Title: Development of a novel risk prediction and risk stratification score for polycystic ovary syndrome (PCOS)

Harshal Deshmukh[1], Maria Papageorgiou[1], Eric S Kilpatrick[2], Stephen L Atkin[3], Thozhukat Sathyapalan[1]

[1]Department of Academic Diabetes, Endocrinology and Metabolism, Hull York Medical School, University of Hull, Hull, UK; harshaldeshmukh@nhs.net; M.Papageorgiou@hull.ac.uk (M.P.); Thozhukat.Sathyapalan@hyms.ac.uk (T.S.)

[2]Department of Pathology, Sidra Medical and Research Center, Doha, Qatar; ekilpatrick@sidra.org (E.S.K)

[3]Weill Cornell Medicine in Qatar, Education City, P.O. Box 24144, Qatar; Sla2002@qatar-med.cornell.edu (S.L.A)

Short Title: Risk prediction in PCOS

Key words: PCOS, AMH, FAI, 17-OHP, risk-score

Word count (2631)

Correspondence: Professor Thozhukat Sathyapalan, Department of Academic Diabetes, Endocrinology and Metabolism, Hull York Medical School, University of Hull, Anlaby Road HU3 2RW, Hull, UK, T. (01482) 675387; Email address: Thozhukat.Sathyapalan@hyms.ac.uk.

Abstract

Objective: The aim of this study was to develop a simple phenotypic algorithm that can capture the underlying clinical and hormonal abnormalities to help in the diagnosis and risk stratification of PCOS. **Methods:** The study consisted of 111 women with PCOS fulfilling the Rotterdam diagnostic criteria and 67 women without PCOS. A Firth's penalised logistic regression model was used for independent variable section. Model optimism, discrimination and calibration were assessed using bootstrapping, area under the curve (AUC) and Hosmer-Lemeshow statistics, respectively. The Prognostic index (PI) and risk score for developing PCOS was calculated using independent variables from the regression model. **Results**: Firth penalised logistic regression model with backward selection identified 4 independent predictors of PCOS namely, free androgen index [Beta 0.30(0.12), p=0.008], 17-OHP [Beta=0.20(0.01), P=0.026], anti-mullerian hormone (AMH) [Beta=0.04(0.01) p<0.0001], and waist-circumference [Beta=0.08(0.02), p<0.0001]. The model estimates indicated high internal validity (minimal optimism on 1000-fold bootstrapping), good discrimination ability (bias corrected c-statistic=0.90) and good calibration (Hosmer–Lemeshow chi-squared=3.7865). PCOS women with a high risk score (q1+q2+q3 *vs.* q4) presented with a worse metabolic profile characterised by a higher 2-hour glucose (p=0.01), insulin (p=0.0003), triglycerides (p=0.0005), C-reactive protein (p<0.0001) and low HDL-cholesterol (p=0.02) as compared to those with lower risk score for PCOS. **Conclusion:** We propose a simple 4-variable model, which captures the underlying clinical and hormonal

abnormalities in PCOS and can be used for diagnosis and metabolic risk stratification in women with PCOS.

Introduction

Polycystic ovary syndrome (PCOS) is one of the most common endocrine disorders affecting up to 20% of reproductive-aged women of (1, 2). There are three available diagnostic criteria for PCOS; the National Institute of Health (NIH) (3), ESHRE/ASRM Rotterdam consensus criteria (4) and the androgen excess and PCOS society (AES)(5). Biochemical hyperandrogenism is a common component of each of three criteria and can be assessed by using a variety of assays to test for relevant biomarkers in serum and/or saliva including serum levels of total testosterone (TT), free T, androstenedione, and dehydroepiandrosterone sulphate (DHEAS) or by calculating available indices such as free androgen index. This plethora of available androgen biomarkers and indices in combination with the current little guidance on cut-offs indicative of androgen excess in the PCOS guidelines (3-5) contribute to diagnosis- and risk stratification- related uncertainties. FAI is commonly used to define hyperandrogenemia in the diagnosis of PCOS. However, recent data (6) shows that FAI is not a reliable indicator of free T when sex hormone binding globulin (SHBG) concentration is low and hence can misclassify women who are being investigated for PCOS. Clinical hyperandrogenemia, characterised by the presence of hirsutism is recommended as a substitute of biochemical hyperandrogenemia in the current guidelines but this can often be unreliable due to wide inter-observer variation and ethnic variations (7) . While the focus has been placed upon biochemical and clinical hyperandrogenemia for the diagnosis of PCOS, recent data by our group (8) and others (9) have shown that elevated levels of anti-Mullerian hormone (AMH), a surrogate measure of follicle count on ultrasound, can be an important supplement to the hormonal parameters used in the diagnosis of PCOS. While PCOS is a diagnosis of exclusion, the diagnosis can often be challenging, given the presentation of this syndrome as a spectrum of clinical features and metabolic abnormalities in the affected

patients, rather than the presence of a single unified entity, PCOS. The aim of this study was to use relevant biochemical markers and quantifiable clinical features to derive a risk score that can capture the entire PCOS disease spectrum. This simple risk score has the potential to assist in diagnosis, severity prediction of the disease risk stratification of PCOS women.

Methods

Study population

This was a cross sectional study involving 111 well characterised women with PCOS and 67 women without PCOS who presented sequentially and prospectively at the Department of Academic Diabetes, Endocrinology and Metabolism. All patients gave written informed consent. This study was approved by the Newcastle & North Tyneside Ethics committee (ISRCTN70196169) and was conducted in accordance to the Declaration of Helsinki and local regulations. The diagnosis of PCOS was based on at least two out of three of the diagnostic criteria of the Rotterdam consensus, namely clinical and biochemical evidence of hyperandrogenism (Ferriman-Gallwey score >8; free androgen index >4, total testosterone >1.5 nmol/l), oligomenorrhea or amenorrhea and polycystic ovaries on transvaginal ultrasound. Non-classical 21-hydroxylase deficiency, hyperprolactinemia, Cushing's disease and androgen-secreting tumours were excluded by appropriate tests. The study and study measurements are described in detail in our previous publication(8) . In summary we measured body mass index (BMI) (kg/m2), waist circumference (cm), hip circumference (cm), AMH (pmol/l), salivary testosterone (pmol/l), total testosterone (nmol/L), salivary androstenedione (pmol/l), serum androstenedione (nmol/L), SHBG (nmol/L) , FAI (%) , follicle stimulating hormone (FSH) (IU/L), Leutenizing hormone (LH) (IU/L) , fasting glucose (mmol/L), 2-Hour glucose (mmol/L), insulin (µIU/ml) according to established protocols in women with PCOS and controls. We also ascertained oral contraceptive use and

history of menstrual irregularity/amenorrhoea. All of the control women had regular periods, no clinical or biochemical hyperandrogenism, no polycystic ovaries on ultrasound, no significant background medical history and none of them were on any medications including oral contraceptive pills or over the counter medications

