

Predictive value and reference ranges of anogenital distance for determining fetal gender in the first trimester

A retrospective cohort study

Abdulrahman M. Alfuraih, MS, PhD, Samiah A. Alotaiby, BS, MS, Mohammed J. Alsaadi, MS, PhD, Hanifa A. Bukhari, MD, Ali M. Aldhebaib, MS, PhD, Rafat S. Mohtasib, MS, PhD.

ABSTRACT

الأهداف: تقييم قدرة المسافة الشرجية التناسلية (AGD) في التنبؤ بجنس الجنين باستخدام الموجات فوق الصوتية في الأجنة من جنسية سعودية خلال الأشهر الثلاثة الأولى وتوفير المعويات المرجعية الطبيعية.

المنهجية: تم عمل هذه الدراسة كدراسة جماعية بأثر رجعي في مستشفى الملك فيصل التخصصي ومركز الأبحاث في الرياض في الفترة ما بين نوفمبر 2020 ومايو 2021م. تم جمع فحوصات الموجات فوق الصوتية لـ 313 حالة حمل بين 11 إلى 13+6 أسابيع من الحمل قبل تأكيد الجنس لاحقاً عند الولادة. تم قياس AGD من قبل قارئين على أنه المسافة من القاعدة السفلية للحدبية التناسلية إلى الردف. تم استخدام الانحدار اللوجستي ذي الحدين ومنحنيات خصائص تشغيل المستقبل (ROC) لتقييم أداء AGD في تنبؤ الجنس.

النتائج: كان هناك فرق كبير في متوسط AGD بين الأجنة الإناث (5.92 م [6.70، 6.14]) والذكور (6.80 م [7.00، 6.61]) بحوالي 15% ($p < 0.001$) مرتبط بشكل كبير مع عمر الحمل ($r = 0.573$; $p < 0.001$) ومع طول الجنين (CRL) ($r = 0.562$; $p < 0.001$). أشار الانحدار اللوجستي بأن AGD كان متنبئاً مؤثراً للجنس ($p > 0.001$). ومع ذلك، أظهر تحليل ROC أن الدقة الكلية كانت منخفضة عند 68% ($p = 0.001$) و 70% ($p < 0.001$) و 64% ($p = 0.017$) لأعمار 11 و 12 و 13 أسبوعاً من الحمل على التوالي. كان AGD لدى الحوامل السعوديات أطول مما ورد في الأدبيات من الأجناس الأخرى. أظهرت قابلية استنساخ القارئ الداخلي اتفاقاً جوهرياً عند 89.6% [95% CI 0.845-0.930].

الخلاصة: كان تقييم AGD باستخدام الموجات فوق الصوتية لـ AGD ممكناً وموثوقاً به خلال الثلث الأول من الحمل. كان هنالك فرق في طول AGD بين الجنسين من الذكور والإناث والفرق لم يسفر عن دقة تنبؤية عالية. يجب أن تأخذ الأبحاث المستقبلية في الاعتبار العوامل العرقية عند تقييم AGD.

Objectives: To assess the usefulness of sonographically measured anogenital distance (AGD) in predicting fetal gender in Saudi fetuses during the first trimester and to provide normal reference centiles for AGD.

Methods: A retrospective cohort study was conducted at King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia between November 2020 and May 2021. The ultrasound scans of 313 singleton pregnancies between 11–13 plus 6 gestational weeks and their gender-at-birth were collected. Anogenital distance was measured from the inferior base of the genital tubercle to the rump. Binominal logistic regression and receiver operating

characteristic curves (ROC) evaluated the predictive performance of AGD for determining fetal gender.

Results: There was a significant difference of approximately 15% in mean AGD between female (5.92 mm [95% CI= 6.70, 6.14]) and male (6.80 mm [95% CI= 6.61,7.00]) fetuses ($p < 0.001$). Anogenital distance significantly correlated with gestational age ($r = 0.573$, $p < 0.001$) and crown-rump length ($r = 0.562$, $p < 0.001$). The logistic regression determined AGD as a significant predictor of fetal gender ($p < 0.001$). However, ROC analysis showed that overall accuracies were low at 68% ($p = 0.001$) for 11 weeks, 70% ($p < 0.001$) for 12 weeks, and 64% ($p = 0.017$), and for 13 weeks. The average AGD of our Saudi cohort was longer than what the literature reported from other populations.

Conclusion: The first-trimester ultrasound evaluation of AGD was feasible and reliable. It showed a difference between the genders but did not yield high predictive accuracy. Future research should consider racial factors when evaluating AGD.

