labioscrotal
1		10–20	Labioscrotal	Inguino-Labioscrotal	Inguino-Labioscrotal
0.5				Inguinal	Inguinal
0	Unfused	< 10	Perineal	Impalpable	Impalpable

Abbreviations: EGS, External Genitalia Score; GT, genital tubercle

TIDES method include placement of the baby in the middle of the bed instead of at the edge to allow the same position in premature babies in the incubator. For the same reason, the fixed end of the caliper is held at the center of the anus, and the sliding part of the caliper is moved while measuring the AGDs. The sliding part touches but does not compress the skin at the anterior base of the genital tubercle to standardize measurements in male and female infants. The examiner does not use a marker for the mid-anus position but chooses a wrinkle in the center of the anus to use for the measurement of the 2 AGDs. The average of 2 measurements is used for analysis instead of the average of 3 measurements. AGDl was measured from the center of the anus to the base of the labioscrotal border and AGDu was measured from the center of the anus to the anterior base of the genital tubercle. PS was determined by visual inspection and EMS, EGS, and AGDl/u were calculated based on the obtained scores and AGD measurements.

Participants

First, the inter-observer reliability of PS, EMS, EGS and inter- and intra-observer reliability of AGDl/u were determined by 2 observers from 2 different centers in 35 babies with typical genitalia (12 female, 23 male; 12 preterm, 23 term). Subsequently, the reliability of these parameters was assessed by 2 observers in 4 different centers in 66 babies with atypical genitalia: males with "mild nonspecific undermasculinization" (ie, isolated hypospadias [n=29] or isolated cryptorchidism [n=8]), and babies with 46,XY DSD (n=22), sex chromosome DSD (n=2), and 46,XX DSD (n=5) (23).

A collaborative multicenter study was then conducted in 8 European clinical centers from July 2016 to July 2018, to establish reference data for the EGS in typical genitalia (Table 2). For this purpose, the external genitalia of preterm infants, term infants up to 1 month and babies from 1 to 24 months of age were assessed by 1 observer per center and PS, EMS, EGS, GTL, and AGDl/u were determined. The following potential covariates were noted: maternal age, ethnicity, virilization and medication use in pregnancy, exposure to toxic products, smoking in pregnancy, history of consanguinity, gestational age at birth, weight and length at birth, weight and length at assessment. Children with a major congenital malformation (central, cardiac, pneumologic, urologic) were excluded. In total, 181 (105 male and 76 female) preterm (< 37 weeks) neonates, 378 (178 male and 200 female term neonates), and 308 (153 male and 155 female) babies aged 1–24 months were assessed (a total of 686 babies). In 4 clinical centers the PS, EMS, EGS, and AGD l/u were obtained in babies with atypical genitalia (23).

Statistical analyses

The inter-observer reliability of the PS, EMS, and EGS and the intra- and inter-observer reliability for AGDs were assessed by Intraclass Correlation Coefficient (ICC) estimates and their 95% confidence intervals (CI), based on absolute-agreement, 2-way random-effect model (24). The median (10th– 90th centile) for EGS and EMS were generated. Spearman's rho determined the correlation between EGS and EMS since both have a skewed distribution. The Bland–Altman analysis was used to assess agreement between the EGS and EMS. This method calculates the mean difference between 2 methods, and 95% limits (2 SD) of agreement of the differences between the 2 methods (25). The mean (SD, 10th– 90th centile) for the different AGDs and AGD-ratios was calculated in typical and in atypical genitalia. Correlations of the different AGDs and the AGDl/u with weight, length and age were assessed by Pearson analysis. Potential covariates of the different AGDs and AGDl/u were assessed by linear regression. A Spearman's correlation

Table 2. EGS in Female and Male Babies With Typical Genital Phenotypes in Different Gestational Age, Birth Weight, and Age Groups