Study measurements

Blood samples were centrifuged within 5 min of collection and were stored frozen at −80 °C pending analysis. All study measurements and analysis were performed in accordance with the relevant guidelines and regulations. Serum T and A were measured by LC/MS/MS on an Acquity UPLC system coupled to a Quattro Premier XE mass spectrometer (Waters, Manchester, UK). Sex hormone binding globulin (SHBG) was measured by an immunometric assay with fluorescence detection on the DPC Immulite 2000 analyzer using the manufacturer's recommended protocol (upper limit of the reference range 2.0 nmol/l). The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG. Serum insulin was assayed using a competitive chemiluminescent immunoassay performed on the manufacturer's DPC Immulite 2000 analyzer (Euro/DPC, Llanberis, UK). The analytical sensitivity of the insulin assay was 2 µU/ml, the coefficient of variation was 6%, and there was no stated cross-reactivity with proinsulin. Plasma glucose was measured using a Synchron LX 20 analyzer (Beckman-Coulter), using the manufacturer's recommended protocol. The coefficient of variation for the assay was 1.2% at a mean glucose value of 5.3 mmol/liter. The insulin resistance was calculated using the HOMA method [HOMA-IR = (insulin × glucose)/22.5]. Anti-Müllerian hormone was measured using a Beckman Coulter Access automated immunoassay. A number of AMH immunoassays have been developed: we used the Beckman Coulter Access automated immunoassay from Beckman Coulter, as studies have shown good correlation between the Gen II ,Elecsys assays and the new Acesss AMH assay(10). 17-OHP was measured in the early morning sample and if on

the higher side of the normogram, congenital adrenal hyperplasia was excluded with ACTH stimulation test. The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG

Collection and handling of saliva samples

This has been detailed previously for the saliva collection and for the salivary androgen measurement methodology7. In brief, participants were asked to spit or drool directly into a 4 mL sealable polystyrene tube and to provide at least 3 mL of saliva. Unstimulated saliva samples were used to avoid any assay interference. The "passive drool" technique was used for the collection of saliva rather than the 'salivette' method. Salivary testosterone and salivary androstendione were measured by LC-MS/MS analysis performed using a Waters Acquity UPLC system coupled to a Waters Xevo TQS mass spectrometer, giving a lower limit of quantification of 5 pmol/L for salT and 6.25 pmol/l for salA with an inter and intra-assay precision coefficient of variation of <4% and <7.5%, respectively.

Statistical analysis

All the study variables were log transformed if they were not normally distributed. After the log transformation we imputed the missing values using an iterative imputation method *missForest* (11) . *missForest* is an implementation of random forest algorithm. It is a non-parametric imputation method, which builds a random forest model for each variable and subsequently uses the model to predict missing values in the variable with the help of observed values. To evaluate androgen levels between PCOS cases and controls, univariate comparative analyses were performed using the non-parametric Mann-Whitney tests on the imputed datasets. Means (standard deviations) or medians (interquartile range) were used to

summarize continuous variables as appropriate while proportions and frequencies were used to summarize categorical variables.

Risk prediction

In logistic regression models, if the sample size is small or if a predictor is strongly associated with one of the possible outcomes the estimated coefficients may be biased. To overcome this issue, we used logistic regression model with Firth's bias-adjusted estimates. The basic idea of the Firth's logistic regression (Firth 1993) is to introduce a more effective score function by adding a term that counteracts the first-order term from the asymptotic expansion of the bias of the maximum likelihood estimation—and the term will go to zero as the sample size increases (12). Model selection with Firth's bias adjustments was done using R package '*logistf*' (12). Firstly, we included all the relevant variables in a model such as age, BMI, waist-circumference, menstrual irregularity (yes/no), use of oral contraceptives (yes/no), serum testosterone, salivary testosterone, serum androstenedione, salivary androstenedione, oestradiol, SHBG, DHEAS, LH, FSH, Prolactin, 17-OHP, FAI and AMH levels. We did not include menstrual disturbances in the model as it is extremely difficult to quantify the extent duration and severity of menstrual disturbances and simply entering a yes/no variable can lead to model overfitting. Next, we used *backward* in *logistf* in R to identify best model from a set of candidate predictor variables by entering predictors based on p value cut-off of 0.05. The variable selection in *logistf* is simply performed by repeatedly calling add 1 or drop 1 methods for *logistf* and is based on penalized likelihood ratio test. In order to assess the stability of the model thus obtained compared this stepwise model based on P-values to a model using *forward* selection. As the apparent predictive performance (performance in the development cohort) usually overestimates the performance in other patients, owing to overfitting and peculiarities in the development cohort (13), we internally validated the model through bootstrapping using package *boot* in R. A bootstrap analysis with 1000 simulations

was performed to compare the measures of effect obtained from the original model with the bootstrapped model.

We assessed model discrimination using area under the receiver operator curve (AUC) in a logistic regression model. Values greater than 0.7 indicate good predictive performance and values greater than 0.8 indicate excellent predictive performance of the model. Goodness-of-fit were assessed using calibration plot and Hosmer-Lemeshow statistics.

In order to calculate an individual patient's risk of having PCOS, we first calculated their prognostic index (14) (PI). To achieve this, the estimated coefficients were multiplied by the values of the predictor variables of the patient and the sum of these multiplications were added to the intercept of the model. Using the PI we then calculated the risk of PCOS as exp(PI)/(1+exp(PI)).

For ease of interpretation we back-transformed the significant variables retained in the model and presented the effect estimates and P-values associated with these. We did a sensitivity analysis using 1) untransformed raw variables with missing values and 2) untransformed raw variables with imputed values to assess model stability.

Results

The anthropometric and hormonal characteristics of women with PCOS and controls from the Hull UK PCOS biobank are shown in **Table 1**. Women with PCOS were younger (P=0.01) had higher BMI (P<0.0001), waist circumference (P<0.0001), and overall, greater levels of all markers indicating hyperandrogenemia compared to controls. Women with PCOS also had significantly higher levels of 17-OHP (P=0.03) and AMH (P<0.0001).

The logistic regression with backward selection model revealed four variables independently associated with PCOS namely, FAI [Beta 0.30(0.12), P=0.008)], 17-OHP [Beta=0.20(0.01), P=0.026], AMH [Beta=0.04(0.01), P<0.0001], and waist-circumference [Beta=0.08(0.02),

P<0.0001] (Table 2). Relaxation and restriction of the removal criterion for backward selection to P<0.20 and P<0.10, respectively, did not change the final model. Similar results were also seen in a model with forward selection. A bootstrap analysis with 1000 simulations indicated minimal bias and model optimism in estimated effect sizes **(Supplementary table 1).** Bootstrap estimates of several discrimination indices to quantify the model are presented in **Supplementary Table 2**. The optimism corrected estimate of the Somers' D was 0.81 (Supplementary Table 2) with a corresponding bias corrected c-statistic of ((1+0.8193)/2) = 0.90. The model with the 4 predictor variables had a high discrimination ability with a c-statistics of AUC=0.91 (0.88-0.95). The AUCs for FAI, AMH, 17-OHP and WC were 0.81 (0.75-0.87), 0.75 (0.68-0.82), 0.59 (0.51-0.67) and 0.91(0.88-0.95), respectively (Figures 1a-1e). Model calibration was assessed using the Hosmer–Lemeshow statistics and a calibration plot (Fig 2). The model shows good calibration with Hosmer–Lemeshow chi-squared of 3.7865, and a p-value of 0.87.

