Published in final edited form as:

Fertil Steril. 2014 January ; 101(1): 280–287.e1. doi:10.1016/j.fertnstert.2013.10.001.

Follicle number, and not assessments of the ovarian stroma, represents the best ultrasonographic marker of polycystic ovary syndrome

Jacob P Christ, B.S.¹, Amy D Willis, BActSt², Eric D Brooks, B.S.¹, Heidi Vanden Brink, MSc¹, Brittany Y Jarrett, B.S.¹, Roger A Pierson, PhD³, Donna R Chizen, MD³, and Marla E Lujan, PhD^{1,*}

¹Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA

²Department of Statistical Science, Cornell University, Ithaca, NY, USA

³Department of Obstetrics, Gynecology & Reproductive Sciences, University of Saskatchewan, Saskatoon, SK, Canada

Abstract

Objective—To compare the diagnostic potential of ultrasonographic markers of ovarian morphology, used alone or in combination, to predict PCOS.

Design—A diagnostic test study using cross-sectional data collected from 2006 to 2011.

Setting—Academic hospital and clinical research unit.

Participants—82 women with PCOS and 60 healthy female volunteers.

Interventions—None.

Main Outcome Measures—Follicle number per ovary (FNPO), ovarian volume (OV), follicle number per single cross-section (FNPS), follicle distribution pattern, stromal area, ovarian area, stromal to ovarian area ratio (S:A) and stromal index (SI).

Results—FNPO best predicted PCOS ($R^2=67\%$) with 85% sensitivity and 98% specificity, followed by OV ($R^2=44\%$), and FNPS ($R^2=36\%$). Neither S:A or SI had predictive power for PCOS. In combination, FNPO+S:A and FNPO+SI most significantly predicted PCOS ($R^2=74\%$ vs. 73% , respectively). The diagnostic potentials of OV and FNPS were substantially improved when used in combination (OV+FNPO, $R^2=55\%$).

Conclusions—As a single metric, FNPO best predicted PCOS. While the addition of S:A or SI improved the predictive power of FNPO, gains were marginal suggesting limited use in clinical practice. When image quality precludes a reliable estimation of FNPO, measurements of OV +FNPS provide the next closest level of diagnostic potential.

*Corresponding Author and Reprint Request Contact: Marla E Lujan, PhD, Assistant Professor, Division of Nutritional Sciences, Cornell University, 216 Savage Hall, Ithaca NY 14853. Tel. (607) 255 3153; Fax (607) 255 1033; MEL245@cornell.edu.

Disclosure Summary: The authors have no conflict of interests to disclose.

Keywords

Polycystic ovary syndrome; Ultrasound; Follicle; Stroma; Hyperandrogenism

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a heterogeneous disorder, originally characterized by Stein and Leventhal in 1935 given the consistent presence of abnormal ovarian morphology among patients (1). The significance of ovarian morphology to PCOS has since been debated. The 1990 National Institutes of Health (NIH) consensus statement on diagnostic criteria for PCOS did not include ovarian morphology, indicating that it was only suggestive – and not definitive – evidence of a diagnosis (2). However, more recent diagnostic criteria proposed by the American Society of Reproductive Medicine (ASRM) and European Society of Human Reproduction and Embryology (ESHRE) in 2003(3, 4), as well as by the Androgen Excess and PCOS Society in 2006(5), have reasserted the importance of ovarian morphology to the diagnosis of PCOS.

Since the 1980s, ultrasonography has allowed for the non-invasive assessment of polycystic ovarian morphology (6). The earliest studies assessing polycystic ovaries commented on three seemingly unique characteristics of the ovarian stroma: hypertrophy, increased echogenicity, and a tendency for the peripheral distribution of small ovarian follicles much like a “string of pearls”(6–9). Although this subjective assessment of the ovarian stroma was once common in the diagnosis of polycystic ovaries (7), it has since been replaced by more quantitative measures of follicle number per ovary (FNPO) and ovarian volume (OV). Increased FNPO and OV continue to be favored over other morphological characteristics of the polycystic ovary, despite criticism that these metrics are not purely specific to PCOS. In healthy women, OV has been found to naturally vary with age (10, 11), stage of puberty (12, 13), body mass index (BMI)(14), and insulin levels (15). Multiple reports have indicated that the threshold for FNPO supported by the ASRM/ESHRE Rotterdam consensus (12 follicles) has contributed to an increased prevalence of polycystic ovaries among healthy women of reproductive age (16–18). Additionally, significant intra- and inter-observer variability exists when counting follicles throughout the entire ovary (19, 20), suggesting that assessments of FNPO might be too subjective to serve as part of the diagnosis of polycystic ovaries. There is also no uniform consensus on whether follicles should be counted throughout the entire ovary or in a single cross-sectional view of the ovary.