EGS Values in Female and Male Babies With a Typical Genital Phenotype				
FEMALE				
	N	**Median**	**P10**	**P90**
All gestational ages, birth weights, age 0-24 months	**424**	**0**	**0**	**0**
MALE				
Gestational age (weeks)	N	Median	P10	P90
<28	11	10	8.6	11.5
28-32,9	23	11.5	9.2	12
33-36,9	31	11.5	10.5	12
>37	178	12	10.5	12
Birth weight (g)	N	Median	P10	P90
<1000	12	10	8.7	11.9
1000–1499	13	11.5	8.4	11.5
1500–2499	33	11.5	10.5	12
2500–4000	163	12	10.5	12
>4000	22	12	10.8	12
Age (months)	N	Median	P10	P90
0–1	178	12	10.5	12
1–6	98	12	11.5	12
6–12	39	12	11	12
12–24	16	12	11.9	12

Abbreviations: EGS, External Genitalia Score

Table 3. Genital Tubercle Length, AGDl/u in Male (light grey) and Female (dark grey) Babies With a Typical Genital Phenotype

Genital Tubercle Length in Male Infants With Typical Genital Phenotype					
Gestational age (weeks)	N	Mean (mm)	SD	P10	P90
<28	11	21.5	4.7	12.0	27.4
28–33	21	27.5	4.7	20.3	27.4
33–37	27	28.3	4.7	21.8	34.0
>37	174	31.2	5.4	23.5	38.0
Age Group (months) full term infants					
0–1	174	31.2	5.4	23.5	38.0
1–6	96	31.1	5.2	26.0	38.6
6–12	39	33.4	5.0	26.7	39.6
12–24	13	37.8	3.7	32.4	44.5
AGDl/u in male infants with typical genital phenotype					
Gestational age (weeks)					
<28	11	0.45	0.1	0.35	0.58
28–33	23	0.44	0.1	0.36	0.52
33–37	31	0.50	0.1	0.38	0.63
>37	178	0.52	0.1	0.41	0.65
Age Group (months) full term infants					
0–1	178	0.52	0.1	0.42	0.65
1–6	97	0.48	0.1	0.38	0.56
6–12	39	0.49	0.1	0.41	0.60
12–24	15	0.49	0.1	0.34	0.63
AGDl/u in full-term male and female infants with typical genital phenotype					
AGD ratio's					
AGDl/u	178	0.49	0.1	0.42	0.65
AGDl/u	200	0.39	0.1	0.29	0.48

Abbreviations: AGD, anogenital distance; AGDl/u, lower/upper AGD ratio; AGDu (AGDupper), measured from the center of the anus to the anterior base of the genital tubercle; AGDl (AGDlower), measured from the center of the anus to the base of the labioscrotal border.

was done to determine the relationship between EGS and AGDl/u. An independent-samples t-test was conducted to compare AGDl/u in typical and atypical genitalia. All analyses were performed using the SPSS statistical package version 25.

Ethics

The study was approved by the local ethical committees of each participating center (Local IDs: Ghent: B670201628499, Medical University of Vienna 1872/2014), Rotterdam: MEC-2016–706, Copenhagen: H-15014876 and RH-2015-210-04146, Katowice: KNW/0022/KB1/158/I/16/17/18, Stockholm Karolinska University Hospital 2008/167-31/3, 2009-01-13, 10-12-16. Messina: MEC 104/16. Informed consent was obtained from at least one parent or legal guardian for each child.

Results

Inter-observer reliability of EGS in comparison with EMS and PS, and of the various AGDs

As the EGS is a more refined modification of the EMS, we compared its reliability with the original EMS, and to the PS, which is historically the most widely used score. Inter-observer ICC (n=35) for EGS showed no case of disagreement between any scorer (excellent) in typical genitalia (ICC=1) and, the interobserver variability in atypical genitalia (n=66) was good (ICC=0.89; CI, 0.82–0.93). Likewise, inter-observer ICC for PS and EMS also showed no case of disagreement in typical and were moderate and good in atypical genitalia. Inter-observer ICC for the different AGDs and genital tubercle length were moderate for AGDu and good for AGDl and genital tubercle length in typical male genitalia and good for AGDu, AGDl and genital tubercle length in atypical genitalia. Inter-observer ICC were good for AGDl and, AGDu in typical female genitalia. Intra-observer ICC for the different AGDs and genital tubercle length were good or excellent in both typical and atypical genitalia (23).