Based on the penalised regression coefficient, we calculated a prognostic index (PI) for each of the PCOS cases using the formula (-9.77 + (0.07*WC) + (0.04*AMH) + (0.3*FAI) + (0.01*17OHP)) and calculated a risk score for each case of PCOS with formula exp(PI)/(1+exp(PI))*100. The metabolic profile of women with PCOS in the top 3 quartiles (q1-q3) of this risk score (classified as low-risk score) was compared with the metabolic profile of PCOS women in the bottom quartile (q4) of the risk score (classified as high risk score). PCOS women with a high risk score, had a worse metabolic profile with significantly higher 2-hour glucose (P=0.01), baseline insulin (P=0.0003), TG (P=0.0005) and CRP (<0.0001) levels and lower HDL-C levels (P=0.02), as compared to those with a low-risk score **(Table 3)**. We have constructed a mobile phone application for easy usage of this risk score in clinical settings. (**Supplementary Figure 1**)

Discussion

The diagnosis of PCOS is often challenging given the wide range of hormonal markers and derived indices used to measure hyperandrogenism and variations in clinical presentations. We developed and internally validated a simple 4-variable model (i.e., FAI, 17-OHP, AMH and waist circumference) for predicting the risk of having PCOS in clinical settings. This model showed good discrimination ability and good calibration. Each of the 4 variables reported in our model have been previously associated with PCOS (6, 9, 15-17).

In line with differential diagnoses of conditions causing hyperandrogenism in females, in this we measured 17-OHP levels to rule out a potential diagnosis of non-classical congenital adrenal hyperplasia (NCCAH), which is another disorder of hyperandrogenism. The normal levels of 17-OHP in females are well defined and the baseline mean level of 17-OHP in those with NCCAH is around 20 ng/ml (60 nmol/L) (18). In this study the PCOS women had mean baseline 17-OHP levels of 1.6ng/ml (5 nmol/L) safely ruling out NCCAH. A baseline 17-OHP cut-off of 2ng/ml is suggested for the screening NCCAH, however, it is not unusual for patients with PCOS to have levels of 17-OHP higher than this cut-off. A study by Pall et.al (19) comparing the 17-OHP levels in PCOS and NCCAH showed that 25% of lean patients with PCOS, 21% of obese patients with PCOS, and 7% of controls had basal 17-OHP levels above the cut-off level 2 ng/ml. Patients with PCOS have also been showed to have higher 17-OHP levels as compared to those without PCOS (17). For example, 17-OHP levels have been shown to be significantly higher in pre- and postmenopausal PCOS women as compared to controls(15, 16) , with the levels being highest in those with severe phenotype of PCOS (15) Interestingly, a subgroup of PCOS patients with exaggerated 17-OHP response to GnRH agonist presented with severe hyperandrogenemia, glucose-stimulated β-cell insulin secretion, and worse insulin resistance (20). The excess 17-OHP in patients with PCOS is thought to be of the result of excess stimulation of theca interna cells- by luteinizing hormone (LH)(15). In this study, for the first time, we showed that 17-OHP are independently

associated with PCOS, after adjustments of FAI, AMH and waist circumference. However, the discriminatory capacity of 17-OHP to detect PCOS were small and if not readily available, can be excluded from the model.

We also show that AMH was independently associated with PCOS diagnosis after adjustments for FAI, WC and 17-OHP. AMH is produced in the granulosa cells by the preantral and small antral follicles and it appears to inhibit the action of FSH on aromatase, and therefore, it contributes to the development of a single follicle for ovulation (21) . AMH is elevated in PCOS due to the increased count of small antral follicle and increased secretion of AMH per follicle (22) . We have recently shown that those with raised AMH have up to 4-fold increased risk of having PCOS(8) . It has also been suggested that serum AMH reflects ovarian size in PCOS patients and can be used as surrogate for transvaginal ultrasound in the diagnosis of PCOS (9).

The associations of FAI and waist circumference with PCOS are well-documented in the literature (6, 23) . Waist circumference, a measure of central adiposity, is a marker of severity of PCOS and has been suggested to be a better surrogate of glucose and lipid metabolism in PCOS than the disease status per se (23) . Menstrual dysfunction is a common symptom in PCOS and is a consequence of anovulation. Ovulatory dysfunction can also be seen in women who have regular menstrual cycle (24) (25) and as a result menstrual history alone is insufficient in defining PCOS. The prevalence of non-specific menstrual dysfunction in high in women, especially in adolescent population where it can be as high has 30%, 1 year post menarche (26). It is difficult to identify real anovulation related menstrual dysfunction and many of the women are already on oral-contraceptive pills which makes it difficult to ascertain the history of menstrual dysfunction. Hence we decided not to include this variable in our model.

In this study we showed that those with a high risk score derived from a model, which included waist circumference, FAI , AMH and 17-OHP, had a poor metabolic profile, as evidenced by a higher 2h-glucose , raised TG levels, basal insulin, CRP and lower HDL-cholesterol. Thus, this risk score can not only identify patients who are at high risk of PCOS, but it can also risk stratify patients and identify those who are more likely to experience adverse PCOS-related metabolic outcomes. Collectively, the four variables in our model capture the full spectrum of PCOS, wherein, FAI reflects androgens excess, AMH grasps the ovarian size and/or follicle count, 17-OHP represents the alteration in LH-FSH ratio and WC indicates the presence of metabolic abnormalities in PCOS. FAI, free testosterone and SHBG are routinely measured as a part of the diagnostic workup for PCOS, while 17-OHP is measured as per the endocrine society guidelines to rule out congenital adrenal hyperplasia. AMH measurement is routinely done in these patients as a part of their fertility work-up, hence, no additional testing is required when this model is used. On the other hand, using this model, may eliminate the need for testing additional androgen markers such as salivary testosterone and androstenedione, and hence, it can reduce the cost associated with these tests.

Given the high prevalence of metabolic syndrome in PCOS, guidelines issued by the American College of Obstetricians and Gynaecologists and the Endocrine Society recommend that all women with PCOS should undergo screening for impaired glucose tolerance and dyslipidaemia with a 2 hour 75 g oral glucose tolerance test and fasting lipid profile upon diagnosis, with repeat screening of each test every 2-5 years (27). However, there is no guidance on how to identify women who are at high risk for developing metabolic syndrome and not all women with PCOS get metabolic syndrome screening in primary or secondary care. The advantage of this scoring system is that it may assist in the diagnosis of PCOS and highlights those women who are at high risk of developing metabolic syndrome to help prevent future metabolic complications.

Our study has several limitations. Our 4 variable risk model for PCOS is not externally validated. We have attempted to overcome this problem by bootstrapping, and the effects size of our model indicate very little optimism and good calibration. However, further external validation of this model in an ethnically diverse population is warranted. Secondly, although the mean levels of 17-OHP in our study are significantly lower than those seen in patients with CAH and NCCAH, it is possible to have NCCAH with a normal 17-OHP level. The sample size of our study was modest with 111 PCOS and 67 controls. However, this a very well characterized cohort of PCOS- and control women which measures all the androgen and related markers (including salivary markers) and unique in the sense that all the participants had classical PCOS whereby all the three criteria for diagnosis of PCOS namely oligomenorrhea, hyperandrogenism, and PCO morphology on ultrasound were met. Nonetheless, this model will need further validation in large prospective cohorts from different ethnicities for its validation. Another limitation of our study is that all the patients in our study had Classical PCOS oligomenorrhea, hyperandrogenism, and PCOS as designated in the Rotterdam criteria. The other sub-phenotypes include ovulatory PCOS (hyperandrogenism, PCO, and regular menstrual cycles), non-PCO PCOS (oligomenorrhea, hyperandrogenism, and normal ovaries) and mild PCOS (oligomenorrhea, PCO, and normal androgens). Hence we were not able to evaluate our model for the other 3 phenotypes. However, the classical PCOS phenotype represents the largest subgroup of patients with PCOS, with an estimated prevalence of up to 80% amongst the PCOS population (28) and this model can be generalised to the largest subgroup of the PCOS population. The strength of the study on the other hand is that it provides a simple 4 variable model and calculator which can predict the risk of PCOS in clinical settings and identify those with unfavourable PCOS-related metabolic consequences. Furthermore, this study consisted of a homogenous

group of Caucasian women who fulfilled Rotterdam diagnostic criteria of PCOS, thus providing a robust database for model development.