While there appears to be inherent limitations to assessments of ultrasound features of polycystic ovaries, numerous groups have recently reported improvements in reliability of these measures given advancements in imaging technology. We developed a systematic approach for counting ovarian follicles that showed exceptional reproducibility in follicle counts made in polycystic ovaries (21). As a result, we and others have recently revised diagnostic thresholds for FNPO and OV in polycystic ovaries, which are obviating the artificial increase in polycystic ovarian morphology in the general population (22–24). Objective assessments of the ovarian stroma are now possible as a result of improved ultrasonographic imaging software. Fulghesu and colleagues (25) have developed a metric

for assessing the ovarian stroma – called the stromal to total area ratio (S:A) – which appears highly specific to PCOS. Digital technology now also enables an unbiased method of quantifying stromal echogenicity by evaluating the mean pixel intensity of selected regions in the ovary (26). Finally, quantitative assessments of follicle distribution pattern by multiple observers were shown to be associated with fair levels of agreement suggesting that distribution pattern evaluations may not be as subjective as once considered (20).

Although ultrasound metrics appear to have diagnostic utility on their own, few have quantified the predictive power gained when using these factors in combination to detect PCOS (9). Doing so may allow for an even more sensitive and specific diagnosis based on ultrasound. Thus, the primary objective of this study was to identify the diagnostic potential of each sonographic criterion to predict PCOS. The secondary objective was to determine the diagnostic capabilities of these morphological characteristics in combination to predict PCOS.

METHODS

Subjects

Eighty-two women diagnosed with PCOS by the NIH criteria as having both oligoamenorrhea and hyperandrogenism were recruited to the study. Oligoamenorrhea was defined as a history of unpredictable menstrual cycles shorter than 21 days or longer than 36 days. Hyperandrogenism was defined as a modified hirsutism score ≥ 7 (internally validated value having 83% sensitivity and 96% specificity to distinguish between PCOS and controls) and/or an elevated total testosterone concentration ≥ 114.12 ng/dL (internally validated value having 87% sensitivity and 100% specificity to distinguish between PCOS and controls). Sixty women from the general population with no hyperandrogenism and regular menstrual cycles served as controls. Participants were recruited from the general population using ads seeking healthy women of reproductive age or women with concerns over outward features of PCOS such as irregular periods, obesity, excess hair growth, and/or infertility. Volunteers ranged in age from 18 to 38 years and could not have used hormonal contraception, fertility medications or insulin sensitizers in the 3 months prior to enrollment. Subjects were ineligible if they had a previous history of ovarian surgery or current abnormalities in cortisol, prolactin, thyroid hormone, dehydroepiandrosterone sulfate or 17-hydroxyprogesterone secretion. Of the 62 women presenting for evaluation as controls, 2 were excluded because of a hirsutism. Of the 134 women presenting for evaluation over PCOS features 98 met the NIH criteria for PCOS. Of the 60 controls that proceeded for an ultrasound scan (described below), all were judged to have sufficient image quality to be included in the study. Of the 98 PCOS participants, 16 were excluded because of borderline image quality thereby generating a PCOS cohort of 82 participants.

Ultrasound Procedures and Measurements

Participants were evaluated by transvaginal ultrasonography by one of two experienced ultrasonographers. Control subjects were scanned on day 2 – 5 of the menstrual cycle and women with PCOS were scanned at an unspecified time. Ovaries were scanned from their inner to outer margins in the longitudinal plane using a 5 – 9 MHz transducer on an

Ultrasonix RP System (Vancouver, BC, Canada) or a 6 – 12 MHz transducer on a GE Voluson E8 System (Milwaukee, WI, USA). The age of the equipment varied by 3.5 years. A post-hoc comparison of ultrasound measures noted only a 4.5% difference in FNPO (absolute difference ± 1.27 follicles) and 7.7% difference in OV (absolute difference $\pm 0.88 \text{ cm}^3$) for the pooled data set. Because this difference in measurements among machines was considered within the error for these parameters, data obtained using both machines were pooled. Digital cine loops throughout each ovary (DICOM file format) and static images of the largest cross-sectional view of each ovary (JPEG file format) were archived for off-line analysis using Santesoft DICOM Editor (©Emmanouil Kanellopoulus, Athens, Greece).

Largest cross-sectional views of the ovary were evaluated by a single investigator for follicle distribution pattern, stromal area (SA), ovarian area (OA), S:A, mean stromal echogenicity, mean total echogenicity, and the stromal index (SI). Follicle distribution pattern was determined in ovaries if a follicle $\geq 10\text{mm}$ in size was not present. Participants were considered to have peripherally distributed follicles if the largest cross-sectional plane of both ovaries contained ≥ 9 follicles in a clear aggregation around the periphery with no more than 1 central follicle. OA was measured by outlining the external limits of the ovary with an electronic caliper tool. Stromal area was measured by outlining the peripheral profile of the stroma, avoiding antral follicles represented by anechoic structures in the ovary. The outline was extended to the periphery of the ovary when no follicles were present around that peripheral portion of the ovary. Mean echogenicity was defined as the sum of the product of each intensity level and the number of pixels for that intensity concentration, divided by the total number of pixels in the measured area. The SI was calculated by dividing the mean stromal echogenicity by the echogenicity of the entire ovary. This corrected for cases where gains were due to variations in image quality.

