CLINICAL RESEARCH ARTICLE

Endocrine-sensitive physical endpoints in newborns: ranges and predictors

Rachana Shah [1,2], Belal Alshaikh[3,4,5], Joan I. Schall[3], Andrea Kelly[1,2], Eileen Ford[3], Babette S. Zemel[2,3], David M. Umbach[6], Margaret Adgent[6] and Virginia A. Stallings[2,3]

BACKGROUND: In neonates, endocrine-sensitive physical endpoints, including breast and reproductive tissues, may reflect effects of fetal environmental exposure. Studies using standardized measurement techniques that describe demographic and clinical variability in these endpoints are lacking.
METHODS: Three hundred and eighty-eight healthy term newborns <3 days old were evaluated, 69% African American and 25% White. Measures included breast bud diameter, anogenital distance (AGD), stretched penile length (SPL), and testicular volume (TV).
RESULTS: Breast buds were larger in females than males bilaterally (right: 13.0 ± 4.0 vs. 12.0 ± 4.0 mm, $p = 0.008$; left: 13.0 ± 4.0 vs. 11.0 ± 3.0 mm, $p < 0.001$). Breast bud size correlated positively with gestational age (regression coefficient $= 0.46 \pm 0.12$ mm, $p < 0.001$) and weight Z-score (0.59 ± 0.24 mm, $p = 0.02$), and negatively with White race (-1.00 ± 0.30 mm, $p = 0.001$). AGD was longer in males (scrotum-to-anus) than females (fourchette-to-anus) (21.0 ± 4.0 vs. 13.0 ± 2.0 mm, $p < 0.001$) and did not differ by race. SPL was shorter in White infants (35.0 ± 5.0 vs. 36.0 ± 5.0 mm, $p = 0.04$). Median TV was 0.5 cm^3, and larger in White males (odds ratio 1.71, 95% confidence interval: 1.02–2.88)
CONCLUSIONS: This study provides a range of physical measurements of endocrine-sensitive tissues in healthy infants from the United States, and the associations with demographic and clinical characteristics.

IMPACT:

- This study reports physical measurements for endocrine-sensitive endpoints in healthy US newborns, including breast buds, AGD, SPL, and TV.
- Associations of measurements to demographic and clinical factors (including race, gestational age, and newborn length and weight) are presented.
- Contemporary ranges and identification of predictive factors will support further study on effects of pre- and postnatal exposures to endocrine-sensitive tissues in the infant.

INTRODUCTION

Breast and fetal reproductive tissues appear to be susceptible to endocrine-disrupting effects of environmental chemicals.[1–3] Increasing awareness of the harmful effects of these chemicals on the fetus prompted the international workshop "Assessing Endocrine-related Endpoints." The report emphasized both the need for reference data on endocrine-sensitive endpoints in the first few years of life and the use standardized methods and adjustments for sex, weight, age, and length.[4] Endocrine-sensitive endpoints in newborns and infants include tissues that can be evaluated by physical examination, as well as tissue and organs requiring imaging or blood tests to fully assess. Tissues available for physical examination include breasts and external genitalia in both male and female infants.

Neonatal breast tissue, or breast bud, consists of parenchymal and stromal elements. The parenchyma forms a system of

branching ducts with hormone-sensitive acini, lacteals, and alveolar cells, whereas stroma consists mainly of adipose tissue. While early stages of mammary gland embryogenesis are hormone-independent, hormones and other regulatory factors are important for breast development after the first trimester.[5] In mice and humans, pre- or postnatal exposure to endocrine disruptors has been reported to alter mammary gland development.[6,7]

Anogenital distance (AGD) has been long recognized as a measure of prenatal androgen exposure in animal models.[8] Differentiation of human fetal genitalia requires appropriate androgen production and response from targeted tissues. Increased fetal androgen concentrations are the cause for longer AGD in males when compared to females. In contrast, fetal exposure to anti-androgens, such as phthalates, results in a

[1]Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3]Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [4]Department of Pediatrics, Alberta Children's Hospital, Calgary, AB, USA; [5]Community Health Sciences, University of Calgary, Calgary, AB, USA and [6]National Institute of Environmental Health Sciences, National Institutes of Health, Philadelphia, PA, USA
Correspondence: Rachana Shah (shahr@email.chop.edu)

Endocrine-sensitive physical endpoints in newborns: ranges and predictors
R Shah et al.

661

reduction of AGD in both sexes.[1,9,10] The effects of environmental exposures may differ in White and African-American infants.[11]

Experimental studies have recognized two critical windows for fetal development of male genitalia:[12,13] a narrow "masculinization programing window" at which low androgen production leads to hypospadias and undescended testis; and a wide "growth window" when blocking androgen action results in micropenis. Stretched penile length (SPL) at birth has been studied in various racial, geographic, and ethnic populations,[14–17] although little has been reported on SPL association with other maternal and fetal characteristics. Evidence suggests that AGD is established early in fetal development and, as opposed to SPL, is unaffected by androgen concentrations later in gestation;[18] thus, these two measurements serve as markers of disruption during different stages of fetal reproductive development.