Conclusions

In summary, we have developed a simple model consisting of FAI, 17-OHP, AMH and waist circumference for risk prediction and risk stratification in PCOS, with these variables previously associated with PCOS This model will have to be externally validated in populations across different ethnicities before a widespread clinical application.

Disclosure Statement: All the authors do not have anything relevant to disclose with respect to this manuscript.

Funding statement: This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

1. Azziz R, Woods KS, Reyna R, Key TJ, Knochenhauer ES, Yildiz BO. The prevalence and features of the polycystic ovary syndrome in an unselected population. J Clin Endocrinol Metab. 2004;89(6):2745-9.
2. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Fertil Steril. 2018.
3. Bani Mohammad M, Majdi Seghinsara A. Polycystic Ovary Syndrome (PCOS), Diagnostic Criteria, and AMH. Asian Pac J Cancer Prev. 2017;18(1):17-21.
4. Rotterdam EA-SPCWG. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. Fertil Steril. 2004;81(1):19-25.
5. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril. 2009;91(2):456-88.
6. Keevil BG, Adaway J, Fiers T, Moghetti P, Kaufman JM. The free androgen index is inaccurate in women when the SHBG concentration is low. Clin Endocrinol (Oxf). 2018;88(5):706-10.

7. Williamson K, Gunn AJ, Johnson N, Milsom SR. The impact of ethnicity on the presentation of polycystic ovarian syndrome. Aust N Z J Obstet Gynaecol. 2001;41(2):202-6.
8. Sathyapalan T, Al-Qaissi A, Kilpatrick ES, Dargham SR, Atkin SL. Anti-Mullerian hormone measurement for the diagnosis of polycystic ovary syndrome. Clin Endocrinol (Oxf). 2018;88(2):258-62.
9. Matsuzaki T, Munkhzaya M, Iwasa T, Tungalagsuvd A, Yano K, Mayila Y, et al. Relationship between serum anti-Mullerian hormone and clinical parameters in polycystic ovary syndrome. Endocr J. 2017;64(5):531-41.
10. Pearson K, Long M, Prasad J, Wu YY, Bonifacio M. Assessment of the Access AMH assay as an automated, high-performance replacement for the AMH Generation II manual ELISA. Reprod Biol Endocrinol. 2016;14:8.
11. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112-8.
12. Heinze G, Ploner M, Beyea J. Confidence intervals after multiple imputation: combining profile likelihood information from logistic regressions. Stat Med. 2013;32(29):5062-76.
13. Bell S, Dekker FW, Vadiveloo T, Marwick C, Deshmukh H, Donnan PT, et al. Risk of postoperative acute kidney injury in patients undergoing orthopaedic surgery--development and validation of a risk score and effect of acute kidney injury on survival: observational cohort study. BMJ. 2015;351:h5639.
14. Royston P, Altman DG. Visualizing and assessing discrimination in the logistic regression model. Stat Med. 2010;29(24):2508-20.
15. Tsang BK, Taheri A, Ainsworth L, Downey BR. Secretion of 17 alpha-hydroxyprogesterone, androstenedione, and estrogens by porcine granulosa and theca interna cells in culture. Can J Physiol Pharmacol. 1987;65(9):1951-6.
16. Piouka A, Farmakiotis D, Katsikis I, Macut D, Gerou S, Panidis D. Anti-Mullerian hormone levels reflect severity of PCOS but are negatively influenced by obesity: relationship with increased luteinizing hormone levels. Am J Physiol Endocrinol Metab. 2009;296(2):E238-43.
17. Maas KH, Chuan SS, Cook-Andersen H, Su HI, Duleba A, Chang RJ. Relationship between 17-hydroxyprogesterone responses to human chorionic gonadotropin and markers of ovarian follicle morphology in women with polycystic ovary syndrome. J Clin Endocrinol Metab. 2015;100(1):293-300.
18. Ambroziak U, Kepczynska-Nyk A, Kurylowicz A, Malunowicz EM, Wojcicka A, Miskiewicz P, et al. The diagnosis of nonclassic congenital adrenal hyperplasia due to 21-hydroxylase deficiency, based on serum basal or post-ACTH stimulation 17-hydroxyprogesterone, can lead to false-positive diagnosis. Clin Endocrinol (Oxf). 2016;84(1):23-9.
19. Pall M, Azziz R, Beires J, Pignatelli D. The phenotype of hirsute women: a comparison of polycystic ovary syndrome and 21-hydroxylase-deficient nonclassic adrenal hyperplasia. Fertil Steril. 2010;94(2):684-9.
20. Pasquali R, Patton L, Pocognoli P, Cognigni GE, Gambineri A. 17-hydroxyprogesterone responses to gonadotropin-releasing hormone disclose distinct phenotypes of functional ovarian hyperandrogenism and polycystic ovary syndrome. J Clin Endocrinol Metab. 2007;92(11):4208-17.
21. Dewailly D, Andersen CY, Balen A, Broekmans F, Dilaver N, Fanchin R, et al. The physiology and clinical utility of anti-Mullerian hormone in women. Hum Reprod Update. 2014;20(3):370-85.
22. Bhide P, Dilgil M, Gudi A, Shah A, Akwaa C, Homburg R. Each small antral follicle in ovaries of women with polycystic ovary syndrome produces more antimullerian hormone than its counterpart in a normal ovary: an observational cross-sectional study. Fertil Steril. 2015;103(2):537-41.

23. Pazderska A, Kyaw Tun T, Phelan N, McGowan A, Sherlock M, Behan L, et al. In women with PCOS, waist circumference is a better surrogate of glucose and lipid metabolism than disease status per se. Clin Endocrinol (Oxf). 2018;88(4):565-74.
24. Norman RJ, Dewailly D, Legro RS, Hickey TE. Polycystic ovary syndrome. Lancet. 2007;370(9588):685-97.
25. Lujan ME, Chizen DR, Pierson RA. Diagnostic criteria for polycystic ovary syndrome: pitfalls and controversies. J Obstet Gynaecol Can. 2008;30(8):671-9.
26. Legro RS, Lin HM, Demers LM, Lloyd T. Rapid maturation of the reproductive axis during perimenarche independent of body composition. J Clin Endocrinol Metab. 2000;85(3):1021-5.
27. Salley KE, Wickham EP, Cheang KI, Essah PA, Karjane NW, Nestler JE. Glucose intolerance in polycystic ovary syndrome--a position statement of the Androgen Excess Society. J Clin Endocrinol Metab. 2007;92(12):4546-56.
28. Clark NM, Podolski AJ, Brooks ED, Chizen DR, Pierson RA, Lehotay DC, et al. Prevalence of Polycystic Ovary Syndrome Phenotypes Using Updated Criteria for Polycystic Ovarian Morphology: An Assessment of Over 100 Consecutive Women Self-reporting Features of Polycystic Ovary Syndrome. Reprod Sci. 2014;21(8):1034-43.