Reference data for genital tubercle length, EGS, AGDs and AGDl/u

As a new measuring instrument, we established reference data for EGS, including data in pre- and dysmature babies who present more often with atypical genitalia (12). In addition, we determined AGDs and AGDl/u to investigate correlations of EGS with other measures of genital virilization. In male term infants with typical genitalia, the mean (SD) genital tubercle length (n=174), AGDl, and AGDu (n=178) were 31.2 (5.4), 24.6 (4.7), and 47.6 (5.8) mm respectively. In female term infants with typical genitalia (n=200), the mean (SD) length of AGDl and AGDu were 14.8 (3.5) and 37.8 (4.5) mm, respectively. AGDl/u was independent of body weight

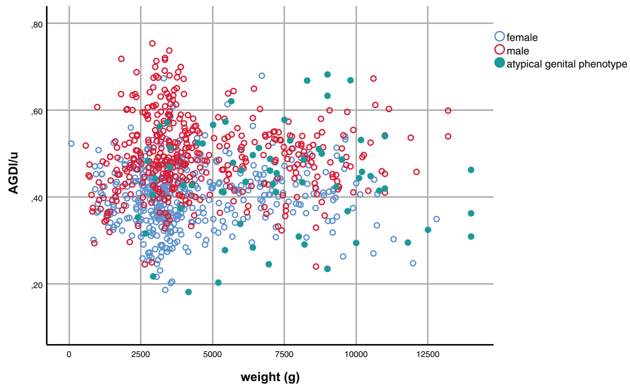

Figure 2. Correlation between AGDl/u and weight in babies with typical genitalia and atypical genitalia. Abbreviations: ratio AGDl/u, lower/upper AGD ratio; AGDl, measured from the center of the anus to the base of the labioscrotal border; AGDu, measured from the center of the anus to the anterior base of the genital tubercle.

(Fig. 2). Although mean (SD) AGDl/u in male infants [0.49 (0.1)], significantly differs from AGDl/u in female infants [0.39 (0.1)], large overlap exists between both groups (Table 3, Fig. 2). AGD l/u in male neonates positively correlates with gestational age (r (243) = 0.3; $P < 0.05$). No univariate or bivariate correlation was detected between AGD l/u and any of the other covariates (maternal age, ethnicity, center, virilization and/or medications used in pregnancy, exposure to toxic products or smoking during pregnancy). In typical male infants, the median and 10th centile EGS gradually rise with increasing gestational age and birth weight due to increasing genital tubercle length and descent of the testes (Fig. 3A and 3B). In addition, the EGS 10th centile gradually increases with age up to 24 months. Median EGS in typical female premature and full-term babies up to 24 months is 0 (0-0) (Table 2).

Genital tubercle length, EGS, AGDs, and AGDl/u in children with atypical external genitalia

In babies with atypical genitalia, the EGS covers the whole phenotypic spectrum, resulting in scores ranging from 0 to 12 with large overlap between the various DSD categories (46,XX DSD, 46,XY DSD, and 45,X/46,XY DSD) (Table 4 and Fig. 3C). In male babies with atypical genitalia (46,XY DSD and "mild nonspecific undermasculinization"), AGDl/u (M=0.43, SD=0.11) is significantly shorter than AGDl/u in typical males (M=0.49, SD=0.09); t (95.1) = 4.8; $P < 0.05$), however AGDl/u widely varies in babies with atypical genitalia, with a mean 0.43 (0.1 SD) not different from mean AGDl/u 0.45 (0.1) in babies with typical genitalia (Fig. 2).