Keywords: anogenital distance, fetal gender, first trimester, ultrasound

Saudi Med J 2021; Vol. 42 (10): 1057-1064
doi: 10.15537/smj.2021.42.10.20210468

From the Radiology and Medical Imaging Department (Alfuraih, Alsaadi) College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Kharj; from the Obstetrics and Gynecology Department (Alotaiby, Bukhari), from the Department of Biomedical Physics, Molecular and Functional Imaging (Mohtasib), King Faisal Specialist Hospital and Research Center; and from the Radiologic Sciences Department (Aldhebaib), College of Applied Medical Sciences, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Kingdom of Saudi Arabia.

Received 13th June 2021. Accepted 1st September 2021.

Address correspondence and reprint request to: Dr. Abdulrahman M. Alfuraih, Radiology and Medical Imaging Department, College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Al Kharj, Kingdom of Saudi Arabia. E-mail: a.alfuraih@psau.edu.sa
ORCID ID: <https://orcid.org/0000-0002-4655-7248>

Fetal gender identification in early pregnancy is of great interest to expecting parents. However, physicians also take interest in gender identification for fetuses at higher risks of inheriting gender-dependent genetic disorders. Several X-linked recessive inherited diseases are male-specific, while adrenal congenital hyperplasia can affect female fetuses. Hence, early determination of fetal gender could be vital for early diagnosis and interventions.

Early gender identification during the first trimester can be achieved with sonography and genetic testing. Definitive results can be achieved with the latter by sampling the chorionic villus under ultrasonography guidance.¹ However, it is an invasive procedure with an associated 1.46 relative risk of pregnancy loss before 28 weeks.² Alternatively, analysis of cell-free fetal DNA is non-invasive but rather expensive and limited in availability.³ In contrast, ultrasonography can be used non-invasively to determine fetal gender with great accuracy in the second and third trimesters by visual identification of the penis or labia majora and minora.⁴ Ultrasonography can also be used to determine gender by measuring the genital tubercle angle, with 100% sensitivity at 13 weeks but lower sensitivity between 11 and 12 weeks.^{5,6}

Anogenital distance (AGD) is a recently introduced sonographic marker of fetal gender.⁷⁻¹¹ It is based on measuring the distance between the anus and the base of the genital tubercle in the perineal region.¹² Anogenital distance is sexually dimorphic since its length is dependent on hormonal levels. Hence, AGD in male fetuses can be longer than that in female fetuses.^{12,13} In one study, prenatal AGD differences between male and female fetuses were found to be maintained across their life spans.¹² Experimental preclinical studies suggested that androgen exposure affects AGD during masculinization.¹⁴ In humans, sonographically measured AGD successfully dichotomized male and female fetuses.^{7-11,15} In these studies, AGD measurement was a safe, non-invasive, and relatively cheap method of early gender identification. However, AGD measurement lacks the required validation in large samples from different populations. Recent studies reported significantly different AGD values between races, recommending that population-specific

normative values are needed for accurate clinical assessment.^{9,10} These factors support the rationale for conducting this study.

The primary objective of this study was to evaluate the accuracy of AGD measurement via ultrasound for the early determination of fetal gender during the first trimester. Secondary objectives included evaluating the inter-rater agreement in AGD measurement and identifying population-specific normative AGD values for each gender.

Methods. This study was conducted as a retrospective cohort study at King Faisal Specialist Hospital and Research Center (KFSHRC), Riyadh, Saudi Arabia between November 2020 and May 2021. The KFSHRC Research Ethics Committee approved the study and granted a waiver of informed consent (approval memo EC ref: C380/308/42). The local records of obstetric ultrasound scans performed between January 2015 and January 2019 were screened consecutively to identify eligible cases based on the inclusion criteria: i) singleton pregnancies from 11 weeks to 13 weeks plus 6 days of gestation; ii) acquisition of the ultrasound image transabdominally in the mid-sagittal plane with the fetus lying in a natural position (neck and spine are neither hyper-flexed nor hyper-extended); and iii) Saudi nationality. The exclusion criteria were as follows: i) cases where the rump and genital tubercle were not captured clearly; and ii) cases where information on the gender at birth was not available (such as, delivery occurred at a different hospital)

All scans were performed by accredited sonographers using 2 calibrated ultrasound systems (Philips EpiQ 7 and Philips Affiniti 70). An accredited obstetric sonographer screened the database and consecutively recruited eligible cases until the required sample size was attained. The sonographer exported a single mid-sagittal plane image for each case. Next, the definitive gender of the fetus after birth was documented. Data on additional variables were collected, including gestational age, maternal age, use of assisted reproduction technology, history of diabetes or polycystic ovarian syndrome, and birth weight. Sonographic information included fetal heart rate and crown-rump length (CRL).