Ultrasound scans were also evaluated for the number of antral follicles (2–10mm) per ovary (FNPO), follicle number in the largest cross-sectional plane (FNPS), and OV. Reliable follicle counts were achieved for each ovary by imposing a programmable grid onto the viewing window as previously described (21). Based on an intra-class correlation coefficient analysis, the level of inter-observer agreement for FNPO and FNPS by three observers was 0.84 and 0.94, respectively. Ovarian volume was estimated using the equation: $\pi/6$ (transverse diameter) $\times$ (anteroposterior diameter) $\times$ (longitudinal diameter). When all follicles in both the left and right ovary were $<10\text{mm}$ in size, a value for OV was designated as the mean recorded values of both ovaries. When a follicle $\geq 10\text{mm}$ was present in a single ovary, the OV for the other ovary was used, and when present in both ovaries the OV was excluded as per the method previously reported by Johnstone et al.(17). The level of inter-observer agreement for OV by three observers was 0.96. A value for SA, OA, S:A, SI, FNPO, and FNPS for each participant was designated as the mean recorded values of the left and right ovaries. Pictorial representations for each sonographic endpoint are presented in Figure 1.

Endocrine Assays

Sex hormone-binding globulin was measured by a commercially available two-site chemiluminescent immunogenic assay (Siemens Healthcare Diagnostics, Deerfield, IL,

USA). Total testosterone was measured by isotope dilution liquid chromatography tandem mass spectrometry as previously described (27). Intra- and interassay coefficients of variation were <10% for both assays. Free androgen index was calculated by dividing the total testosterone level by the SHBG level and multiplying by 100.

Statistical Analyses

JMP 10 Statistical Software (SAS Institute Inc., Cary, NC, USA) and the Statistical Package for the Social Sciences 20 (SPSS Inc., Chicago, IL, USA) were used to perform the analyses. Descriptive statistics (median and inter-quartile ranges) were tabulated for baseline characteristics. Mann-Whitney tests were performed to compare continuous parameters between women with PCOS and controls. A two-sample Z test for equal proportions was used to compare follicle distribution pattern assignments between groups. Receiver Operating Characteristic (ROC) curve analysis was used to evaluate the accuracy of sonographic endpoints in the diagnosis of PCOS. Diagnostic thresholds for sonographic features were based on Youden's index, which balanced maximum test sensitivity and test specificity.

Multiple logistic regression analysis was used to determine the diagnostic potential of sonographic markers in combination. A systematic approach was first taken to find the best model that predicted PCOS. Then, because FNPO, FNPS, and OV are markers most currently recommended for evaluation in clinical practice (28, 29), we analyzed if any additional features improved the diagnostic potential of these markers alone. To accomplish these goals, combinations of FNPO, FNPS, OV, OA, SA, S:A, SI and distribution pattern were entered in a stepwise manner to the logistic regression. Markers were retained in the model if the addition of a marker significantly improved the model fit based on a likelihood ratio test. Comparisons between models were done by assessing both the model Akaike Information Criterion (AIC), and R^2 values of each model. AIC is a model selection tool that balances model size with explanatory power. AIC decreases with increased explanatory power and increases with model size. As such, a low AIC is indicative of increased model strength for a given model size (30). R^2 corresponds to the percent variability in outcomes (i.e. PCOS diagnosis) that can be explained by factors entered into each model. A higher R^2 is indicative of increased model strength. Based on a probability greater than 50% that a woman was correctly classified as PCOS, diagnostic equations were generated using the parameter estimates of each logistic regression. Equations were generated so that a positive diagnosis was obtained if the inclusion of ultrasound parameters specified was greater than or equal to the model estimate. Each equation's sensitivity and specificity for this sample of women was calculated based on the ability of the equation to correctly diagnose PCOS using the current study population.

Ethical Considerations

This study was approved by the University of Saskatchewan Biomedical Research Ethics Review Board and Cornell University's Institutional Review Board. Interactions with human participants occurred at the Royal University Hospital within the Department of Obstetrics, Gynecology and Reproductive Sciences, University of Saskatchewan (Saskatoon, SK, Canada) from 2006 to 2008, and in the Division of Nutritional Sciences' Human Metabolic

Research Unit, Cornell University (Ithaca, NY, USA) from 2009 to 2011. Informed, written consent was obtained from all study participants.

RESULTS

Baseline clinical and hormonal features in women with PCOS and controls are presented in Table 1. Women with PCOS were similar in age to controls, but had higher BMI. As expected, women with PCOS had longer lengths between menses, higher hirsutism scores, total testosterone, free androgen index, FNPO, FNPS, OV and OA. Although women with PCOS had a larger stromal area, S:A and SI did not differ between groups. A significantly larger proportion of women with PCOS (26.6%) were classified as having peripherally distributed follicles compared to healthy women (1.7%; $P<.0001$). Scatterplots of ultrasonographic endpoints are presented in Figure 2.

The level of diagnostic accuracy for each ultrasound metric is presented in Table 2. FNPO, FNPS, OV, OA, SA and follicle distribution pattern had significant diagnostic potential (as judged by area under the ROC curve), while S:A and SI did not predict PCOS. A threshold of 28 follicles for FNPO had the best combined sensitivity (85%) and specificity (98%) of any marker followed by an OV of 10cm^3 (Sensitivity=90% and Specificity=86%).

Multiple logistic regression analyses performed using FNPO, FNPS, OV, OA, SA, S:A, SI, and follicle distribution pattern as possible independent variables produced two models which best predicted PCOS (Table 3). A combination of FNPO+S:A, as well as FNPO+SI, were the most significant predictive factors for PCOS. The addition of a third metric to either model did not further improve the models predictive power (data not shown). Estimates of diagnostic potential based on ROC curve analysis for both models showed similar predictive power for PCOS (FNPO+S:A=0.980 and FNPO+SI=0.978). Using the generated diagnostic equations, FNPO+SI had a combined sensitivity of 94% and specificity of 93%, while FNPO+S:A had a combined sensitivity of 91% and specificity of 93%.