Figure legends

Figure 1: Graphs showing AUC for for Antimullerian Hormone (AMH); Free Androgen Index (FAI); 17-OHP, 17α-Hydroxyprogesterone and waist circumference (WC) individually and combined. The c-statistics for the complete model was 0.91(0.88-0.95)

Figure 2: Graph showing a plot of the expected event probabilities against the predicted event probabilities with a perfect predictive ability shown on the graph by the diagonal red straight line at 45

Table 1: Baseline characteristics and hormonal parameters of women with and without PCOS in the Hull UK PCOS biobank

	PCOS (n=67) Median (IQR)	Control (n=111) Median (IQR)	P-value*
Age	27.68 (11)	29.92 (11)	0.01
BMI	34.15 (9.9)	26.86 (6.2)	<0.0001
Waist Circumference (cm)	101 (21.2)	78 (14.5)	<0.0001
Testosterone (nmol/L)	1.30 (0.85)	0.94 (0.45)	<0.0001
Salivary Androstenedione (pmol/L)	146.4 (88.65)	185.8 (112.4)	0.0002
Oestradiol (pmol/)L	190 (295)	180 (165)	0.43
SHBG (nmol/L)	27 (18)	47 (31)	<0.0001
TSH (mU/L)	1.9 (1.2)	1.5 (1.1)	0.03
DHEAS (umol/l)	5.2 (3.8)	4.6 (4)	0.04
Androstenedione (nmol/L)	9.5 (5.8)	7.3 (4.4)	<0.0001
Prolactin	250 (165)	260 (126)	0.68
LH	6.2 (5.6)	4.1 (4.3)	0.003
FSH	4.9 (2.7)	5.5 (3.2)	0.09
FAI	4.5 (4.8)	1.98 (1.4)	<0.0001
17-OHP (nmol/L)	4.4 (3)	3.9 (2)	0.03
AMH	37 (41)	18.1 (24.5)	<0.0001

* P-values based on Mann–Whitney U test

AMH, Antimullerian Hormone; BMI, Body Mass Index; DEAS, Dehydroepiandrosterone, FAI, Free Androgen Index; FSH, Follicle Stimulating Hormone; LH, Luteinizing hormone; SHBG, sex hormone-binding globulin; TSH. Thyroid Stimulating Hormone; 17-OHP, 17α-Hydroxyprogesterone.

Table 2: Independent predictors of PCOS from penalized logistic regression model

	GLM based Logistic regression estimates			Firth's Penalised logistic regression estimates		
	Beta	SE	P-value	Beta	SE	P-value
FAI	0.32	0.12	0.008	0.30	0.12	0.008
17-OHP	0.21	0.09	0.026	0.20	0.09	0.026
AMH	0.04	0.01	<0.0001	0.04	0.01	<0.0001
Waist Circumference	0.08	0.01	0.0003	0.07	0.02	<0.0001

AMH, Antimullerian Hormone; FAI, Free Androgen Index; 17-OHP, 17α-Hydroxyprogesterone.

Table 3: Metabolic Profile of PCOS patients with low (q1-q3) and high risk (q4) score based on penalised regression model

	PCOS Cases with low risk score (Q1-Q3) (n=84)	PCOS cases with high-risk score (Q4) (n=27)	P-value
	Mean (SD)	Mean (SD)	
Baseline Glucose	4.73 (0.48)	5.19 (1.91)	0.41
2-Hour Glucose	5.51 (1.30)	7.73 (3.39)	0.01
Insulin	13.01 (8.27)	27.25 (21.98)	0.0003

LDL-c	2.88 (0.90)	2.99 (0.73)	0.94
HDL-c	1.26 (0.32)	1.10 (0.18)	0.02
TG	1.26 (0.62)	2.46 (2.22)	0.0005
TC	4.72 (0.98)	4.95 (0.95)	0.19
CRP	3.64 (3.73)	8.45 (6.61)	<0.0001

LDL, Low density lipoprotein; HDL, high density lipoprotein; TG, Triglycerides; TC: Total Cholesterol and CRP, C-reactive protein*P-values based on Mann–Whitney U test

Figure 1: Receiver Operator curves for Antimullerian Hormone (AMH); Free Androgen Index (FAI); 17-OHP, 17α-Hydroxyprogesterone and waist circumference (WC)

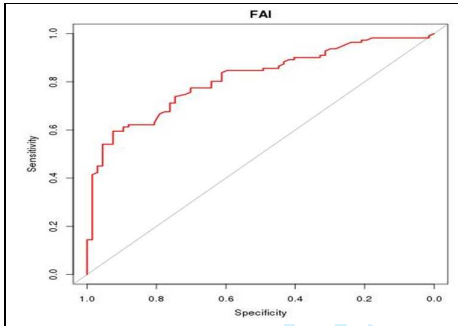


Fig 1a AUC=0.81 (0.75-0.87)

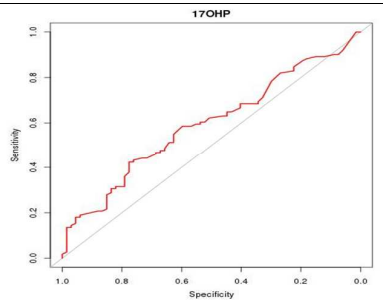


Fig 1b , AUC=0.59 (0.51-0.67)

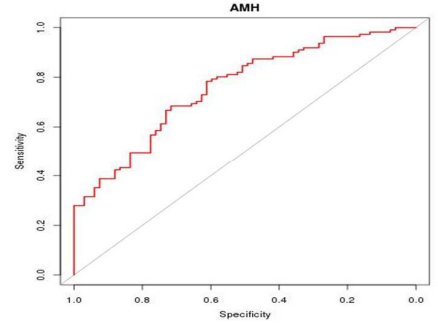


Fig 1c AUC=0.75 (0.68-0.82)

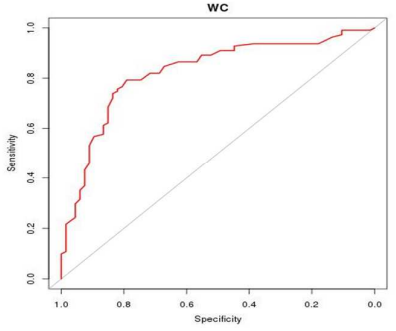


Fig 1d , AUC=0.82 (0.76-0.88)

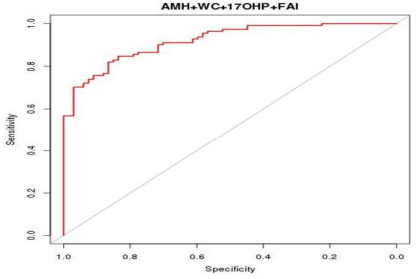
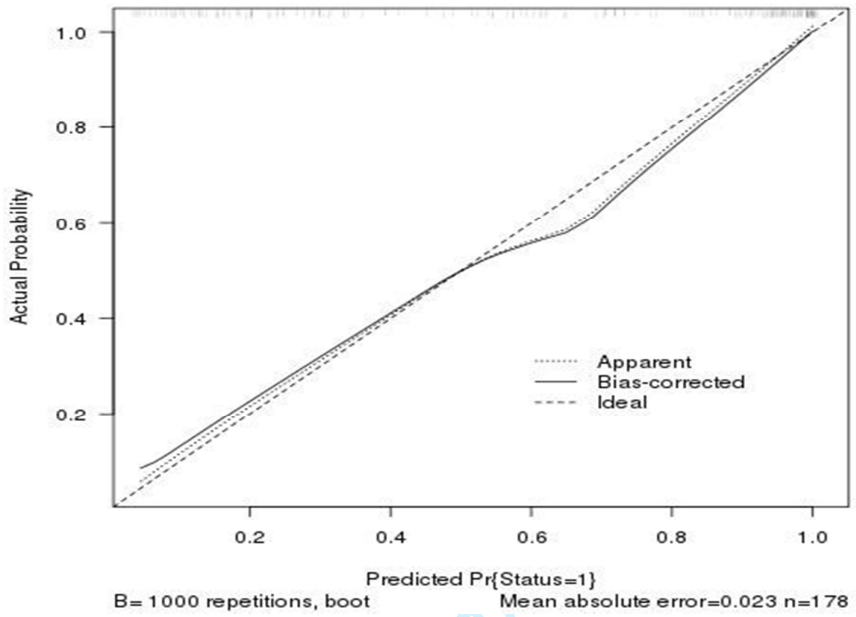