The fetal testis produces androgen in addition to its being a target for other hormones. Many reports have assessed testicular volume (TV) during childhood and puberty, but data on testicular size in newborns are scarce.[19] Our group recently published ultrasound-based measurements of TV in the Infant Feeding and Early Development (IFED) cohort along with ultrasound measurements of breast buds in both sexes and ovaries and uterus in females.[20] The current report focused on measurements possible without ultrasonography.

Previous reports on tissue and organ measurements in human neonates are limited by lack of race-specific reference ranges, use of non-standard techniques, and concerns for measurement reliability. Having readily available ranges for these endocrine-sensitive physical endpoints in the newborn is crucial for determination of when to consider further evaluation for endocrine disorders seen in newborns and for studying the effects of environmental exposure in the fetus. Understanding how race, maternal age, and other factors influence these measurements is essential in interpreting studies of environmental exposures that may differ between populations and lead to bias.

Data for the current study were collected as part of a prospective study to evaluate estrogen-sensitive endpoints in relation to the use of soy-protein-based formula compared to cow-milk-based formula and breast milk, in healthy term neonates born to healthy mothers.[2] The purpose of this report was to provide ranges for these endocrine-sensitive endpoints among newborns (<3 days old) in this cohort, and to explore maternal and neonatal characteristics that may affect these endpoints.

MATERIALS AND METHODS
Population
The IFED study was conducted at the Children's Hospital of Philadelphia (CHOP) in Philadelphia, Pennsylvania, USA and was designed to compare endocrine-sensitive endpoints in infants fed soy formula as compared to cow-milk formula or breast milk. Healthy mother and newborn infant pairs were enrolled from the postpartum and well-baby nurseries from seven hospitals in the Philadelphia region. The institutional review boards at CHOP, Virtua Hospitals, Abington Memorial Hospital, and the National Institute of Environmental Health Sciences approved the protocol.

Mothers were fluent in the English language and 18 years of age or older at enrollment for the study. Eligible mothers had decided before contact by study staff to exclusively feed their infants soy formula, cow-milk formula, or breast milk from birth. Exclusion criteria for the mothers included endocrine disorders such as diabetes (type I, type II, and gestational diabetes), thyroid conditions, polycystic ovarian syndrome, Cushing's syndrome, congenital adrenal hyperplasia, and Addison's disease. In addition, mothers taking steroids (with the exception of inhaled steroids) or immunosuppressant medication during pregnancy were

excluded. Healthy singleton infants born between 37- and 42-week gestation as documented in the medical record and weighing between 2500 and 4500 g at birth were eligible. Exclusion criteria for infants included congenital malformations, chromosomal anomalies, significant illness affecting feeding, growth, or development, a sibling previously enrolled in IFED, and males with one or both non-palpable testes, or urogenital anomalies, including hypospadias or chordee. All mother/infant pairs were enrolled and completed the study visit within 72 h of infant birth.

Race classification for the mother and infant was by maternal report. Mothers chose from six categories: American Indian or Alaskan Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, or other. Mothers also stated whether she considered herself or her infant to be Hispanic or Latina. Due to small numbers, all categories except "Black" and "White" were grouped into "Other" for demographic analysis. For analyses of correlations, we further combined "Other" and "Black" categories into "White" and "non-White."

Measurements
Breast buds were measured twice and averaged using beads custom made for the IFED study in 2 mm diameter increments from 4 to 28 mm (Precision Machine and Tool Co., Beltsville, MD). Using the index and middle finger, study staff trained in this procedure palpated the areola to locate the breast bud while holding the beads in the other hand next to the breast bud to match the size of the bud with one of the beads. A third measurement was performed if the first two measurements differed by more than 2 mm (one bead length), and the three values were averaged.

AGDs were measured using a digital caliper (Wiha digiMax, Swiss Precision, Monticello, MO) to the nearest 0.1 mm as depicted in Fig. 1. Measurement required two observers—one to position the infant and another to take the measurement. The infant was placed in the dorsal decubitus position on a table with one observer at each end. A blanket roll was placed under the infant's buttocks to optimize visualization and measurement. The caliper was positioned so that the measuring surfaces were immediately adjacent to the anus and the genitalia (fourchette [f-a] or clitoris [c-a] in females, or scrotum [s-a], anterior [ap-a], or posterior penis base [pp-a] in males). The fixed measuring edge of the caliper was placed at the center of the anus and the moveable edge at the genitalia landmark. Two independent measurements of the distance from the anus to the genitalia were obtained and averaged. If the second measurement varied from the first by more than 2 mm, a third measurement was taken and all three were averaged. Based on existing literature,[21,22] an expected range was set for selected measures, and values outside these ranges warranted an additional measurement for verification. Out-of-range measures included AGD f-a <3.5 mm or >15.0 mm, and AGD s-a <7.5 mm or >33.0 mm. There were no expected lower or upper ranges set for AGD c-a, pp-a, and ap-a.
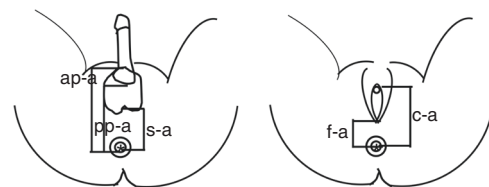

Fig. 1 Schematic diagram of anogenital distances by sex. ap-a: Anogenital distance between anterior penis and anus, 2: pp-a: anogenital distance between posterior penis and anus, s-a: anogenital distance between scrotum and anus, f-a: anogenital distance between fourchette and anus, and c-a: anogenital distance between clitoris and anus.