When assessing if any additional factors could improve the diagnostic potential of OV or FNPS, the best model included a combination of both FNPS and OV (Table 3). FNPS+OV had an area under the ROC curve of 0.941 and, using the proposed diagnostic equation, a combined sensitivity of 89% and specificity of 93%. Although this model had improved predictive power (AIC=85.82) and better accounted for variability in PCOS ($R^2=55\%$) compared to either FNPS (AIC=127.50, $R^2=36\%$) or OV alone (AIC=102.98, $R^2=44\%$), power was not improved compared to FNPO alone (AIC=64.90, $R^2=69\%$). Further, while the addition of markers to FNPS+OV yielded no improvements, the combination of FNPS with other sonographic features produced models with improved diagnostic potential compared to FNPS alone (Table 3).

DISCUSSION

The main objective of the current study was to contrast the diagnostic potential of multiple sonographic criteria, alone or in combination, to predict PCOS. These evaluations are particularly timely given the urgent need to redefine polycystic ovarian morphology and obviate the growing misconception that polycystic ovaries are a highly prevalent finding in

the general population. Using a subset of women judged to have superior image quality from our recently published study cohort (22), we re-evaluated the diagnostic potential of FNPO, FNPS and OV, in addition to exploring the potential of stromal assessments to predict PCOS. As expected, we noted similar findings to our earlier study with regards to the diagnostic potential, specificity and sensitivity of FNPO, FNPS and OV to detect PCOS (22). Of note is the repeated finding that a substantially higher threshold for FNPO (more than twice that supported by the 2003 ESHRE/ASRM Rotterdam consensus) is needed to detect PCOS. In fact, had we considered 12 follicles consistent with a definition of polycystic ovaries then 73% of the control population would have met this definition. Data supporting the notion that the Rotterdam recommendations for FNPO are obsolete are consistent with other groups who have reported substantially higher thresholds for FNPO using newer imaging technology (23, 24). While our proposed threshold for FNPO was higher compared to earlier studies (7, 9, 23, 24, 31–37), our thresholds for OV (10 cm³) and FNPS (9 follicles) were remarkably similar to past reports (7, 9, 23, 24, 34, 35, 38, 39). This would suggest that improvements in spatial resolution have not substantially impacted measurements dependent on a single cross-sectional view of the ovary. However, the finding that FNPO had improved diagnostic potential compared to OV and FNPS alone supports the conclusion that use of FNPO over these two parameters should be employed whenever possible. It must be noted that we employed a very precise assessment of FNPO which may slightly overestimate follicle counts obtained in a busy clinical practice setting.

We noted limited potential for measurements of the ovarian stroma to predict PCOS. Of the four markers assessed, only two (SA and follicle distribution pattern) had diagnostic potential, better than chance alone, to detect PCOS. However, their diagnostic potential was substantially lower than that for FNPO, FNPS and OV which limits their utility in clinical practice. Unexpectedly, we did not note any difference in S:A or SI between women with PCOS and controls – nor did we detect any diagnostic potential for these parameters. These findings conflict with a previous study showing S:A to have 100% sensitivity and 100% specificity to diagnose PCOS (25) and to reports showing that both subjective scoring of stromal echogenicity and SI distinguished well between PCOS and healthy women (9, 26, 31). A potential explanation for discrepancies between studies might relate to differences in the composition of the cohorts used. In our current study, we compared healthy women to women with PCOS as identified by the NIH criteria. Inclusion of participants was not based on any prior examination of the ovaries. We felt this approach was warranted as it eliminated any bias with regards to the actual clinical spectrum of PCOS and the physiological range of ovarian morphology in women from the general population which may have included asymptomatic women with polycystic ovaries. This was in contrast to those former studies (25, 26) whose PCOS cohorts had confirmed presence of polycystic ovaries at inclusion. By excluding women without polycystic ovaries, the differences seen between PCOS and control populations may have been exaggerated in their studies. Additionally, we cannot exclude the possibility that our stromal measurements might have varied slightly from earlier studies (25, 40, 41) – despite our attempts to replicate the methodology as closely as possible. In the case of S:A, we believe our method may be more similar to that reported by Dewailly and colleagues because we extended our stromal measurements to areas between follicles (42). This group subtracted the area of all

individual “microcysts” in the largest cross-sectional plane to produce the stromal area. The ratio between the stromal and total area found by Dewailly et al. approximated 0.74 for women with PCOS and 0.71 for controls, which is more in line with our S:A values.