Fig 1e AUC =0.91(0.88-0.95)

Figure 2: Calibration plot for the 4-variable logistic regression model



Supplementary Table 1: Estimates of bias for logistic regression model from 1000 bootstrap estimates

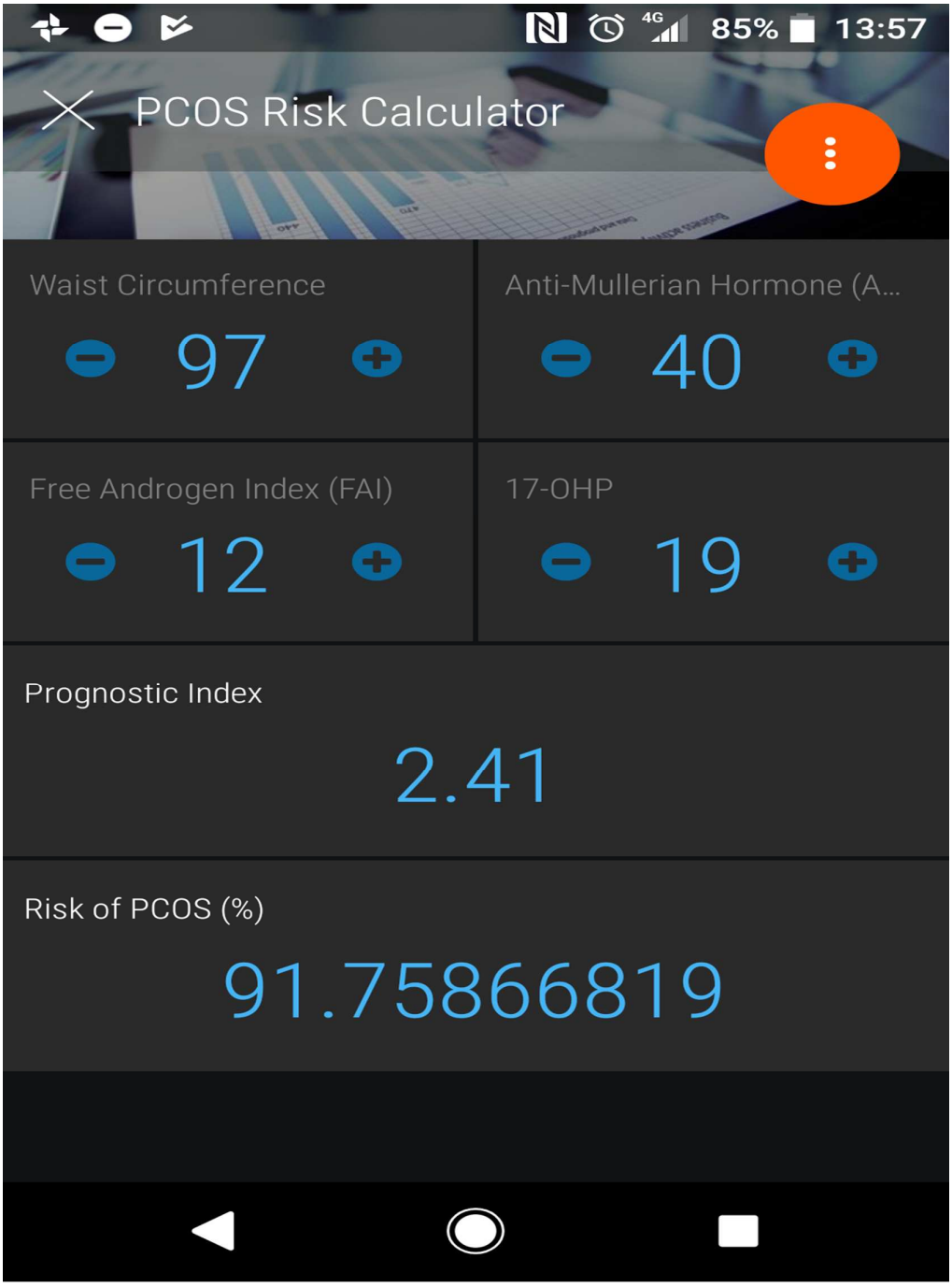
	Original	Bias
FAI	0.30	0.029
17-OHP	0.20	0.002
AMH	0.08	0.002
Waist Circumference	0.04	0.001

AMH, Antimullerian Hormone; FAI, Free Androgen Index; 17-OHP, 17α-Hydroxyprogesterone.

Supplementary Table 2: Bootstrap estimates of several discrimination indexes to quantify the model

	index.orig	**training**	**test**	**optimism**	**index.corrected**	**n**
Dxy	0.8384	0.8442	0.8252	0.019	0.8193	1000
R2	0.6297	0.6461	0.616	0.0301	0.5995	1000
Intercept	0	0	-0.003	0.003	-0.003	1000
Slope	1	1	0.9185	0.0815	0.9185	1000
Emax	0	0	0.0195	0.0195	0.0195	1000
D	0.6147	0.6386	0.5963	0.0422	0.5725	1000
U	-0.0112	-0.0112	0.0034	-0.0146	0.0034	1000
Q	0.6259	0.6498	0.5929	0.0569	0.5691	1000
B	0.1161	0.1107	0.1209	-0.0101	0.1262	1000
g	3.5059	3.8221	3.4621	0.36	3.1459	1000
gp	0.3936	0.3959	0.3892	0.0067	0.3869	1000

Supplementary Figure 1: Phone based PCOS risk calculator application



We are thankful to the reviewers for their useful comments. This has certainly increased the clarity of the manuscript and has improved the overall quality.

Reviewer: 1

1-No enough information regarding the women without PCOS was given. They need to explain the method via they selected the control women.

Response: We are thankful to reviewers for pointing this out. We have addressed this by including the following in the methods section **(lines 95-98)**

"All of the control women had regular periods, no clinical or biochemical hyperandrogenism, no polycystic ovaries on ultrasound, no significant background medical history and none of them were on any medications including oral contraceptive pills or over the counter medications.

2-FAI should be defined for non-endocrinologist readers.

Thanks. We have now addressed this by including the following in the methods section. **(lines 120-121)**

"*The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG*"

3-17OHP is a well-known marker in the diagnosis of NCAH due to 21OH deficiency. It should be measured early in the morning and during early follicular phase in order to rule out the adrenal contribution. So, the details about the measurement of 17OHP are necessary and it should be given in Method section.