To measure SPL, the infant was placed in the dorsal decubitus position on the exam table. One end of a tongue depressor with a measuring strip (with 0.01 cm gradation) attached was placed on the pubic symphysis above and immediately adjacent to the penis while firmly applying pressure against the bone. The tongue depressor and penis were held at a 90° angle away from the infant's body. The penis was held and gently stretched while, at the same time, lowering the foreskin until the urethral meatus was visible. The distance from the lower end of the tongue depressor to the tip of the penis was measured to the nearest 0.1 cm. The protocol included two independent measurements, with a third measurement obtained if the initial measurements differed by >0.2 cm or if the measurement was below the 3rd percentile (<2 cm) or above the 97th percentile (>4.5 cm) as previously reported for term infants.[23]

Testes were measured using Praeder orchidometer beads in 0.5 cm^3 volume increments (0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 cm^3) (Precision Machine and Tool Company, Beltsville, MD).[24] Each testis was measured twice and averaged. If the two measures differed by >0.5 cm^3 (one bead), a third measurement was obtained and the three values were averaged. While study design excluded infants with non-descended or non-palpable testes, those with high scrotal or inguinal testes were included in the general analysis.

Infant length was measured to the nearest 0.1 cm using a Harpenden (Holtain Limited, Crymch, UK) infantometer. A fixable plastic measuring tape with 0.01 cm gradations was used to obtain head circumference (HC) (McCoy Health Science Supply, Maryland Heights, MO). Weight was measured with clothing and diaper removed, using a digital infant scale (Scaletronix, White Plains, NY) to the nearest 0.01 kg. All measures were completed twice. A third measurement was obtained if initial measurements varied by more than 1.0 cm for length, 0.01 kg for weight, and 0.5 cm for HC.

A pediatric endocrinologist (A.K.) trained the study staff in measurement of AGD, and testes and breast bud measurement technique, and anthropometry experts (B.S.Z. and J.I.S.) trained the staff in growth measurement technique. Staff certification in measurement technique consisted of observation as well as duplicate measurements within accepted variance ranges in comparison to experts at time intervals during the study.

Statistical analysis
Descriptive statistics, including means and standard deviations, or medians, percentiles, and ranges were used. Proportions and frequencies were used for categorical variables. Cross-sectional comparisons were performed for continuous variables using unpaired Student's t test for breast bud, SPL, and AGD and Mann–Whitney U test for TV. We investigated associations of breast bud diameter, AGD, or penile length with clinical characteristics using linear regression analysis. Potential covariates identified in previous studies, including maternal age, gestational age, race, birth size including weight, length, and HC,[22,25–27] were included in all regression models.

Because of skewness and limited variation, testicular size was grouped into three categories of 0.5 cm^3 (68.0%), 1.0 cm^3 (25.0%), and ≥1.5 cm^3 (7.0%). We used ordinal logistic regression with baseline categories to evaluate their correlation with clinical characteristics.

We assessed intra-observer reliability using interclass correlation coefficient based on the first and second measurements on each subject (third measurements, if present, were ignored for these analyses). An intra-class coefficient near 1 indicates that the variance attributable to measurement error is small relative to the total variance (measurement variance plus between-subject variance) of the endpoint. Repeat measures between expert and staff were performed throughout the IFED longitudinal study, but few were performed exclusively on newborns and thus not reported here.

All statistical analyses were performed using STATA 12.0 (College Station, TX). Results were considered significant at p value < 0.05.

RESULTS
A total of 405 mother–newborn pairs were enrolled; 388 subjects (197 males and 191 females) were included in the present report. The 17 subjects were excluded due to consent withdrawal ($n = 5$), birth weight <2500 g ($n = 5$), undescended testes in males ($n = 3$), missing anthropometric measurements at birth ($n = 3$), and exceeding the 72-h window before the measurements were acquired ($n = 1$). Baseline characteristics of all subjects in the study are summarized in Table 1. The sample was predominantly African-American (68.8%) and White (24.5%) ancestry; the remaining 6.7% of subjects were of Other or mixed ancestry. Weight, length, and HC were lower in females, and all mean Z-scores slightly shifted towards negative values.

Reliability
Differences of >2 mm between the first and second measurement in AGD s-a, AGD pp-a, AGD ap-a, AGD f-a, and AGD c-a were observed in 6.6%, 3.0%, 1.5%, 3.1%, and 4.1% of subjects, respectively. Out-of-reference range values occurred in 1.5% of males and 1.6% of females for AGD s-a and AGD f-a, respectively.

For breast bud diameter, intra-rater reliability was 0.96 based on duplicate measurements on both right and left breasts for females

Table 1. Maternal and neonatal characteristics of all subjects at birth and by sex.