While use of multiple ultrasonographic criteria to detect PCOS was proposed in the 1980s (7), few studies have quantified the diagnostic potential of this approach despite advancements in technology (9, 38). We found that the best models for predicting PCOS were a combination of FNPO+SI and FNPO+S:A. These results are consistent with those by Atiomo et al.(9) who found that a combination of FNPO and elevated stromal echogenicity best predicted PCOS. Although we would have expected that only metrics that differed between women with PCOS and controls would add diagnostic potential, the addition of S:A and SI may control for a factor that helps distinguish between the overlap in FNPO among populations. This agrees with the analyses by Belosi et al.(43) which found that S:A ratio was able to identify a subset of patients that fell between the NIH and the Rotterdam criteria but clearly exhibited abnormal clinical and hormonal profiles compared to controls. We hypothesize that S:A and SI reflect the distribution and density of follicles located around the periphery of the ovary. That these factors improved the diagnostic model, while patterning did not, is likely due to their quantitative nature which provided a more graduated evaluation than possible when using a categorical measure of patterning. In our opinion, the addition of S:A or SI to assessments of FNPO may not be clinically relevant. While both models had improved diagnostic potential compared to FNPO alone, the contribution was minimal and specificity was decreased by 5% in both models.

It is important to acknowledge that in women meeting both clinical and endocrine evidence of PCOS, the relevance of detecting polycystic ovaries is uncertain. While confirmation of polycystic ovaries may add confidence to a challenging clinical diagnosis, at present the presence of polycystic ovaries does not appear to yield additional relevant information for the diagnosis in this population (44–47). In our current study, we noted that approximately 15% of women with endocrine-based PCOS did not have polycystic ovaries. Whether women with “Non-PCO PCOS” differ in long-term health risks compared to those with polycystic ovaries has not been resolved. Another limitation of our study was that we did not assess ovarian morphology in populations at risk for PCOS – namely, women with oligoanovulation or hyperandrogenism alone. Certainly, these are clinical populations in which a definition of polycystic ovarian morphology is most needed. Future studies should resolve whether thresholds for ovarian morphology established in women with endocrine-based PCOS adequately discriminate women with milder PCOS phenotypes. It is possible that milder PCOS phenotypes may require different thresholds for either morphological or functional markers (i.e. anti-Müllerian hormone) of polycystic ovaries.

In summary, FNPO in combination with SI or S:A had the best diagnostic potential for the detection of PCOS on ultrasonography. However, gains in diagnostic potential using these combined metrics were marginal and resulted in some loss of specificity. Such as, we believe it is more prudent to use FNPO alone because of the increased time to evaluate both FNPO and the ovarian stroma. When an assessment of FNPO is not possible, due to reduced image quality or dependence on older equipment, a combined metric of FNPS and OV

provides significant predictive power in detecting PCOS comparable to that of FNPO alone. We recommend against the use of stromal evaluations on their own to predict PCOS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grants and Funding: This study was funded by Cornell University and fellowship awards from the Saskatchewan Health Research Foundation and Canadian Institutes of Health Research.

REFERENCES

1. Stein IF, Leventhal ML. Amenorrhea associated with bilateral polycystic ovaries. *Am J Obstet Gynecol.* 1935; 29:181–191.
2. Zawadzki, JADA. Diagnostic criteria for polycystic ovary syndrome: towards a rational approach. In: Dunaif, A.; Givens, JR.; Haseltine, FP.; Merriam, GR., editors. *Polycystic Ovary Syndrome.* Boston: Blackwell Scientific; 1992. p. 377–384.
3. Group REA-SPCW. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril.* 2004; 81:19–25.
4. Group REA-SPCW. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Hum Reprod.* 2004; 19:41–47. [PubMed: 14688154]
5. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterwiet W, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil Steril.* 2009; 91:456–488. [PubMed: 18950759]
6. Swanson M, Sauerbrei EE, Cooperberg PL. Medical implications of ultrasonically detected polycystic ovaries. *J Clin Ultrasound.* 1981; 9:219–222. [PubMed: 6787087]
7. Adams J, Franks S, Polson DW, Mason HD, Abdulwahid N, Tucker M, et al. Multifollicular ovaries: clinical and endocrine features and response to pulsatile gonadotropin releasing hormone. *Lancet.* 1985; 2:8469–8470.
8. Takahashi K, Ozaki T, Okada M, Uchida A, Kitao M. Relationship between Ultrasonography and Histopathological Changes in Polycystic Ovarian Syndrome. *Hum Reprod.* 1994; 9:2255–2258. [PubMed: 7714140]
9. Atiomo WU, Pearson S, Shaw S, Prencice A, Dubbins P. Ultrasound criteria in the diagnosis of polycystic ovary syndrome (PCOS). *Ultrasound Med Biol.* 2000; 26:977–980. [PubMed: 10996697]
10. Erdem A, Erdem M, Biberoglu K, Hayit O, Arslan M, Gursoy R. Age-related changes in ovarian volume, antral follicle counts and basal FSH in women with normal reproductive health. *J Reprod Med.* 2002; 47:835–839. [PubMed: 12418067]
11. Erdem M, Erdem A, Biberoglu K, Arslan M. Age-related changes in ovarian volume, antral follicle counts and basal follicle stimulating hormone levels: comparison between fertile and infertile women. *Gynecol Endocrinol.* 2003; 17:199–205. [PubMed: 12857427]
12. Ersen A, Onal H, Yildirim D, Adal E. Ovarian and uterine ultrasonography and relation to puberty in healthy girls between 6 and 16 years in the Turkish population: a cross-sectional study. *J Pediatr Endocrinol Metab.* 2012; 25:447–451. [PubMed: 22876537]
13. Razzaghy-Azar M, Ghasemi F, Hallaji F, Ghasemi A, Ghasemi M. Sonographic measurement of uterus and ovaries in premenarcheal healthy girls between 6 and 13 years old: correlation with age and pubertal status. *J Clin Ultrasound.* 2011; 39:64–73. [PubMed: 20572067]
14. Bastos CA, Oppermann K, Fuchs SC, Donato GB, Spritzer PM. Determinants of ovarian volume in pre-, menopausal transition, and post-menopausal women: A population-based study. *Maturitas.* 2006; 53:405–412. [PubMed: 16111843]