The images for all eligible cases were extracted in Digital Imaging and Communications in Medicine (DICOM) format. Two experienced sonographers (AMA and MJA) measured the AGD to assess its inter-reader reproducibility using the OSIRIX DICOM viewer. A proximal caliper was placed at the inferior base of the genital tubercle and a distal caliper was placed at the most prominent rump location similar to the distal

Disclosure. Authors have no conflict of interests, and the work was not supported or funded by any drug company.

CRL measurement.^{7,9} The sonographers were blinded to all clinical information except for the ultrasound images. To standardize measurement acquisition, the sonographers practiced on a random subset of cases, and then reviewed the technique for measuring AGD.

Based on previously published data,⁷ to detect a minimum AGD difference of 10% on independent sample t-test with alpha level of 0.05 and power (1-β) of 0.95, the minimum sample size is 250, assuming an expected even allocation ratio between the male and female groups.

Statistical analysis. Descriptive and inferential statistics were carried out using SPSS version 27 for Mac (IBM Corp., Armonk, N.Y., USA). To compare the AGD difference between male and female fetuses, independent samples student's t-test was conducted. An alpha value of 0.05 was considered significant. To evaluate the performance of AGD in identifying fetal gender at specific gestational ages, the data were sub-grouped into the following: a) Group 1: women with gestational age of 11 weeks to 11 weeks, 6 days; b) Group 2: women with gestational age of 12 weeks to 12 weeks, 6 days; and c) Group 3: women with pregnancy age of 13 weeks to 13 weeks, 6 days.

Centiles for AGD reference ranges were calculated using a previously described method.¹⁶ Briefly, the required centiles for AGD per gestational week were calculated using centile= mean + K*SD, where mean refers to the mean AGD, SD is the standard deviation of the mean, and K is the corresponding centile of the standard distribution. For example, determination of the 25th and 75th centiles requires that K= ±0.675.

Binominal logistic regression was used to evaluate the predictive performance of AGD for gender determination. To specify valid and relevant independent variables, primary maternal and fetal characteristics were evaluated for normality, linearity in the logit (via the Box-Tidwell procedure), and multicollinearity.¹⁷ Relevant variables that satisfied all assumptions were entered into the model. Receiver operating characteristic curves were used to define optimum cut-offs for AGD in each group and to assess diagnostic performances. The AGD cutoff that yielded the highest Youden index was selected as the best AGD cutoff.

The inter-reader reproducibility was analyzed using the Bland-Altman plot and intraclass correlation coefficients (ICC). The results were interpreted as follows: 0.00-0.20, 'poor agreement'; 0.21-0.40, 'fair agreement'; 0.41-0.60, 'moderate agreement'; 0.61-0.80, 'substantial agreement'; and >0.80, 'almost perfect agreement'.¹⁸

Results. A total of 322 pregnancy cases was collected, of which 9 (2.7%) were later excluded due to poor image quality in the genital pedicle region. Hence, 313 cases were finally included in the analysis. The descriptive statistics for the main characteristics are presented in **Table 1**. Among the included cases, 23 (7.3%) pregnancies occurred following assisted reproduction technologies, 44 (14.1%) mothers had diabetes, and only 2 mothers had history of polycystic ovarian syndrome. All these 3 factors were not significantly associated with AGD ($p>0.1$). No fetus had pathological development in the genitalia at birth. When AGD was analyzed, male fetuses had longer measurements than those of female fetuses by 14.8% ($p<0.001$). The mean AGD measurements per gestational week were also significantly different, ranging from 10.7% in week 13 to 17.4% in week 11 (**Table 2**). The calculated 2.5th to 95th centile reference ranges for normal AGD per gender are listed in **Appendix 1**. An example of the AGD measurement is illustrated in **Figure 1**. A comparison between the current values to measurements from different populations is shown in **Table 3**. AGD significantly correlated with gestational age ($r=0.573$, $p<0.001$) and CRL ($r=0.562$, $p<0.001$). There were no significant correlations of AGD with maternal age ($r=0.085$, $p=0.134$), fetal heart rate ($r=-0.031$, $p=0.584$), and birth weight ($r=0.012$, $p=0.836$). The charts in **Figures 2** and **3** show the changes in AGD relative to fetal age.