Correlation and agreement between scores and measures

AGDl, AGDu, and AGDl/u positively correlate with EGS in typical male full-term neonates as well as in

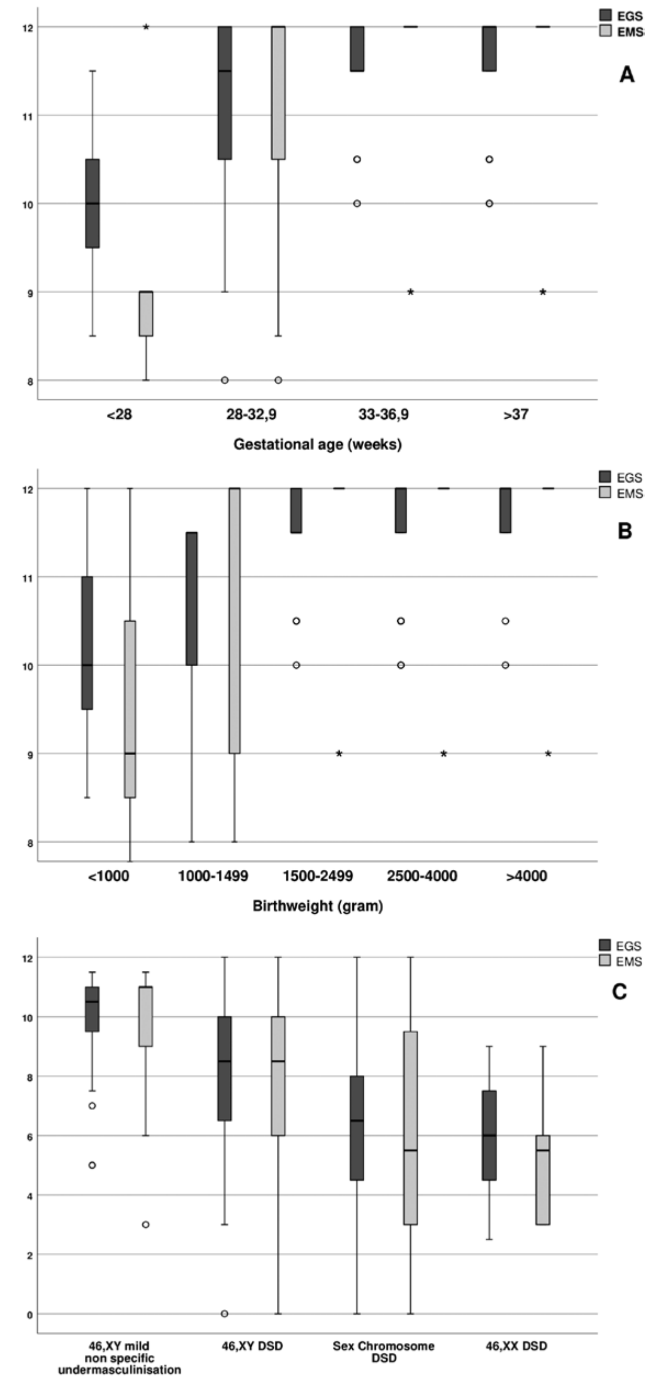

Figure 3. Boxplot with median and interquartile range of EGS (dark grey) in comparison with EMS (light grey). A) Results for typical male babies according to gestational age; B) Results for typical male babies according to birth weight; C) Results for babies with atypical genitalia and various DSD groups. Abbreviations: mild nonspecific undermasculinization, refers to isolated hypospadias or isolated cryptorchidism.

babies with atypical genital phenotypes (r_s (243) = 0.19; $P < 0.05$ and r_s (78) = 0.35; $P < 0.05$ respectively) (23). As expected, there is a strong, positive correlation between EGS and EMS in typical (r_s (853) = 0.97; $P < 0.05$) and atypical genitalia (r_s (110) = 0.9; $P < 0.05$) (23). The Bland-Altman analysis shows that optimal agreement

Table 4. EGS Median,10th–90th Centile Scores and AGDl/u in Babies With Atypical Genital Phenotypes

EGS Scores in Babies With Atypical Genital Phenotypes

Group	N	Median	P10	P90
46,XY DSD	46	8.5	5.5	11.5
Sex Chromosomal DSD	9	6.5		
Mild, nonspecific undermasculinization*	45	10.5	7.5	11.5
46,XX DSD	10	6	2.7	9