15. Frajndlich R, Spritzer PM. Association between ovarian volume and serum insulin levels in ovulatory patients with idiopathic hirsutism. *Fertil Steril*. 2005; 83:1561–1564. [PubMed: 15866607]
16. Duijkers, IJM; Klipping, C. Polycystic ovaries, as defined by the 2003 Rotterdam consensus criteria, are found to be very common in young healthy women. *Gynecol Endocrinol*. 2010; 26:152–160. [PubMed: 19757237]
17. Johnstone EB, Rosen MP, Neril R, Trevithick D, Sternfeld B, Murphy R, et al. The polycystic ovary post-Rotterdam: a common, age-dependent finding in ovulatory women without metabolic significance. *J Clin Endocrinol Metab*. 2010; 95:4965–4972. [PubMed: 20719841]
18. Kristensen SL, Ramlau-Hansen CH, Ernst E, Olsen SF, Bonde JP, Vested A, et al. A very large proportion of young Danish women have polycystic ovaries: is a revision of the Rotterdam criteria needed? *Hum Reprod*. 2010; 25:3117–3122. [PubMed: 20940139]
19. Lujan ME, Chizen DR, Peppin AK, Dhir A, Pierson RA. Assessment of ultrasonographic features of polycystic ovaries is associated with modest levels of inter-observer agreement. *J Ovarian Res*. 2009; 2:6. [PubMed: 19515259]
20. Lujan ME, Chizen DR, Peppin AK, Kriegler S, Leswick DA, Blotski TG, et al. Improving inter-observer variability in the evaluation of ultrasonographic features of polycystic ovaries. *Reprod Biol Endocrinol*. 2008; 6:30–41. [PubMed: 18638401]
21. Lujan ME, Brooks ED, Kepley AL, Chizen DR, Pierson RA, Peppin AK. Grid analysis improves reliability in follicle counts made by ultrasonography in women with polycystic ovary syndrome. *Ultrasound Med Biol*. 2010; 36:712–718. [PubMed: 20381953]
22. Lujan ME, Jarrett BY, Brooks ED, Reines JK, Peppin AK, Muhn N, et al. Updated ultrasound criteria for polycystic ovary syndrome: reliable thresholds for elevated follicle population and ovarian volume. *Hum Reprod*. 2013; 28:1361–1368. [PubMed: 23503943]
23. Dewailly D, Gronier H, Poncelet E, Robin G, Leroy M, Pigny P, et al. Diagnosis of polycystic ovary syndrome (PCOS): revisiting the threshold values of follicle count on ultrasound and of the serum AMH level for the definition of polycystic ovaries. *Hum Reprod*. 2011; 26:3123–3129. [PubMed: 21926054]
24. Allemand MC, Tummon IS, Phy JL, Foong SC, Dumesic DA, Session DR. Diagnosis of polycystic ovaries by three-dimensional transvaginal ultrasound. *Fertil Steril*. 2006; 85:214–219. [PubMed: 16412756]
25. Fulghesu AM, Ciampelli M, Belosi C, Apa R, Pavone V, Lanzzone A. A new ultrasound criterion for the diagnosis of polycystic ovary syndrome: the ovarian stroma/total area ratio. *Fertil Steril*. 2001; 76:326–331. [PubMed: 11476780]
26. Buckett WM, Bouzayen R, Watkin KL, Tulandi T, Tan SL. Ovarian stromal echogenicity in women with normal and polycystic ovaries. *Hum Reprod*. 1999; 14:618–621. [PubMed: 10221685]
27. Lujan M, Blotski TG, Chizen DR, Lehotay DC, Pierson RA. Digit ratios do not serve as anatomical evidence of prenatal androgen exposure in clinical phenotypes of polycystic ovary syndrome. *Hum Reprod*. 2010; 25:204–211. [PubMed: 19855107]
28. American Institute of Ultrasound in M. AIUM practice guideline for the performance of pelvic ultrasound examinations. *J Ultrasound Med*. 2010; 29:166–172. [PubMed: 20040793]
29. Scoutt L, Wax J, Cohen HL, Diacon L, Fox JC, Fulgham P, et al. AIUM Practice Guideline for the Performance of a Focused Reproductive Endocrinology and Infertility Scan. *J Ultrasound Med*. 2012; 31:1865–1874. [PubMed: 23248819]
30. Neter, J.; Wasserman, W.; Kutner, MH. *Applied linear statistical models : regression, analysis of variance, and experimental designs*. 3rd ed.. Homewood, Ill: Irwin; 1990.
31. Pache TD, Wladimiroff JW, Hop WC, Fauser BC. How to discriminate between normal and polycystic ovaries: transvaginal US study. *Radiology*. 1992; 183:421–423. [PubMed: 1561343]
32. Fox R. Transvaginal Ultrasound Appearances of the Ovary in Normal Women and Hirsute Women with Oligomenorrhoea. *Aust N Z J Obstet Gynaecol*. 1999; 39:63–68. [PubMed: 10099753]
33. Jonard S, Robert Y, Cortet-Rudelli C, Pigny P, Decanter C, Dewailly D. Ultrasound examination of polycystic ovaries: is it worth counting the follicles? *Hum Reprod*. 2003; 18:598–603. [PubMed: 12615832]