Response: We have now included a study measurements section which addresses this. 17-OHP was measured as a part of early morning sample and if on the higher side was excluded CAH was excluded with ACTH stimulation test. We have included following paragraphs in the methods sections**. (lines 99-132)**

Blood samples were centrifuged within 5 min of collection and were stored frozen at −80 °C pending analysis. All study measurements and analysis were performed in accordance with the relevant guidelines and regulations. Serum T and A were measured by LC/MS/MS on an Acquity UPLC system coupled to a Quattro Premier XE mass spectrometer (Waters, Manchester, UK). Sex hormone binding globulin (SHBG) was measured by an immunometric assay with fluorescence detection on the DPC Immulite 2000 analyzer using the manufacturer's recommended protocol (upper limit of the reference range 2.0 nmol/l). The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG. Serum insulin was assayed using a competitive chemiluminescent immunoassay performed on the

manufacturer's DPC Immulite 2000 analyzer (Euro/DPC, Llanberis, UK). The analytical sensitivity of the insulin assay was 2 μU/ml, the coefficient of variation was 6%, and there was no stated cross-reactivity with proinsulin. Plasma glucose was measured using a Synchron LX 20 analyzer (Beckman-Coulter), using the manufacturer's recommended protocol. The coefficient of variation for the assay was 1.2% at a mean glucose value of 5.3 mmol/liter. The insulin resistance was calculated using the HOMA method [HOMA-IR = (insulin × glucose)/22.5]. Anti-Müllerian hormone was measured using a Beckman Coulter Access automated immunoassay. A number of AMH immunoassays have been developed: we used the Beckman Coulter Access automated immunoassay from Beckman Coulter, as studies have shown good correlation between the Gen II ,Elecsys assays and the new Acesss AMH assay (Reprod Biol Endocrinol. 2016 Kylie Pearson). 17-OHP was measured in the early morning sample and if on the higher side of the normogram, congenital adrenal hyperplasia was excluded with ACTH stimulation test. The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG

Collection and handling of saliva samples

This has been detailed previously for the saliva collection and for the salivary androgen measurement methodology7. In brief, participants were asked to spit or drool directly into a 4 mL sealable polystyrene tube and to provide at least 3 mL of saliva. Unstimulated saliva samples were used to avoid any assay interference. The "passive drool" technique was used for the collection of saliva rather than the 'salivette' method. Salivary testosterone and salivary androstendione were measured by LC-MS/MS analysis performed using a Waters Acquity UPLC system coupled to a Waters Xevo TQS mass spectrometer, giving a lower limit of quantification of 5 pmol/L for salT and 6.25 pmol/l for salA with an inter and intra-assay precision coefficient of variation of <4% and <7.5%, respectively.

4-As far as I know there is no perfect method for measurement of AMH. Do the authors think that this a problem in the development of such a model?

A number of AMH immunoassays have been developed and they show good correlation with each other. We have used the new Beckman Coulter Access automated immunoassay from Beckman Coulter which shows good correlation with the other commonly used assays. We have addressed this comment by including the following it the methods section. **(Lines 115-119)**

A number of AMH immunoassays have been developed: we used the Beckman Coulter Access automated immunoassay from Beckman Coulter, as studies have shown good correlation between the Gen II ,Elecsys assays and the new Acesss AMH assay (Reprod Biol Endocrinol. 2016 Kylie Pearson).

5-According to the model developed by the authors, the following hormones should be measured:17-OHP, AMH, SHBG and total testosterone, the last two for the calculation of FAI. I am wondering whether this model is feasible or not in clinical practice.

Response: FAI, free testosterone and SHBG are routinely measured as a part of diagnostic workup for PCOS, while 17-OHP is measured as per the Endocrine society guidelines to rule out congenital adrenal hyperplasia. AMH on the other hand is routinely done in these patients as a part of their fertility work-up; therefore, no additional testing is required when this model is used. Hence using this model does not require any additional testing and can be used in clinical practice.

6-Do the authors make a comment about the cost-effectiveness of the model?

Response: FAI, free testosterone and SHBG are routinely measured as a part of diagnosis for PCOS, while 17-OHP is measured as per the Endocrine society guidelines to rule out Congenital adrenal hyperplasia. AMH on the other hand is routinely done in these patients as a part of fertility work-up; therefore, no additional testing is required when this model is used. On the other hand, using this model, eliminates the need for testing additional androgen markers such as salivary testosterone, androstenedione and hence, it can reduce the cost associated with these tests. We have not formally measured the cost-effectiveness of this model - but believe that no additional costs will be required to implement this model as all the 3 tests are done routinely in women with PCOS in most centres.

We have address the **comments 5 and 6** by including the following in the discussion section **(lines 273-279)**

"*FAI, free testosterone and SHBG are routinely measured as a part of the diagnostic workup for PCOS, while 17-OHP is measured as per the Endocrine society guidelines to rule out congenital adrenal hyperplasia. AMH measurement is routinely done in these patients as a part of their fertility work-up, hence, no additional testing is required when this model is used. On the other hand, using this model, may eliminate the need for testing additional androgen markers such as salivary testosterone and androstenedione, and hence, it can reduce the cost associated with these tests".*

7-They suggest that this simple 4 variable model identifies unfavourable PCOS-related metabolic consequences. Can they say that we will be able to decide whether metabolic parameters in PCOS patients should be measured or not when we used the model? If no, what about the advantages of the model?

Response: Given the high prevalence of metabolic syndrome in PCOS, guidelines issued by the American College of Obstetricians and Gynecologists and the Endocrine Society recommend that all women with PCOS undergo screening for impaired glucose tolerance and dyslipidemia with a 2 hour 75 g oral glucose tolerance test and have a fasting lipid profile upon diagnosis, with repeat screening of each test every 2-5 years. However, there is no guidance on how to identify women who are at high risk for developing metabolic syndrome and not all women with PCOS get metabolic syndrome screening in primary or secondary care. There is an ongoing debate if we should be screening all pregnant women for Type 2 diabetes. The advantage of this scoring system is that it can highlight the women who are at high risk of developing metabolic syndrome to help prevent future metabolic complications in those women.

We have addressed the above comment by including the following paragraph in the discussion: **(lines 280-289)**

"*Given the high prevalence of metabolic syndrome in PCOS, guidelines issued by the American College of Obstetricians and Gynaecologists and the Endocrine Society recommend that all women with PCOS should undergo screening for impaired glucose tolerance and dyslipidaemia with a 2 hour 75 g oral glucose tolerance test and fasting lipid profile upon diagnosis, with repeat screening* of each test every 2-5 years. *(Kelsey E. S. Salley The Journal of Clinical Endocrinology & Metabolism 2007). However, there is no guidance on how to identify women who are at high risk for developing metabolic syndrome and not all women with PCOS get* metabolic syndrome *screening in primary or secondary care. The advantage of this scoring system is that it may assist in the diagnosis of PCOS and highlights those women who are at high risk of developing metabolic syndrome to help prevent future metabolic complications."*

8- There are some minor grammatical errors such as; page 10, line 198, showed should be shown.

Response: Thanks we have now corrected this.