AGDl/u in Babies With Atypical Genital Phenotypes

Group	N	Mean	SD	P10	P90
46,XY DSD	33	0.44	0.13	0.23	0.60
mild, nonspecific undermasculinization*	43	0.42	0.10	0.28	0.54
46,XX DSD	7	0.45	0.10	0.32	/

Abbreviations: AGDu, measured from the center of the anus to the anterior base of the genital tubercle; AGDl, measured from the center of the anus to the base of the labioscrotal border; AGDl/u, lower/upper AGD ratio; EGS, External Genitalia Score. *mild nonspecific undermasculinization refers to males with isolated cryptorchidism or isolated hypospadias

between the 2 methods is reached for EMS/EGS results < 3 and > 9.5 (23).

Discussion

The EMS, developed by Ahmed et al in 2000 (22) provides an objective and standardized tool to describe external genitalia in male babies and has been correlated with various DSD-related outcomes (10–14). A major limitation of the EMS in the workup of an infant with atypical genitalia is that it cannot be applied in assigned females because of the gender-specific design and vocabulary (eg, micropenis yes/no, scrotal fusion yes/no). Also, EMS does not capture the full phenotypic spectrum of genital variation that characterizes DSD conditions due to its dichotomous nature. To overcome these problems, COST Action BM1303 Working Group 1 modified the EMS in a gender-neutral and more refined categorical scale, that better reflects the naturally occurring variation (eg, by introducing the option *posterior labioscrotal fusion*). The resulting tool was termed the EGS and was subsequently validated in a large European multicenter study. EGS can be applied in both typical male and female babies and in babies who have variations in their genital characteristics. We provide normative data for premature, low-birth-weight and full-term babies until the age of 2 years for a mixed European population. Such data are of particular relevance given the frequent association in males of intrauterine growth retardation with genital undermasculinization and the difficulties in assessing genital variation in preterm infants whose testes have not yet descended and whose penis has not yet reached its full-term length. Although the EGS can be used for the initial evaluation of babies with atypical genitalia, it cannot fully replace a more detailed qualitative genital description. The EGS does not inform on the presence of other atypical genital features such as complete or partial penoscrotal transposition, scrotal anomalies or degree of penile curvature. Moreover, EGS, like EMS, does not provide information on important internal genital characteristics in the context of DSD, such as the presence of a urogenital sinus or the location of the vaginal confluence in 46,XX babies who have CAH. Bland-Altman analysis reveals that EGS and EMS have least agreement in the group of children with atypical genitalia, that is, children who have an EGS between 3 and 9.5. In our data from 66 children with a DSD, the interquartile ranges are smaller for EGS compared with EMS, supporting our hypothesis that the EGS enables a more refined description of genital virilization. In addition, EGS is easy to use, helps to assess important landmarks of the external genitalia, can be implemented by physicians who do not regularly examine infants with variant genital development, and it serves as an alternative to genital photography, which has ethical constraints. Due to its objectivity and simple design, it is also very instrumental for the exchange of data on genital phenotypes between centers and researchers, for example through large-scale registries such as I-DSD. Future research and clinical use of EGS will reveal whether specific EGS outcomes can be allocated to specific diagnoses/mutated genes, but based on our preliminary data, it is expected that EGS will have little predictive value regarding the underlying diagnosis in most cases, given the large overlap between the various DSD categories. Reference data for the EGS in full-term, preterm, and low-birth-weight children, are of high relevance for a broad audience of pediatricians and general practitioners. According to Ahmed et al (9), clinical evaluation by a specialized DSD-team is advised in proximal forms of hypospadias, isolated micropenis, isolated clitoromegaly, any form of familial hypospadias, and in those who have a combination of genital variations resulting in an EMS of less than 11. All these variations will result in a maximal EGS of

10.5, corresponding to P10 in full-term male infants. Therefore, based on our data, we advise referral to a specialized DSD team for any full-term infant who has an EGS > 0 and ≤ 10.5 (or ≤ P10), and of any preterm or low-birth-weight infant who has an EGS > 0 and ≤ P10 for gestational age or birth weight, independent of maternal age, ethnicity, virilization and/or medications used in pregnancy, exposure to toxic products, or maternal smoking. Of note, the obtained EGS will not lead to a specific diagnosis in an infant who has variant genitalia, but it may justify further genetic, biochemical, and hormonal diagnostic investigations. Further research is mandatory to determine whether this recommendation will require adjustments in the future.