In the logistic regression, AGD was the only independent variable that satisfied the model's assumptions for predicting fetal gender. The model was statistically significant: $\chi^2 = 51.7$, $p<0.001$. It explained 20% (Nagelkerke R²) of the variance in gender and

Table 1 - Main maternal and fetal characteristics

| Characteristics | Male at birth (n=172) | Female at birth (n=141) | Total (n=313) |
|------------------------|-----------------------|-------------------------|---------------|
| Maternal age (years) | 31.8 (5.1) | 31.8 (5.9) | 31.8 (5.3) |
| Gestational age (days) | 87.9 (4.7) | 87.7 (4.8) | 87.8 (4.8) |
| Birth weight (g) | 2,945 (539) | 3,000 (509) | 2,970 (526) |
| Crown-rump length (mm) | 60.0 (9.3) | 59.8 (9.5) | 59.9 (9.4) |
| Fetal heart rate (bpm) | 135.7 (10.8) | 136.0 (10.3) | 135.9 (10.6) |

Numbers in parentheses represent standard deviations.

Table 2 - Mean anogenital distance values in each gestational week.

| Week of gestation | Female | Male | Female at birth | Male at birth | Differences in mean and percentage | P-value |
|-------------------|--------|------|----------------------|----------------------|------------------------------------|---------|
| 11 weeks | 33 | 36 | 4.82 [4.50, 5.14] | 5.66 [5.34, 5.98] | 0.84 [0.39, 1.28] (17.4%) | <0.001 |
| 12 weeks | 67 | 85 | 5.77 [5.54, 6.05] | 6.74 [6.52, 6.96] | 0.97 [0.65, 1.29] (16.8%) | <0.001 |
| 13 weeks | 41 | 51 | 6.99 [6.60, 7.38] | 7.70 [7.37, 8.04] | 0.71 [0.20, 1.21] (10.2%) | 0.006 |
| Total | 141 | 182 | 5.92 [6.70, 6.14] | 6.80 [6.61, 7.00] | 0.88 [0.59, 1.17] (14.8%) | <0.001 |

Data between square brackets represent the 95% confidence intervals.

Table 3 - Mean anogenital distance values in first trimester fetuses amongst different populations.

| Population | Female (mm) | Male (mm) | Best AGD cutoff (mm) |
|----------------------|-------------|-------------|----------------------|
| Saudi | 5.92 (1.32) | 6.80 (1.26) | 6.00 |
| French ⁷ | 4.20 (0.95) | 6.00 (1.27) | 4.80 |
| Persian ⁸ | 4.30 (0.81) | 5.14 (0.85) | 4.50 |
| Turkish ⁹ | 3.60* | 5.10* | 4.80 |

Numbers in parentheses represent the standard deviations. *The standard deviations were not reported. AGD: anogenital distance

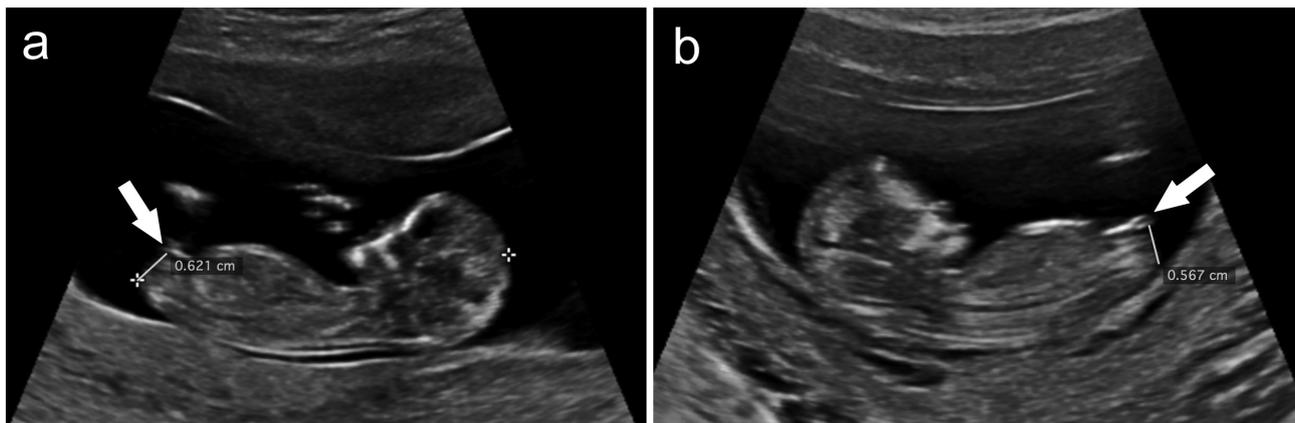


Figure 1 - Examples of anogenital distance measurements for A) male and B) female 12-week-old fetuses.

correctly classified 65% of cases. Anogenital distance was a significant predictor of fetal gender ($p < 0.001$), with larger measurements holding 2.27 times higher odds of being those for male fetuses. Reanalyzing the data for each gestational week independently yielded similar results.