34. Jonard S, Robert Y, Dewailly D. Revisiting the ovarian volume as a diagnostic criterion for polycystic ovaries. *Hum Reprod.* 2005; 20:2893–2898. [PubMed: 16006474]
35. Chen Y, Li L, Chen X, Zhang Q, Wang W, Li Y, et al. Ovarian volume and follicle number in the diagnosis of polycystic ovary syndrome in Chinese women. *Ultrasound Obstet Gynecol.* 2008; 32:700–703. [PubMed: 18773451]
36. Yeh HC, Futterweit W, Thornton JC. Polycystic ovarian disease: US features in 104 patients. *Radiology.* 1987; 163:111–116. [PubMed: 3547491]
37. Kö ü N, Kö ü A, Turhan NÖ, Kamalak Z. Do threshold values of ovarian volume and follicle number for diagnosing polycystic ovarian syndrome in Turkish women differ from western countries? *Eur J Obstet Gynecol Reprod Biol.* 2011; 154:177–181. [PubMed: 21051133]
38. Van Santbrink EJP, Hop WC, Fause BCJM. Classification of normogonadotropic infertility: polycystic ovaries diagnosed by ultrasound versus endocrine characteristics of polycystic ovary syndrome. *Fertil Steril.* 1997; 67:452–458. [PubMed: 9091329]
39. Fulghesu AM, Angioni S, Frau E, Belosi C, Apa R, Mioni R, et al. Ultrasound in polycystic ovary syndrome—the measuring of ovarian stroma and relationship with circulating androgens: results of a multicentric study. *Hum Reprod.* 2007; 22:2501–2508. [PubMed: 17635847]
40. Battaglia C, Battaglia B, Morotti E, Paradisi R, Zanetti I, Meriggiola MC, et al. Two- and three-dimensional sonographic and color Doppler techniques for diagnosis of polycystic ovary syndrome. The stromal/ovarian volume ratio as a new diagnostic criterion. *J Ultrasound Med.* 2012; 31:1015–1024. [PubMed: 22733850]
41. Fulghesu AM, Angioni S, Frau E, Belosi C, Apa R, Mioni R, et al. Ultrasound in polycystic ovary syndrome—the measuring of ovarian stroma and relationship with circulating androgens: results of a multicentric study. *Hum Reprod.* 2007; 22:2501–2508. [PubMed: 17635847]
42. Dewailly D, Robert Y, Hellin I, Ardsens Y, Thomas-Desrousseaux P, Lemaitre L, et al. Ovarian stromal hypertrophy in hyperandrogenic women. *Clin Endocrinol.* 1994; 41:557–562.
43. Belosi C, Selvaggi L, Apa R, Guido M, Romualdi D, Fulghesu AM, et al. Is the PCOS diagnosis solved by ESHRE/ASRM 2003 consensus or could it include ultrasound examination of the ovarian stroma? *Hum Reprod.* 2006; 21:3108–3115. [PubMed: 17053004]
44. Diamanti-Kandarakis E, Panidis D. Unravelling the phenotypic map of polycystic ovary syndrome (PCOS): a prospective study of 634 women with PCOS. *Clin Endocrinol.* 2007; 67:735–742.
45. Loucks TL, Talbott EO, McHugh KP, Keelan M, Berga SL, Guzick DS. Do polycystic-appearing ovaries affect the risk of cardiovascular disease among women with polycystic ovary syndrome? *Fertil Steril.* 2000; 74:547–552. [PubMed: 10973653]
46. Guastella E, Longo RA, Carmina E. Clinical and endocrine characteristics of the main polycystic ovary syndrome phenotypes. *Fertil Steril.* 2010; 94:2197–2201. [PubMed: 20303485]
47. Moran L, Teede H. Metabolic features of the reproductive phenotypes of polycystic ovary syndrome. *Hum Reprod Update.* 2009; 15:477–488. [PubMed: 19279045]

Figure 1. Ultrasound evaluations of ovarian morphology

The number of follicles in a single cross-section of the ovary (FNPS) or throughout the entire ovary (FNPO) were determined by applying a grid and systematically counting and flagging individual follicles per grid section (A). Ovarian volume was calculated by measuring the longest longitudinal and widest anteroposterior dimensions of the ovary (B). The transverse diameter was obtained in the corresponding 90° plane (not shown). Ovarian area was determined by tracing the limits of the ovary using calipers (C). Stromal area was determined by tracing the internal limits of the ovary while avoiding the anechoic antral

follicles using calipers (D). An ovary with a clear aggregation of follicles around the periphery (E) is contrasted with an ovary with a non-peripheral distribution of follicles (F).