	All, $n = 388$	Males, $n = 197$	Females, $n = 191$
Maternal age (years)	26.2 ± 5.6	26.6 ± 5.8	25.8 ± 5.4
Maternal education, number (%)			
Less than high school	75 (19.3)	38 (19.3)	37 (19.4)
High school	146 (37.6)	77 (39.1)	69 (36.1)
Some college	103 (26.6)	53 (27.0)	50 (26.2)
College degree	64 (16.5)	29 (14.7)	35 (18.3)
Race (%)			
White	95 (24.5)	49 (24.9)	46 (24.1)
African American	267 (68.8)	135 (68.5)	132 (69.1)
Other	26 (6.7)	13 (6.6)	13 (6.8)
Completed week gestation, number (%)			
37	31 (8.0)	14 (7.1)	17 (8.9)
38	71 (18.3)	34 (17.3)	37 (19.4)
39	145 (37.4)	72 (36.5)	73 (38.2)
40	94 (24.2)	51 (25.9)	43 (22.5)
41	47 (12.0)	26 (13.2)	21 (11.0)
Anthropometric measures			
Weight (kg)	3.21 ± 0.39	3.26 ± 0.39	3.15 ± 0.38
Weight Z-score	−0.15 ± 0.82	−0.17 ± 0.82	−0.14 ± 0.82
Length (cm)	49.0 ± 1.9	49.4 ± 1.9	48.7 ± 1.9
Length Z-score	−0.36 ± 1.00	−0.36 ± 1.00	−0.36 ± 1.00
Head circumference (cm)	33.9 ± 1.2	34.2 ± 1.7	33.7 ± 1.2
Head circumference Z-score	−0.27 ± 0.96	−0.30 ± 0.92	−0.24 ± 1.00

Results as mean ± SD.

Endocrine-sensitive physical endpoints in newborns: ranges and predictors
R Shah et al.

663

and males combined. In males, intra-rater reliabilities for SPL, AGD s-a, AGD pp-a, and AGD ap-a were 0.94, 0.91, 0.94 and 0.93, respectively. In females, AGD f-a and AGD c-a had intra-rater reliabilities of 0.85 and 0.88, respectively.

Breast bud, AGD, SPL, and TV measurements
Mean breast bud diameters were larger in females than in males on both sides (right: 13 ± 4 vs. 12 ± 4 mm, $p = 0.008$; left: 13 ± 4 vs. 11 ± 3 mm, $p < 0.001$). Right breast bud was larger than left breast bud on average in males ($p = 0.001$), but no right-left difference was evident in females ($p = 0.72$). In both sexes, African-American infants had larger left breast bud on average than White infants (12 ± 4 vs. 11 ± 3 mm, $p = 0.04$), while the difference in right breast bud was smaller and not significant (12 ± 4 vs. 12 ± 3 mm, $p = 0.12$).

All AGD measurements by sex were similar between White and African-American newborns. Using AGD between anus and the base of genitalia, AGD s-a in males was longer than AGD f-a in females (13 ± 2 vs. 21 ± 4, $p < 0.001$). SPL was shorter in White newborns compared to African American (35 ± 5 vs. 36 ± 5 mm, $p = 0.04$). TV was grouped into three categories of 0.5 cm^3 (68.0% of subjects), 1 cm^3 (25.0%), and $\geq$1.5 cm^3 (7.0%) and did not differ by race. Difference of 0.5 cm^3 in TV between the two testicles was found in 22.1% of subjects. Table 2 describes means, standard deviations (SDs), ranges (minimum-maximum), and percentiles between 3% and 97% for breast bud and all AGD measurements by sex, and SPL and TV in males.

Association of measurements with maternal and infant characteristics
Associations of SPL and AGD with maternal age, gestational age, race, and weight and length Z-scores are summarized in Table 3. As noted above, SPL was increased in African-American infants compared to White infants. SPL had a positive association to length Z-score (β-coef. $\pm$ SE = 0.1 ± 0.1, $p = 0.02$). AGD in males was correlated positively with both weight Z-score (1.2 ± 0.6, $p = 0.04$ for ap-a and 1.6 ± 0.5, $p = 0.002$ for pp-a) and length Z-scores (0.9 ± 0.4, $p = 0.04$ for ap-a), while in females, correlated only with weight Z-score (1.4 ± 0.5, $p = 0.004$ for c-a). There was no correlation of SPL or AGD with gestational age in our cohort of term infants. SPL correlated positively with AGD ap-a in males (coef. $\pm$ SE = 0.20 ± 0.08, $p = 0.008$).

Relationship between breast bud and previous covariates are reported in Table 3 as well. Breast bud size was positively correlated with gestational age in males (0.62 ± 0.15, $p < 0.001$), and with weight Z-score in females (1.02 ± 0.21, $p < 0.001$). Breast bud diameter was negatively correlation with White race in both males and females (-0.89 ± 0.42, $p < 0.001$ and -1.20 ± 0.36, $p < 0.001$). Correlation with length Z-score differed by sex; there was positive correlation in males (0.56 ± 0.26, $p = 0.03$), but negative correlation in females (-0.43 ± 0.20, $p = 0.03$). There was no correlation of breast bud diameter with AGD, testicular size, or SPL.