When considering different cutoffs in the receiver operator characteristic analysis, the best overall cutoff

value of 6.00 mm achieved the highest Youden index of 29, with sensitivity of 69%, specificity of 60%, and area under the curve of 0.686. This can be interpreted as poor discrimination according to Hosmer et al.¹⁹ The cutoffs and their predictive performances for each gestational week are listed in **Appendix 2**. The accuracy describes the overall number of correctly identified fetal genders. Sensitivity refers to the percentage of correctly

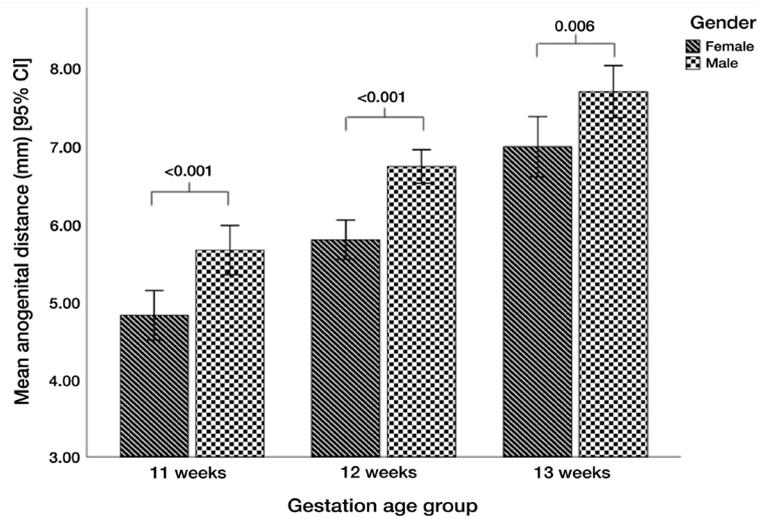


Figure 2 - A bar chart of anogenital distance (mm) against fetal gestational age by gender. CI: confidence interval

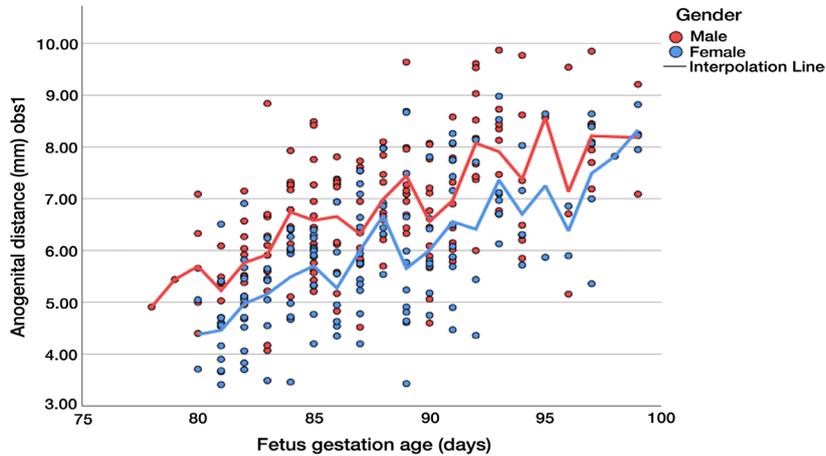


Figure 3 - A scatter plot of anogenital distance measurements with interpolated means for male and female fetuses.

identified male fetuses, whereas specificity is the percentage of correctly identified female fetuses. The positive predictive value is the chance of being male when the AGD is above or equal to the cutoff, and negative predictive value is the chance of being female when the AGD is above or equal to the cutoff. The positive likelihood ratio is the increase in the probability that the fetus is expected to be male provided that the AGD is higher than the cutoff, and the negative likelihood ratio is the increase in the probability that the fetus is expected to be female provided that the AGD is higher than the cutoff.

As for the inter-reader reproducibility of the AGD measurements, the mean difference was -0.14 mm

[95% confidence interval (CI)= -0.32, 0.49]. The interclass correlation coefficient demonstrated almost perfect agreement at 0.896 [95% CI= 0.845, 0.930].