The anogenital distance has been shown to be a surrogate marker of prenatal androgen exposure and has been correlated to various endocrine-reproductive outcomes (26–28). Although it adds to the description of the external genitalia (29), its clinical use is limited as it is relatively time-consuming and measurements are hard to standardize among different observers. As AGD is known to correlate with anthropometric variables, which was confirmed in our study, the AGD ratio may represent a more useful marker. In our study, AGDl/u followed a normal distribution and did not correlate with any of the anthropometric variables. Moreover, while mean AGDl/u significantly differs between typical males and typical females, this measure underscores the naturally encountered variation in genital phenotypes, both in typical males and females and in children who have a DSD, as becomes obvious from Fig. 2. As expected, AGDl/u correlates with EGS in undermasculinized infants, both measurements independently reflecting the degree of prenatal androgen exposure.

A major strength of our study is its multicenter design, allowing data collection in a large European sample in a relatively short period. At the same time, this multicenter approach may constitute a weakness, since some measurements, such as AGD and genital tubercle length are prone to larger inter-observer variability. This was also confirmed by the variable ICC scores obtained for these measures in our study and this may explain the relatively large SD obtained for these parameters. In addition the assessment of children with atypical genitalia was performed in 4 out of 8 centers, which could have led to recruitment bias.

In conclusion, the EGS is a reliable and easy-to-use tool that allows objective and detailed description of typical and variant external genitalia in neonates and infants. This facilitates clinical management and data exchange across centers, to study outcomes or draw genotype-phenotype correlations. We provide European reference data for term and premature neonates, for neonates who have low birth weight and for toddlers up to 24 months.

Acknowledgments

The authors would like to thank Mrs Ellen Deschepper for statistical support.

Financial Support: This article is based upon work from COST Action DSDnet, supported by COST (European Cooperation in Science and Technology), grant BM1303. M.C. holds a senior clinical investigator grant from the Research Foundation Flanders, A.G., K.S., and M.B. are supported by statutory work of Medical University of Silesia (number KNW-1–011/N/8/K). M.L.L, A.S.B. and A.J. are supported by COPENHAGEN Minipuberty Study (R146-A5644) and The Danish Environmental Protection Agency and Aase and Ejnar Danielsens Foundation (10-00101874). C.E.F. is supported by the Swiss National Science Foundation (grant number 320030-146127).

Author Contributions: Saskia van der Straaten coordinated overall data collection, performed the interobserver validation study, performed local data collection, analysed results and drafted the manuscript. Alexander Springer developed the EGS, designed the study, supervised the interobserver validation study, coordinated local data collection, analysed results and revised the manuscript. Aleksandra Zecic supervised local data collection and revised the manuscript. Doris Hebenstreit performed the interobserver validation study, performed local data collection, and revised the manuscript. Ursula Tonnhofer, Aneta Gawlik, Malgorzata Baumert, Kamila Szeliga, Sara Debulpaep, An Desloovere, Lloyd Tack, Koen Smets, Malgorzata Wasniewska, Domenico Corica, Mariarosa Calafiore, Marie Lindhardt Ljubicic, Alexander Siegfried Busch, Anders Juul, Jon Sigurdsson, Tanja Haamberg, Stefanie Graf, Sabine E. Hannema and Katja P. Wolffenbuttel performed local data collection, took part in analyses of the results and revised the manuscript. Anna Nordenström and Christa E. Flück developed the EGS, designed the study, coordinated local data collection, analysed results and revised the manuscript. Olaf Hiort coordinated the COST Action, supervised the overall project and revised the manuscript. S. Faisal Ahmed developed the EGS, analysed results and revised the manuscript. Martine Cools developed the EGS, designed the study, performed local data collection, analysed results, drafted the manuscript and supervised the overall project. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Additional Information

Correspondence and Reprint Requests: Martine Cools, MD, PhD, Ghent University Hospital, Princess Elisabeth Childrens' Hospital, Building 3K12D, room 319, Corneel Heymanslaan 10, 9000 Ghent, Belgium. E-mail: martine.cools@ugent.be.