Using ordinal logistic regression and compared to the 0.5 cm^3 category, large TV ($\geq$1.5 cm^3) was more likely in White newborns [odds ratio (OR) = 1.71; 95% confidence interval (CI): 1.02–2.88]

Table 2. Breast bud, stretched penile length, and anogenital distance means and percentiles (in mm) in term newborns[a] by sex.

	Mean ± SD	Range	3rd	10th	25th	Median	75th	90th	97th
Males									
Right breast bud									
White	11 ± 3	4–20	7	8	8	11	12	14	16
Non-White	12 ± 4	4–22	6	8	10	12	14	16	20
Left breast bud									
White	11 ± 3	6–18	6	8	9	10	12	14	18
Non-White	12 ± 4	4–20	6	6	9.5	12	14	16	18
Stretched penile length									
White	35 ± 5	27–43	27	28	30	35	39	41	41
Non-White	36 ± 5	24–50	27	30	33	36	40	44	46
Anogenital distance									
Scrotum-anus	21 ± 4	7–35	13	16	19	21	24	26	28
Posterior penis-anus	46 ± 5	34–61	38	40	43	46	49	52	56
Anterior penis-anus	50 ± 4	40–62	42	44	47	50	53	55	58
Ratio	0.42 ± 0.07	0.17–0.58	0.24	0.33	0.38	0.43	0.47	0.51	0.55
Females									
Right breast bud									
White	13 ± 3	4–18	6	8	10	12	15	18	18
Non-White	13 ± 4	4–26	7	8	10	12	16	17	20
Left breast bud									
White	12 ± 3	4–18	8	8	10	12	15	16	18
Non-White	13 ± 4	4–18	6	9	10	12	15	17	18
Anogenital distance									
Fourchette-anus	13 ± 2	8–22	9	10	11	13	14	16	18
Clitoris-anus	35 ± 3	26–45	28	31	33	35	37	39	42
Ratio	0.37 ± 0.07	0.24–0.60	0.26	0.29	0.33	0.37	0.41	0.47	0.50

Non-White includes African American and other. AGD ratio in males is between scrotum-anus and anterior penis-anus distances. AGD ratio in females is between fourchette-anus and clitoris-anus.
[a]Thirty-seven- to 43-week gestation with birth weight 2.5–4.5 kg.

Endocrine-sensitive physical endpoints in newborns: ranges and predictors
R Shah et al.

664

Table 3. Association of maternal and infant factors with stretched penile length, anogenital distances, and breast bud diameter.

| | Stretched penile length (males) | | Anogenital distance in males | | | | | | Anogenital distance in females | | | | Breast bud in males | | Breast bud in females | |
| | | | Ant. penis-anus | | Post. penis-anus | | Scrotum to anus | | Fourchette-anus | | Clitoris-anus | | Diameter | | Diameter | |
	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value	Coef. ± SE	P value
Maternal age (years)	−0.01 ± 0.1	0.86	0.03 ± 0.5	0.53	−0.1 ± 0.1	0.49	0.1 ± 0.1	0.27	0.04 ± 0.03	0.25	0.03 ± 0.04	0.44	−0.05 ± 0.03	0.09	0.01 ± 0.03	0.85
Gestational age (weeks)	0.5 ± 0.4	0.09	0.1 ± 0.3	0.70	0.5 ± 0.4	0.18	0.1 ± 0.3	0.86	0.2 ± 0.2	0.34	0.2 ± 0.2	0.45	**0.62 ± 0.15**	**0.001**	0.23 ± 0.14	0.09
White race	**−2.2 ± 0.8**	**0.01**	0.5 ± 0.7	0.50	0.4 ± 0.8	0.62	−0.9 ± 0.7	0.21	−0.8 ± 0.4	0.07	−0.5 ± 0.6	0.44	**−0.89 ± 0.42**	**0.03**	**−1.20 ± 0.36**	**0.001**
Weight Z-score	−0.8 ± 0.7	0.24	**1.2 ± 0.6**	**0.04**	−0.1 ± 0.7	0.92	−0.1 ± 0.6	0.91	0.6 ± 0.3	0.10	**1.4 ± 0.5**	**0.004**	−0.09 ± 0.28	0.76	**1.02 ± 0.21**	**0.001**
Length Z-score	**0.1 ± 0.1**	**0.02**	**0.9 ± 0.4**	**0.04**	**1.6 ± 0.5**	**0.002**	0.5 ± 0.5	0.23	0.2 ± 0.3	0.52	−0.2 ± 0.4	0.52	**0.56 ± 0.26**	**0.03**	**0.43 ± 0.20**	**0.03**

Bold values indicate statistical significance p < 0.05.
All endpoint measures were adjusted for the following variables: maternal age, gestational age, race, weight, length, and head circumference Z-scores.
Coef. regression coefficient, *SE* standard error, *ant. penis* anterior penis, *post. penis* posterior penis.

and in newborns with higher weight Z-scores (OR = 2.12; 95% CI: 1.35–3.34). In contrast, newborns with higher length Z-scores were less likely to have larger testes (OR = 0.62; 95% CI: 0.45–0.87).