Discussion. This study aimed to compare the AGD between male and female fetuses during the first trimester. The most important finding was that the mean AGD in male fetuses was statistically significantly longer than that in female fetuses by approximately 15%. However, the magnitude of the difference was not large enough to yield high accuracy for predicting the genders. The diagnostic metrics tended to yield higher confidence for identifying male fetuses when AGD was higher than the mentioned cutoffs (high sensitivity).

However, when AGD was small, the post-test probability for identifying a female fetus was low (poor specificity).

Our results agree with what Najdi et al⁸ reported approximately 11-week-old fetuses where the sensitivity was 70% and area under the curve was 0.748, corresponding to 72% and 0.738 in our study. Although, the majority of previous studies reported significant predictive performances in contrast with the current findings. Najdi et al⁸ reported >85% sensitivity, specificity, positive predictive value, and negative predictive value for 12- and 13-week-old fetuses. The first study that evaluated first trimester fetuses was carried out by Arfi et al.⁷ They reported sensitivity (identification of male fetuses) of 87% and specificity (identification of female fetuses) of 89%. Later, Sipahi et al⁹ reported 76% accuracy in identifying male fetus and 97% accuracy for female fetus between 11 and 13 weeks.

In older fetuses aged between 26 and 30 weeks, AGD was significantly different between the 2 genders ($p < 0.001$), but the researchers, unfortunately, did not report the predictive performance.¹⁰ Likewise, Gilboa et al¹¹ found significant AGD differences in 20- to 35-week-old fetuses, where the mean difference increased from 2.2 mm at 20 weeks to 11.8 mm at 35 weeks ($r^2 = 0.808$). It should be noted that in such advanced gestational age, AGD is measured differently. It is acquired in an axial plane from the center of the anus to the posterior convergence of the fourchette in female fetuses and to the posterior base of the scrotum in male fetuses. The similarity of this technique to the sagittal plane acquisition is unknown.

An importable and notable observation from the existing literature on AGD is the variability in normal AGD values from different populations. In first-trimester pregnancies, female and male fetuses had mean AGDs of 3.6 mm and 5.1 mm respectively in a Turkish population, 4.0 mm and 4.7 mm in a Persian population, and 4.1 and 5.90 mm in a French study.⁷⁻⁹ Such values expressed 2.2-23% differences in each gender. The Arabic Saudi population had a larger absolute mean AGD compared with values found in previous studies, ranging from 14-36% in male fetuses and from 36-48% in female fetuses. With a standardized ultrasound protocol and no demonstrable concerns over the methodologies used in the recent studies, it is evident that AGD is race-dependent. In neonatal studies, Bedouin neonates had larger AGD than Jewish neonates,²⁰ and Caucasians had larger AGD than those of native Americans and Asians.¹² This indicates that no universal AGD cutoffs can be employed across different populations.

The current study provides normative values for AGD in the Arabic Saudi population. These results are valuable for assessing genital anomalies in the Arabic ethnic group that we evaluated. A recent study demonstrated that AGD is longer in mothers who had polycystic ovarian syndrome and in those who used assisted reproduction treatments.²¹ In contrast, a study on suspected isolated abnormal male genitalia showed that fetuses with hypospadias had AGD below the fifth percentile of normal cases.²² These reports suggest that AGD can be a useful biomarker for prenatal hormonal status and genital development.

Despite the low predictive performance of AGD, it demonstrated excellent reliability. The inter-reader reproducibility was relatively similar to that reported by Aydin et al¹⁰ in second-trimester fetuses. Arfi et al⁷ found a higher ICC of 0.97 in first-trimester fetuses. Other studies also confirmed the low variability and high reproducibility of AGD.^{11,15,21} Overall, sonographic AGD is a feasible objective measure, which can be easily acquired by sonographers.