Disclosure summary: The authors have indicated that they have no financial relationships relevant to this article to disclose. All authors have no conflict of interest relevant to this article to disclose.

Data Availability: The datasets generated during and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. Cools M, Nordenström A, Robeva R, Hall J, Westerveld P, Flück C, Köhler B, Berra M, Springer A, Schweizer K, Pasterski V; COST Action BM1303 working group 1. Caring for individuals with a difference of sex development (DSD): a Consensus Statement. *Nat Rev Endocrinol.* 2018;**14**(7):415–429.
2. Lee PA, Nordenström A, Houk CP, Ahmed SF, Auchus R, Baratz A, Baratz Dalke K, Liao LM, Lin-Su K, Looijenga LH 3rd, Mazur T, Meyer-Bahlburg HF, Mouriquand P, Quigley CA, Sandberg DE, Vilain E, Witchel S; Global DSD Update Consortium. Global disorders of sex development update since 2006: perceptions, approach and care. *Horm Res Paediatr.* 2016;**85**(3):158–180.
3. Ogilvy-Stuart AL, Brain CE. Early assessment of ambiguous genitalia. *Arch Dis Child.* 2004;**89**(5):401–407.
4. Hughes IA, Houk C, Ahmed SF, Lee PA; LWPES Consensus Group; ESPE Consensus Group. Consensus statement on management of intersex disorders. *Arch Dis Child.* 2006;**91**(7):554–563.
5. Thyen U, Lanz K, Holterhus PM, Hiort O. Epidemiology and initial management of ambiguous genitalia at birth in Germany. *Horm Res.* 2006;**66**(4):195–203.
6. Loane M, Dolk H, Kelly A, Teljeur C, Greenlees R, Densem J; EUROCAT Working Group. Paper 4: EUROCAT statistical monitoring: identification and investigation of ten year trends of congenital anomalies in Europe. *Birth Defects Res A Clin Mol Teratol.* 2011;**91**(Suppl 1):S31–S43.
7. Springer A, van den Heijkant M, Baumann S. Worldwide prevalence of hypospadias. *Journal of pediatric urology.* 2016;**12**(3):152.e151–157.
8. Hiort O, Birnbaum W, Marshall L, Wünsch L, Werner R, Schröder T, Döhnert U, Holterhus PM. Management of disorders of sex development. *Nat Rev Endocrinol.* 2014;**10**(9):520–529.
9. Ahmed SF, Achermann JC, Arlt W, Balen A, Conway G, Edwards Z, Elford S, Hughes IA, Izatt L, Krone N, Miles H, O'Toole S, Perry L, Sanders C, Simmonds M, Watt A, Willis D. Society for Endocrinology UK guidance on the initial evaluation of an infant or an adolescent with a suspected disorder of sex development (Revised 2015). *Clin Endocrinol (Oxf).* 2016;**84**(5):771–788.
10. van de Grift TC, Cohen-Kettenis PT, de Vries ALC, Kreukels BPC. Body image and self-esteem in disorders of sex development: a European multicenter study. *Health Psychol.* 2018;**37**(4):334–343.
11. De Groote K, Cools M, De Schepper J, Craen M, François I, Devos D, Carbonez K, Eyskens B, De Wolf D. Cardiovascular pathology in males and females with 45,X/46,XY mosaicism. *Plos One.* 2013;**8**(2):e54977.
12. Cox K, Bryce J, Jiang J, Rodie M, Sinnott R, Alkhawari M, Arlt W, Audi L, Balsamo A, Bertelloni S, Cools M, Darendeliler F, Drop S, Ellaithi M, Guran T, Hiort O, Holterhus PM, Hughes I, Krone N, Lisa L, Morel Y, Soder O, Wieacker P, Ahmed SF. Novel associations in disorders of sex development: findings from the I-DSD Registry. *J Clin Endocrinol Metab.* 2014;**99**(2):E348–E355.