DISCUSSION

Here we reported ranges of endocrine-sensitive physical endpoints in healthy term US newborns and demonstrated the reliability of these measurements by trained examiners. In our cohort of 388 infants, breast buds were larger in females than males bilaterally; breast bud size correlated positively with gestational age and negatively with White race (−1.00 ± 0.30 mm, $p = 0.001$). AGD was longer in males (scrotum-to-anus) than females (fourchette-to-anus) and did not differ by race. SPL was shorter in White infants. Median TV was 0.5 cm^3, and larger in White males.

Breast bud
All newborns in the study had measurable breast buds at birth. Our finding that breast bud was larger in females compared to males is in agreement with a study by Francis et al.;[28] however, mean size was 2–3 mm larger in our subjects for both sexes. This difference may be related to the difference in sample composition, as 75% of our subjects were non-White and smaller breast bud size in White infants was evident in our cohort. Previous studies have demonstrated higher testosterone, but similar estradiol and progesterone concentrations in cord blood of males compared to females, suggesting a negative relationship between testosterone and breast size.[28,29]

Breast bud size is included as a component of standard scoring systems for assessment of newborn gestational age.[30] We confirmed a positive relationship between gestational age and size of breast buds, athough other factors like race and birth size influence the breast bud size. The relationship between body weight and size of breast bud was evident in our study. Neonates born small for gestational age are expected to have smaller breast bud size related to decreased embryonic fat tissue[27] as mammary glands continue to grow prenatally according to gestation in these infants.[27] These physical measurements of breast bud size data concur with ultrasound-based measurements in the same cohort.[20] The ultrasound results also showed larger measurements on both sides in females than males, right side vs. left side in males, and non-White vs. White infants. However, while physical measurements showed no difference in size between right and left breast in females, the left side was statistically larger by ultrasound measurement.

Anogenital distance
AGD was sexually dimorphic, as expected, in the present study with males having a greater AGD. This finding is in agreement with Sathyanarayana et al.[31] (in US infants) and Papadopouluo et al.[32] (in Spanish and Greek infants) with comparable AGD ranges. Race was not associated with AGD outcomes in our study, as also shown by Sathyanarayana et al.[31] Birth weight was a significant determinant of AGD (ap-a) and (c-a) in both sexes, while length Z-score was shown to predict AGD (s-a) in males only. There were no associations between AGD outcomes and maternal age, gestational age, or HC. These findings are similar to those reported by Salazar-Martinez et al.,[22] who included only term newborns. Studies that revealed association between gestational age and AGDs included late preterm infants allowing more variability in gestational age.[26,31,32] Associations between birth weight and AGD (ap-a) and (c-a) but not AGD (s-a) and (f-a) suggest that genitalia size may correlate with birth weight.

Penile length
Our study presents ranges for SPL in White and African American infants in our cohort. Assessment of SPL is used to evaluate

Endocrine-sensitive physical endpoints in newborns: ranges and predictors
R Shah et al.

665

micropenis (penile length <2.5 standard deviations (SD) below the mean, defined as 27 mm in term newborns, which can be a sign of critical endocrine conditions, such as hypopituitarism. However, recent studies on Nigerian and Turkish newborns showed that −2 SD of penile length was at 24 and 30 mm, respectively,[33,34] emphasizing the wide variations in SPL by study samples among populations and the need for population-specific normative data. Variation of penile length has been linked to genetic, endocrine, nutritional, and environmental causes.[33] Despite the similarity in the −2 SD definition of SPL by race in the present study, mean SPL was shorter among White newborns. We found that SPL had positive relationship with length at birth; this relationship with stature has been described in neonates and adults, suggesting that there may be overlapping determinants of skeletal and genital growth.[34]

Testicular size
Two-thirds of subjects had TV of 0.5 cm^3, which represented the median size in our cohort. This median was lower than the 1.1 cm^3 that reported in a small sample of 10 White newborns using similar technique.[19] Another study on Turkish newborns showed even larger testes (1.64–1.73 cm^3).[35] This discrepancy may be due to our finding that non-White newborns in our cohort tended to have smaller testes, or due to technique, genetic, and other population-based factors. Measurements by orchidometer include epididymal tissue and skin in addition to testicular tissue, which leads to overestimation of testicular size. In comparison to physical measurement, ultrasonography allows measurement of actual testicular tissue.[36] The mean sonographic volume of the right testis in our subjects was 0.27 ± 0.07 cm^3 and of the left testis was 0.25 ± 0.07 cm^3, which are half that of our physical measurements.[20] The smallest orchidometer used in the current study was 0.5 cm^3, which is twice as large as the sonographic volumes, and technicians were thus not able to record TV <0.5 cm^3. These technical issues limit the sensitivity of orchidometer-based measurements of TV in the neonatal period.