Reviewer 2

1) **The study is quite small, with only 111 PCOS and 67 controls. Hence, it is unclear how powerful will the analysis be.**

Response: We acknowledge that the sample size in our study is small. However, this a very well characterized cohort of PCOS- which measures all the androgen and related markers (including salivary markers) and unique in the sense that all the participants had classical PCOS whereby all the three criteria for diagnosis of PCOS oligomenorrhea, hyperandrogenism, and PCOS on ultrasound were met. A post-hoc power calculation (using the function *pwr.f2.test* in the R package *pwr)* shows that in order to replicate this model (with adjusted r-squared of 0.35) a sample size of 60 cases and 60 controls will give us more than 90% power to replicate the findings. So our discovery cohort seems to be adequately powered. However, we do acknowledge the need to replicate and validate this model in larger prospective cohorts and have addressed this in the paper- by including the following **(Lines 295-301)**

"The sample size of our study was modest with 111 PCOS and 67 controls. However, this a very well characterized cohort of PCOS- and control women which measures all the androgen and related markers (including salivary markers) and unique in the sense that all the participants had classical PCOS whereby all the three criteria for diagnosis of PCOS namely oligomenorrhea, hyperandrogenism, and PCO morphology on ultrasound were met. Nonetheless, this model will need further validation in large prospective cohorts from different ethnicities for its validation".

2) It is unclear what the phenotypes of the PCOS patients that were included are. As the investigators know, the Rotterdam criterion denotes 4 phenotypes (A-D). The predictors/diagnostic markers of each of these phenotypes vary. The investigators need to consider PCOS phenotype in their exercise.

Response: The reviewer has rightly pointed out that there are 4 subtypes of PCOS based on Rotterdam criterion. The Rotterdam and AE-PCOS Society criteria recognize at least 4 unique clinical phenotypes: (A) Classical PCOS (oligomenorrhea, hyperandrogenism, and PCO), (B) Ovulatory PCOS (hyperandrogenism, PCO, and regular menstrual cycles), and (C) Non-PCO PCOS (oligomenorrhea, hyperandrogenism, and normal ovaries) (D) mild PCOS (oligomenorrhea, PCO, and normal androgens)

All the patients in or study had the phenotype A- Classical PCOS- and the diagnosis of PCOS in our study was based on all three diagnostic criteria of the Rotterdam consensus, namely clinical and biochemical evidence of hyperandrogenism (Ferriman-Gallwey score >8; free androgen index >4, total testosterone >1.5 nmol/l), oligomenorrhea or amenorrhea and polycystic ovaries on transvaginal ultrasound.

The prevalence of classical PCOS is reported to be up to 70% of PCOS population (Reprod Sci. 2014 Nina M. Clark) and hence the findings of our study can be generalised to the most prevalent phenotype of PCOS. Since our study population consisted of phenotype A, we are not able to do a subgroup analysis with various sub-groups of PCOS and we have acknowledged this limitation in our discussion section by adding the following paragraph. **(lines 302-310)**

"Another limitation of our study is that all the patients in our study had Classical PCOS oligomenorrhea, hyperandrogenism, and PCOS as designated in the Rotterdam criteria. The other sub-phenotypes include ovulatory PCOS (hyperandrogenism, PCO, and regular menstrual cycles), non-PCO PCOS (oligomenorrhea, hyperandrogenism, and normal ovaries) and mild PCOS (oligomenorrhea, PCO, and normal androgens). Hence we were not able to evaluate our model for the other 3 phenotypes. However, the classical PCOS phenotype represents the largest subgroup of patients with PCOS, with an estimated prevalence of up to 80% amongst the PCOS population (Reprod Sci. 2014 Nina M. Clark) and this model can be generalised to the largest subgroup of the PCOS population"

3) The attempt to develop predictive models for PCOS is not new. The problem is that if the predictive model is based on elements that require invasive testing (i.e. blood tests) or tests that are part of the diagnosis (i.e. androgens), then the predictive model is really a diagnostic model not a predictive model. As such the value of this exercise from a public health or predictive point of view is very limited. Perhaps the investigators are attempting to determine what the minimal elements are for the diagnosis of PCOS – although this is already guided by the diagnostic criteria.

Response: We agree with the reviewer that we proposed to determine the minimal elements required for the diagnosis of PCOS. The current guidelines do not specify which androgen markers should take precedence over the others and also do not include the new marker AMH that has been consistently shown to be associated with PCOS in several recent studies. Also, there is no specific guidance on how to identify women who are at high risk for developing metabolic syndrome in the future and not all women with PCOS get metabolic syndrome screening in primary or secondary care. So the overarching objective of the PCOS risk score was to identify the best available androgen and hormonal markers to assist in the diagnosis of PCOS (when other causes of hyperandrogenemia have been ruled out) and to identify women who are at higher risk of metabolic complications. We have now addressed this comment by including the following in the discussion section. **(lines 280-289)**

"*Given the high prevalence of metabolic syndrome in PCOS, guidelines issued by the American College of Obstetricians and Gynaecologists and the Endocrine Society recommend that all women with PCOS should undergo screening for impaired glucose tolerance and dyslipidaemia with a 2 hour 75 g oral glucose tolerance test and fasting lipid profile upon diagnosis, with repeat screening* of each test every 2-5 years. *(Kelsey E. S. Salley The Journal of Clinical Endocrinology & Metabolism 2007). However, there is no guidance on how to identify women who are at high risk for developing metabolic syndrome and not all women with PCOS get* metabolic syndrome *screening in primary or secondary care. The advantage of this scoring system is that it may assist in the diagnosis of PCOS and highlights those women who are at high risk of developing metabolic syndrome to help prevent future metabolic complications.""*

4) **Minor**:

a. It would be helpful to the reader if the investigators, even briefly, described the methods used hormonal measures, rather than fully referring to the reference #8.

Response: Now we have included the complete methods used in the hormonal measures. **(Lines 99-132)**

Blood samples were centrifuged within 5 min of collection and were stored frozen at −80 °C pending analysis. All study measurements and analysis were performed in accordance with the relevant guidelines and regulations. Serum T and A were measured by LC/MS/MS on an Acquity UPLC system coupled to a Quattro Premier XE mass spectrometer (Waters, Manchester, UK). Sex hormone binding globulin (SHBG) was measured by an immunometric assay with fluorescence detection on the DPC Immulite 2000 analyzer using the manufacturer's recommended protocol (upper limit of the reference range 2.0 nmol/l). The free androgen index (FAI) was calculated as the total testosterone × 100/SHBG. Serum insulin was assayed using a competitive chemiluminescent immunoassay performed on the manufacturer's DPC Immulite 2000 analyzer (Euro/DPC, Llanberis, UK). The analytical sensitivity of the insulin assay was 2 μU/ml, the coefficient of variation was 6%, and there was no stated cross-reactivity with proinsulin. Plasma glucose was measured using a

Synchron LX 20 analyzer (Beckman-Coulter), using the manufacturer's recommended protocol. The coefficient of variation for the assay was 1.2% at a mean glucose value of 5.3 mmol/liter. The insulin resistance was calculated using the HOMA method [HOMA-IR = (insulin × glucose)/22.5]. Anti-Müllerian hormone was measured using a Beckman Coulter Access automated immunoassay; between run precision was <3% across the range measured. 17-OHP was measured in the early morning sample and if on the higher side of normogram CAH was excluded with ACTH stimulation test.

Collection and handling of saliva samples

This has been detailed previously for the saliva collection and for the salivary androgen measurement methodology7. In brief, participants were asked to spit or drool directly into a 4 mL sealable polystyrene tube and to provide at least 3 mL of saliva. Unstimulated saliva samples were used to avoid any assay interference. The "passive drool" technique was used for the collection of saliva rather than the 'salivette' method. Salivary testosterone and salivary androstendione were measured by LC-MS/MS analysis performed using a Waters Acquity UPLC system coupled to a Waters Xevo TQS mass spectrometer, giving a lower limit of quantification of 5 pmol/L for salT and 6.25 pmol/l for salA with an inter and intra-assay precision coefficient of variation of <4% and <7.5%, respectively