13. Lindhardt Johansen M, Hagen CP, Rajpert-De Meyts E, Kjærgaard S, Petersen BL, Skakkebæk NE, Main KM, Juul A. 45,X/46,XY mosaicism: phenotypic characteristics, growth, and reproductive function–a retrospective longitudinal study. *J Clin Endocrinol Metab.* 2012;**97**(8):E1540–E1549.
14. Cools M, Pleskacova J, Stoop H, Hoebeke P, Van Laecke E, Drop SL, Lebl J, Oosterhuis JW, Looijenga LH, Wolffenbuttel KP; Mosaicism Collaborative Group. Gonadal pathology and tumor risk in relation to clinical characteristics in patients with 45,X/46,XY mosaicism. *J Clin Endocrinol Metab.* 2011;**96**(7):E1171–E1180.
15. Lee PA, Mazur T, Danish R, Amrhein J, Blizzard RM, Money J, Migeon CJ. Micropenis. I. Criteria, etiologies and classification. *Johns Hopkins Med J.* 1980;**146**(4):156–163.
16. Cheng PK, Chanoine JP. Should the definition of micropenis vary according to ethnicity? *Horm Res.* 2001;**55**(6):278–281.
17. Sathyanarayana S, Grady R, Redmon JB, Ivicek K, Barrett E, Janssen S, Nguyen R, Swan SH; TIDES Study Team. Anogenital distance and penile width measurements in The Infant Development and the Environment Study (TIDES): methods and predictors. *J Pediatr Urol.* 2015;**11**(2):76.e1–76.e6.
18. Thankamony A, Ong KK, Dunger DB, Acerini CL, Hughes IA. Anogenital distance from birth to 2 years: a population study. *Environ Health Perspect.* 2009;**117**(11):1786–1790.
19. Callegari C, Everett S, Ross M, Brasel JA. Anogenital ratio: measure of fetal virilization in premature and full-term newborn infants. *J Pediatr.* 1987;**111**(2):240–243.
20. Mondal R, Chatterjee K, Samanta M, Hazra A, Ray S, Sabui TK, Banerjee B, Das S, Roychowdhury D, Biswas R. Clitoral length and anogenital ratio in indian newborn girls. *Indian Pediatr.* 2016;**53**(4):299–303.
21. PRADER A. [Genital findings in the female pseudohermaphroditism of the congenital adrenogenital syndrome; morphology, frequency, development and heredity of the different genital forms]. *Helv Paediatr Acta.* 1954;**9**(3):231–248.
22. Ahmed SF, Khwaja O, Hughes IA. The role of a clinical score in the assessment of ambiguous genitalia. *BJU Int.* 2000;**85**(1):120–124.
23. van der Straaten S, Cools M. Supplemental figures and tables of European multicenter validation study of the External Genitalia Score. v2. *Dryad Digital Repository*; 2019. https://doi.org/10.5061/dryad.sbcc2fr23. Date of deposited October 27, 2019.
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;**15**(2):155–163.
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;**1**(8476):307–310.
26. Toppari J, Virtanen HE, Main KM, Skakkebaek NE. Cryptorchidism and hypospadias as a sign of testicular dysgenesis syndrome (TDS): environmental connection. *Birth Defects Res A Clin Mol Teratol.* 2010;**88**(10):910–919.
27. Eisenberg ML, Hsieh MH, Walters RC, Krasnow R, Lipshultz LI. The relationship between anogenital distance, fatherhood, and fertility in adult men. *Plos One.* 2011;**6**(5):e18973.
28. Eisenberg ML, Jensen TK, Walters RC, Skakkebaek NE, Lipshultz LI. The relationship between anogenital distance and reproductive hormone levels in adult men. *J Urol.* 2012;**187**(2):594–598.
29. Bangalore Krishna K, Houk CP, Lee PA. Pragmatic approach to intersex, including genital ambiguity, in the newborn. *Semin Perinatol.* 2017;**41**(4):244–251.