Reliability
Intra-rater reliabilities of all measurements were >0.85, indicating that <15% of the observed variation in our cohort was related to measurement error. AGD reliabilities were slightly lower in females compared to that in males. Studies on AGD in newborns have reported similar reliabilities (>0.85).[26,32]

Our data represent a predominantly White and African-American cohort in the Philadelphia area of the United States. This study oversampled participants whose mothers chose to use soy-protein-based formula as a feeding choice, and this group included more African-American infants, which may limit the generalizability of these results. Data on maternal exposures to over-the-counter medications, smoking, or other endocrine disruptors was not collected. Mean birth weight, length, and HC in our sample were marginally lower than those reported in growth charts of World Health Organization,[37] although other studies from the United States in the past decade have shown similar patterns in these anthropometric measurements.[38,39]

Our data, with measurements of endocrine-sensitive tissues in healthy newborns, and relationships to race, sex, and other characteristics, demonstrate the utility of population-specific comparisons when determining effects of environmental exposures. Our findings further highlight the need for mechanistic inquiry into how between-group differences in measurements are driven by endogenous genetic factors and/or exogenous exposures, perhaps at separate periods during fetal development. Overall, understanding the complex relationships of maternal and infant factors on endocrine-sensitive tissue development should guide the design of future studies of endocrine-disrupting effects.

CONCLUSION
Measurements of breast bud, AGD, SPL and TV of healthy term newborns were presented as means, ranges and percentiles and associations with demographic and clinical characteristics are reported. Identifying ranges for endocrine-sensitive physical endpoints, and predictive factors, in a contemporary United States sample will support future evaluation of prenatal exposure to endocrine disruptors.

AUTHOR CONTRIBUTIONS
Substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data (J.I.S., B.A., A.K., E.F., B.S.Z., D.M.U., M.A., V.A.S.); drafting the article or revising it critically for important intellectual content (R.S., B.A., V.A.S.) and final approval of the version to be published (R.S., B.A., J.I.S., A.K., E.F., B.S.Z., D.M.U., M. A., V.A.S.).

ADDITIONAL INFORMATION
Competing interests: The authors declare no competing interests.

Patient consent: All subjects underwent consent process as approved by IRB, and signed written consents were collected.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. Swan, S. H. et al. Decrease in anogenital distance among male infants with prenatal phthalate exposure. *Environ. Health Perspect.* **113**, 1056–1061 (2005).
2. Adgent, M. A. et al. A longitudinal study of estrogen-responsive tissues and hormone concentrations in infants fed soy formula. *J. Clin. Endocrinol. Metab.* **103**, 1899–1909 (2018).
3. Mammadov, E., Uncu, M. & Dalkan, C. High prenatal exposure to bisphenol a reduces anogenital distance in healthy male newborns. *J. Clin. Res. Pediatr. Endocrinol.* **10**, 25–29 (2018).
4. Arbuckle, T. E. et al. Meeting report: measuring endocrine-sensitive endpoints within the first years of life. *Environ. Health Perspect.* **116**, 948–951 (2008).
5. Robinson, G. W., Karpf, A. B. & Kratochwil, K. Regulation of mammary gland development by tissue interaction. *J. Mammary Gland Biol. Neoplasia* **4**, 9–19 (1999).
6. Wang, W., Tanaka, Y., Han, Z. & Higuchi, C. M. Proliferative response of mammary glandular tissue to formononetin. *Nutr. Cancer* **23**, 131–140 (1995).
7. Zung, A., Glaser, T., Kerem, Z. & Zadik, Z. Breast development in the first 2 years of life: an association with soy-based infant formulas. *J. Pediatr. Gastroenterol. Nutr.* **46**, 191–195 (2008).
8. Hotchkiss, A. K. et al. Prenatal testosterone exposure permanently masculinizes anogenital distance, nipple development, and reproductive tract morphology in female Sprague–Dawley rats. *Toxicol. Sci.* **96**, 335–345 (2007).
9. Zarean, M., Keikha, M., Feizi, A., Kazemitabaee, M. & Kelishadi, R. The role of exposure to phthalates in variations of anogenital distance: a systematic review and meta-analysis. *Environ. Pollut.* **247**, 172–179 (2019).

Endocrine-sensitive physical endpoints in newborns: ranges and predictors
R Shah et al.

666

10. Arbuckle, T. E. et al. Prenatal exposure to phthalates and phenols and infant endocrine-sensitive outcomes: the MIREC study. *Environ. Int.* **120**, 572–583 (2018).

11. Wenzel, A. G. et al. Influence of race on prenatal phthalate exposure and ano-genital measurements among boys and girls. *Environ. Int.* **110**, 61–70 (2018).

12. Clark, R. L. et al. Critical developmental periods for effects on male rat genitalia induced by finasteride, a 5 alpha-reductase inhibitor. *Toxicol. Appl. Pharmacol.* **119**, 34–40 (1993).

13. Welsh, M., MacLeod, D. J., Walker, M., Smith, L. B. & Sharpe, R. M. Critical androgen-sensitive periods of rat penis and clitoris development. *Int. J. Androl.* **33**, e144–e152 (2010).