An alternative technique for gender determination in the late first trimester is the sagittal sign, which is based on calculating the angle of the genital tubercle relative to the lumbosacral skin surface. The fetus would be identified as male if the angle is >30 degree or female if <10 degree.^{5,23} However, a systematic review of this technique demonstrated a high failure rate, ranging from 7.5-40.6%.¹ Its overall accuracy was between 70% and 100%, and the accuracy was particularly poor in fetuses aged 11 and 12 weeks. Indeed, this technique requires post-processing and presents difficulty for sonographers regarding the accuracy of angle placement. Another study employed 3D ultrasound and found poor gender prediction accuracy at 56%.²⁴

Study limitations. Despite using an optimal sample size, the retrospective design of the study prevented us from controlling the mid-sagittal plane acquisition of AGD, in order to ensure accurate detection of the genital pedicles. This however resulted in a failure rate of only 2.7%. The retrospective design also meant that the images were not acquired by only one sonographer. The study also did not include fetuses that were not positioned optimally, which can be frequently encountered in routine scans. Additional contributors to fetal growth, such as maternal anxiety, socioeconomic status, and education, were not controlled due to the unavailability of retrospective data for these variables.²⁵ Future studies are encouraged to further compare the AGD results in Arabic Saudi fetuses to those of fetuses from other ethnic groups. A longitudinal examination of AGD in the Arabic population from prenatal life to

adulthood is needed in order to understand the potential associations of AGD with developmental pathologies.

In conclusion, AGD was significantly different between male and female fetuses in the first trimester. Contrary to the existing literature, the magnitude of the difference was not substantial enough to yield high predictive accuracy. Anogenital distance increased gradually with gestational age and demonstrated reliable readings. It appears that AGD varies among races, and it appears to be longer in Arabic fetuses than the reported measurements from other populations. The study reported the normative centile reference ranges for AGD.

Further longitudinal research is warranted to investigate the usefulness of AGD as an imaging biomarker for congenital genital anomalies and fetal androgen levels.

Acknowledgment. *The authors gratefully acknowledge Vicas Narange form Editage (www.editage.com) for the English language editing.*

References

- Colmant C, Morin-Surroca M, Fuchs F, Fernandez H, Senat MV. Non-invasive prenatal testing for fetal sex determination: is ultrasound still relevant? *Eur J Obstet Gynecol Reprod Biol* 2013; 171: 197-204.
- Mujezinovic F, Alfirevic Z. Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. *Obstet Gynecol* 2007; 110: 687-694.
- Costa JM, Benachi A, Gautier E. New strategy for prenatal diagnosis of X-linked disorders. *N Engl J Med* 2002; 346: 1502.
- Elejalde BR, De Elejalde M, Heitman T. Visualization of the fetal genitalia by ultrasonography: a review of the literature and analysis of its accuracy and ethical implications. *J Ultrasound Med* 1985; 4: 633-639.
- Efrat Z, Akinfenwa OO, Nicolaides KH. First-trimester determination of fetal gender by ultrasound. *Ultrasound Obstet Gynecol* 1999; 13: 305-307.
- Efrat Z, Perri T, Ramati E, Tugendreich D, Meizner I. Fetal gender assignment by first-trimester ultrasound. *Ultrasound Obstet Gynecol* 2006; 27: 619-621.
- Arfi A, Cohen J, Canlorbe G, Bendifallah S, Thomassin-Naggara I, Darai E, et al. First-trimester determination of fetal gender by ultrasound: measurement of the ano-genital distance. *Eur J Obstet Gynecol Reprod Biol* 2016; 203: 177-181.
- Najdi N, Safi F, Hashemi-Dizaji S, Sahraian G, Jand Y. First trimester determination of fetal gender by ultrasonographic measurement of anogenital distance: A cross-sectional study. *Int J Reprod Biomed* 2019; 17: 51-56.
- Sipahi M, Tokgoz VY, Alanya Tosun S. An appropriate way to predict fetal gender at first trimester: anogenital distance. *J Matern Fetal Neonatal Med* 2019; 32: 2012-2016.
- Aydin E, Holt R, Chaplin D, Hawkes R, Allison C, Hackett G, et al. Fetal anogenital distance using ultrasound. *Prenatal Diag* 2019; 39: 527-535.
- Gilboa Y, Kivilevitch Z, Oren M, Cohen YP, Katorza E, Achiron R. Anogenital distance in male and female fetuses at 20 to 35 weeks of gestation: centile charts and reference ranges. *Prenatal Diag* 2014; 34: 946-951.
- 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: 1786-1790.
- Salazar-Martinez E, Romano-Riquer P, Yanez-Marquez E, Longnecker MP, Hernandez-Avila M. Anogenital distance in human male and female newborns: a descriptive, cross-sectional study. *Environ Health* 2004; 3: 1-6.
- Macleod D, Sharpe R, Welsh M, Fiskem M, Scott H, Hutchison G, et al. Androgen action in the masculinization programming window and development of male reproductive organs. *Int J Androl* 2010; 33: 279-287.
- Natsuyama E. Sonographic determination of fetal sex from twelve weeks of gestation. *Am J Obstet Gynecol* 1984; 149: 748-757.
- Royston P, Wright EM. How to construct 'normal ranges' for fetal variables. *Ultrasound Obstet Gynecol* 1998; 11: 30-38.
- Hilbe JM. Practical guide to logistic regression. Boca Raton (FL): CRC Press; 2015. page 72.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174.
- Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons. [Updated 2013. Accessed date?]. Available from: <https://www.wiley.com/en-sa/Applied+Logistic+Regression%2C+3rd+Edition-p->
- Phillip M, De Boer C, Pilpel D, Karplus M, Sofer S. Clitoral and penile sizes of full term newborns in two different ethnic groups. *J Pediatr Endocrinol Metab* 1996; 9: 175-917.
- Perlman S, Toledano Y, Kivilevitch Z, Halevy N, Rubin E, Gilboa Y. Foetal Sonographic anogenital distance is longer in polycystic ovary syndrome mothers. *J Clin Med* 2020; 9: 2863.
- Gilboa Y, Perlman S, Kivilevitch Z, Messing B, Achiron R. Prenatal anogenital distance is shorter in fetuses with hypospadias. *J Ultrasound Med* 2017; 36: 175-182.
- Emerson DS, Felker RE, Brown DL. The sagittal sign. An early second trimester sonographic indicator of fetal gender. *JJ Ultrasound Med* 1989; 8: 293-297.
- Bogers H, Rifouna MS, Koning AH, Husen-Ebbinge M, Go AT, van der Spek PJ, et al. Accuracy of fetal sex determination in the first trimester of pregnancy using 3D virtual reality ultrasound. *J Clin Ultrasound* 2018; 46: 241-246.
- Ellman LM, Schetter CD, Hobel CJ, Chicz-DeMet A, Glynn LM, Sandman CA. Timing of fetal exposure to stress hormones: effects on newborn physical and neuromuscular maturation. *Dev Psychobiol* 2008; 50: 232-241.