Figure 2. A comparison of ultrasonographic endpoints among women with polycystic ovary syndrome (PCOS) and healthy controls

Follicle number per ovary (FNPO; A); follicle number per cross-section (FNPS; B), ovarian volume (OV; C); ovarian area (OA; D) and stromal area (SA; E) were significantly higher in women with PCOS compared to controls. By contrast, no differences in stromal to ovarian area ratio (S:A; F) or stromal index (SI; G) were noted between groups. Boxes represent the 25th and 75th percentiles and the horizontal band within the box represents the median. The 5th–95th percentile range is denoted by the vertical bars.

Table 1

A comparison of clinical, hormonal and ultrasonographic features in women with polycystic ovary syndrome (PCOS) and controls

	Control (n=60)	PCOS (n=82)	P-value
Age (years)	27 (24–31)	28 (24–31)	NS
BMI (kg/m ²)	23.7 (21.9–27.2)	31.2 (23.7–37.8)	<.0001
Menstrual cycle length (d)	29 (28–30)	75 (55–155)	<.0001
Hirsutism score	2 (0–5)	11 (7–15)	<.0001
Total testosterone (ng/dL)	80.7 (63.4–95.1)	100.9 (74.9–141.2)	0.001
Free androgen index (%)	4.6 (3.8–7)	11.9 (6.8–17.4)	<.0001

Median values are presented with 25–75th quartiles in parentheses. Differences between groups were determined using the Mann-Whitney test. NS, not significant

Table 2

Sensitivity and specificity of proposed diagnostic thresholds for the ultrasonographic detection of polycystic ovary syndrome (PCOS)

Criterion	AIC	R ² (%)	AUC (95% CI)	Threshold	Se (%)	Sp (%)
FNPO	64.90	69	0.971 [*] (0.948, 0.993)	28	85	98
FNPS	127.50	36	0.872 [*] (0.816, 0.929)	9	71	90
Ovarian Volume (cm ³)	102.98	44	0.913 [*] (0.859, 0.966)	10	90	86
Ovarian Area (cm ²)	141.66	27	0.822 [*] (0.753, 0.890)	5	85	64
Stromal Area (cm ²)	161.68	16	0.746 [*] (0.664, 0.828)	3	80	57
Stromal to Ovarian Area Ratio	189.64	01	0.417 (0.317, 0.517)	N/A	73	47
Stromal Index	191.46	00	0.554 (0.452, 0.657)	N/A	79	47
Follicle distribution pattern	171.64	10	0.624 [*] (0.532, 0.717)	N/A	27	98

Diagnostic potential was evaluated via logistic regression and ROC curve analysis. Thresholds for criterion were generated based on Youden's index. AIC, Akaike Information Criterion; AUC, Area under Receiver Operating Characteristic curve; FNPO, follicle number per ovary; FNPS, follicle number per cross-section; N/A, not applicable; Se, Sensitivity; Sp, Specificity;

^{*} P<.05 compared with chance alone.

Table 3

Diagnostic models using multiple ultrasound metrics for the detection of polycystic ovary syndrome (PCOS).

Model	AIC	R ² (%)	AUC (95% CI)	Diagnostic Equation [†]	Se (%)	Sp (%)
FNPO + SI	55.69	74	0.978 (0.959, 0.997)	0.73 0.35(FNPO) – 5.87(SI)	94	93
FNPO + S:A	56.79	73	0.980 (0.963, 0.998)	13.82 0.35(FNPO) + 9.21(S:A)	91	93
OV + FNPS	85.82	55	0.941 (0.901, 0.981)	6.75 0.31(OV) + 0.47(FNPS)	89	93
FNPS + FDP + SA	106.70	47	0.919 (0.873, 0.965)	4.97 51(FNPS) ± 1.15(FDP) [‡] + 0.78(SA)	84	81
FNPS + FDP + OA	106.82	47	0.921 (0.877, 0.965)	5.54 46(FNPS) ± 1.00(FDP) [‡] + 0.62(OA)	82	86
FNPS + FDP + SI	109.89	46	0.913 (0.868, 0.959)	–1.06 0.62(FNPS) ± 1.27(FDP) [‡] – 3.46(SI)	82	76
FNPS + OA	111.72	44	0.910 (0.861, 0.959)	6.85 0.48(FNPS) + 0.69(OA)	88	79
FNPS + FDP	112.54	43	0.900 (0.766, 0.906)	3.27 58(FNPS) ± 1.16(FDP) [‡]	82	83
FNPS + SA	112.82	43	0.906 (0.855, 0.958)	6.31 0.56(FNPS) + 0.82(SA)	85	83
FNPS + SI	117.80	41	0.897 (0.847, 0.948)	0.90 66(FNPS) – 2.84(SI)	81	78

Diagnostic potential was evaluated via logistic regression and ROC curve analysis. AIC, Akaike Information Criterion; AUC, Area under Receiver Operating Characteristic curve; FNPO, follicle number per ovary; SI, stromal index; FNPS, follicle number per cross-section; FDP, follicle distribution pattern; SA, stromal area; OA, ovarian area; S:A, stromal to ovarian area ratio; Se, sensitivity; Sp, specificity.

* P<.05 compared with chance alone.

[†] Classification equations are written so that a positive diagnosis is obtained when the equation is true.

[‡] When follicles are peripherally distributed add the constant, when they are not, subtract the constant.