14. Brandt, J. M., Allen, G. A., Haynes, J. L. & Butler, M. G. Normative standards and comparison of anthropometric data of White and Black newborn infants. *Dysmorphol. Clin. Genet.* **4**, 121–137 (1990).

15. Cheng, P. K. & Chanoine, J. P. Should the definition of micropenis vary according to ethnicity? *Horm. Res.* **55**, 278–281 (2001).

16. Feldman, K. W. & Smith, D. W. Fetal phallic growth and penile standards for newborn male infants. *J. Pediatr.* **86**, 395–398 (1975).

17. Fok, T. F. et al. Normative data of penile length for term Chinese newborns. *Biol. Neonate* **87**, 242–245 (2005).

18. Jain, V. G. et al. Anogenital distance is determined during early gestation in humans. *Hum. Reprod.* **33**, 1619–27. (2018).

19. Cassorla, F. G. et al. Testicular volume during early infancy. *J. Pediatr.* **99**, 742–743 (1981).

20. Kaplan, S. L. et al. Size of testes, ovaries, uterus and breast buds by ultrasound in healthy full-term neonates ages 0–3 days. *Pediatr. Radiol.* **46**, 1837–1847 (2016).

21. Thankamony, A., Ong, K. K., Dunger, D. B., Acerini, C. L. & Hughes, I. A. Anogenital distance from birth to 2 years: a population study. *Environ. Health Perspect.* **117**, 1786–1790 (2009).

22. Salazar-Martinez, E., Romano-Riquer, P., Yanez-Marquez, E., Longnecker, M. P. & Hernandez-Avila, M. Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. *Environ. Health* **3**, 8 (2004).

23. Shonfeld, W. A. Primary and secondary sexual characteristics: study of their development in males from birth to maturity, with biometric study of penis and testis. *Am. J. Dis. Child.* **65**, 535–549 (1943).

24. Prader, A. Testicular size: assessment and clinical importance. *Triangle* **7**, 240–243 (1966).

25. Barrett, E. S., Parlett, L. E., Redmon, J. B. & Swan, S. H. Evidence for sexually dimorphic associations between maternal characteristics and anogenital distance, a marker of reproductive development. *Am. J. Epidemiol.* **179**, 57–66 (2014).

26. Romano-Riquer, S. P., Hernandez-Avila, M., Gladen, B. C., Cupul-Uicab, L. A. & Longnecker, M. P. Reliability and determinants of anogenital distance and penis dimensions in male newborns from Chiapas, Mexico. *Paediatr. Perinat. Epidemiol.* **21**, 219–228 (2007).

27. McKiernan, J. F. & Hull, D. Breast development in the newborn. *Arch. Dis. Child.* **56**, 525–529 (1981).

28. Francis, G. L., Hoffman, W. H., Gala, R. R., McPherson, J. C. 3rd & Zadinsky, J. A relationship between neonatal breast size and cord blood testosterone level. *Ann. Clin. Lab. Sci.* **20**, 239–244 (1990).

29. Barry, J. A., Hardiman, P. J., Siddiqui, M. R. & Thomas, M. Meta-analysis of sex difference in testosterone levels in umbilical cord blood. *J. Obstet. Gynaecol.* **31**, 697–702 (2011).

30. Ballard, J. L. et al. New Ballard Score, expanded to include extremely premature infants. *J. Pediatr.* **119**, 417–423 (1991).

31. Sathyanarayana, S., Beard, L., Zhou, C. & Grady, R. Measurement and correlates of ano-genital distance in healthy, newborn infants. *Int. J. Androl.* **33**, 317–323 (2010).

32. Papadopoulou, E. et al. Anogenital distances in newborns and children from Spain and Greece: predictors, tracking and reliability. *Paediatr. Perinat. Epidemiol.* **27**, 89–99 (2013).

33. Jarrett, O. O., Ayoola, O. O., Jonsson, B., Albertsson-Wikland, K. & Ritzen, E. M. Penile size in healthy Nigerian newborns: country-based reference values and international comparisons. *Acta Paediatr.* **103**, 442–446 (2014).

34. Kutlu, A. O. Normative data for penile length in Turkish newborns. *J. Clin. Res Pediatr. Endocrinol.* **2**, 107–110 (2010).

35. Semiz, S., Kucuktasci, K., Zencir, M. & Sevinc, O. One-year follow-up of penis and testis sizes of healthy Turkish male newborns. *Turk. J. Pediatr.* **53**, 661–665 (2011).

36. Rivkees, S. A., Hall, D. A., Boepple, P. A. & Crawford, J. D. Accuracy and reproducibility of clinical measures of testicular volume. *J. Pediatr.* **110**, 914–917 (1987).

37. WHO. *WHO Child Growth Standards* (WHO, 2009).

38. Morisaki, N., Esplin, M. S., Varner, M. W., Henry, E. & Oken, E. Declines in birth weight and fetal growth independent of gestational length. *Obstet. Gynecol.* **121**, 51–58 (2013).

39. Donahue, S. M., Kleinman, K. P., Gillman, M. W. & Oken, E. Trends in birth weight and gestational length among singleton term births in the United States: 1990–2005. *Obstet. Gynecol.* **115**(Part 1), 357–364 (2010).