Appendix 1 - Normal anogenital distances reference ranges for each gestational week: 2.5th to 97th centiles.

| Centile | 11 weeks | 12 weeks | 13 weeks |
|--------------------|----------|----------|----------|
| <i>Male</i> | | | |
| 2.5 th | 3.82 | 4.78 | 5.36 |
| 5 th | 4.12 | 5.10 | 5.73 |
| 10 th | 4.46 | 5.46 | 6.17 |
| 25 th | 5.03 | 6.07 | 6.89 |
| 50 th | 5.66 | 6.74 | 7.70 |
| 75 th | 6.29 | 7.41 | 8.51 |
| 90 th | 6.86 | 8.02 | 9.23 |
| 95 th | 7.20 | 8.38 | 9.67 |
| 97.5 th | 7.50 | 8.70 | 10.04 |
| <i>Female</i> | | | |
| 2.5 th | 3.06 | 3.84 | 4.59 |
| 5 th | 3.35 | 4.15 | 4.97 |
| 10 th | 3.67 | 4.51 | 5.42 |
| 25 th | 4.22 | 5.11 | 6.17 |
| 50 th | 4.83 | 5.78 | 7.00 |
| 75 th | 5.43 | 6.44 | 7.83 |
| 90 th | 5.98 | 7.04 | 8.58 |
| 95 th | 6.31 | 7.40 | 9.02 |
| 97.5 th | 6.59 | 7.72 | 9.41 |

All readings are in millimeters.

Appendix 2 - Predictive performance of the AGD cutoffs.

| Gestation age | AGD cutoff (mm) | Acc | Sens | Spec | PPV | NPV | +LH | -LH | AUC | P-value |
|---------------|-----------------|-----|------|------|-----|-----|------|------|-------|---------|
| 11 weeks | >5.10 | 68% | 72% | 64% | 68% | 68% | 2.00 | 0.50 | 0.738 | 0.001 |
| 12 weeks | >6.07 | 70% | 74% | 64% | 72% | 66% | 2.06 | 0.49 | 0.752 | <0.001 |
| 13 weeks | >7.17 | 64% | 69% | 59% | 67% | 60% | 1.68 | 0.59 | 0.646 | 0.017 |
| Total | >6.00 | 65% | 72% | 56% | 67% | 62% | 1.64 | 0.61 | 0.686 | <0.001 |

A positive result indicates the identification of male gender. Acc: Accuracy, Sens: sensitivity, Spec: specificity, PPV: positive predictive value, NPV: negative predictive value, +LR: positive likelihood ratio, -LR: negative likelihood ratio, and AUC: area under the curve