

Identification and replication of prediction models for ovulation, pregnancy and live birth in infertile women with polycystic ovary syndrome

Hongying Kuang^{1,2}, Susan Jin², Karl R. Hansen³, Michael P. Diamond⁴, Christos Coutifaris⁵, Peter Casson⁶, Gregory Christman⁷, Ruben Alvero⁸, Hao Huang², G. Wright Bates⁹, Rebecca Usadi¹⁰, Scott Lucidi¹¹, Valerie Baker¹², Nanette Santoro⁸, Esther Eisenberg¹³, Richard S. Legro¹⁴, and Heping Zhang^{2,*} for the Reproductive Medicine Network

¹Department of Obstetrics and Gynecology, First Affiliated Hospital and Heilongjiang University of Chinese Medicine, Harbin 150040, PR China
²Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06520, USA ³Department of Obstetrics and Gynecology, University of Oklahoma College of Medicine, Oklahoma City, OK 73104, USA ⁴Department of Obstetrics and Gynecology, Georgia Regents University, Augusta, GA 30912, USA ⁵Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA 19104, USA
⁶Department of Obstetrics and Gynecology, University of Vermont, Burlington, VT 05405, USA ⁷Department of Obstetrics and Gynecology, Shands Hospital, University of Florida, Gainesville, FL 32611, USA ⁸Department of Obstetrics and Gynecology, University of Colorado Denver, Aurora, CO 80045, USA ⁹Department of Obstetrics and Gynecology, University of Alabama Birmingham, Birmingham, AL, USA ¹⁰Department of Obstetrics and Gynecology, Carolinas Medical Center, Charlotte, NC, USA ¹¹Department of Obstetrics and Gynecology, Virginia Commonwealth University, Richmond, VA, USA ¹²Department of Obstetrics and Gynecology, Stanford University Medical Center, Stanford, CA, USA ¹³Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA ¹⁴Department of Obstetrics and Gynecology, Penn State College of Medicine, Hershey, PA 17033, USA

*Correspondence address. Department of Biostatistics, Yale University School of Public Health, New Haven, CT 06520, USA.
Fax: +1-203-785-5145; E-mail: heping.zhang@yale.edu

Submitted on January 21, 2015; resubmitted on June 10, 2015; accepted on June 18, 2015

STUDY QUESTION: Can we build and validate predictive models for ovulation and pregnancy outcomes in infertile women with polycystic ovary syndrome (PCOS)?

SUMMARY ANSWER: We were able to develop and validate a predictive model for pregnancy outcomes in women with PCOS using simple clinical and biochemical criteria particularly duration of attempting conception, which was the most consistent predictor among all considered factors for pregnancy outcomes.

WHAT IS KNOWN ALREADY: Predictive models for ovulation and pregnancy outcomes in infertile women with polycystic ovary syndrome have been reported, but such models require validation.

STUDY DESIGN, SIZE, AND DURATION: This is a secondary analysis of the data from the Pregnancy in Polycystic Ovary Syndrome I and II (PPCOS-I and -II) trials. Both trials were double-blind, randomized clinical trials that included 626 and 750 infertile women with PCOS, respectively. PPCOS-I participants were randomized to either clomiphene citrate (CC), metformin, or their combination, and PPCOS-II participants to either letrozole or CC for up to five treatment cycles.

PARTICIPANTS/MATERIALS, SETTING, AND METHODS: Linear logistic regression models were fitted using treatment, BMI, and other published variables as predictors of ovulation, conception, clinical pregnancy, and live birth as the outcome one at a time. We first evaluated previously reported significant predictors, and then constructed new prediction models. Receiver operating characteristic (ROC) curves were constructed and the area under the curves (AUCs) was calculated to compare performance using different models and data. Chi-square tests were used to examine the goodness-of-fit and prediction power of logistic regression model.

MAIN RESULTS AND THE ROLE OF CHANCE: Predictive factors were similar between PPCOS-I and II, but the two participant samples differed statistically significantly but the differences were clinically minor on key baseline characteristics and hormone levels. Women in PPCOS-II

had an overall more severe PCOS phenotype than women in PPCOS-I. The clinically minor but statistically significant differences may be due to the large sample sizes. Younger age, lower baseline free androgen index and insulin, shorter duration of attempting conception, and higher baseline sex hormone-binding globulin significantly predicted at least one pregnancy outcome. The ROC curves (with AUCs of 0.66–0.76) and calibration plots and chi-square tests indicated stable predictive power of the identified variables (P -values ≥ 0.07 for all goodness-of-fit and validation tests).

LIMITATIONS, REASONS FOR CAUTION: This is a secondary analysis. Although our primary objective was to confirm previously reported results and identify new predictors of ovulation and pregnancy outcomes among PPCOS-II participants, our approach is exploratory and warrants further replication.

WIDER IMPLICATIONS OF THE FINDINGS: We have largely confirmed the predictors that were identified in the PPCOS-I trial. However, we have also revealed new predictors, particularly the role of smoking. While a history of ever smoking was not a significant predictor for live birth, a closer look at current, quit, and never smoking revealed that current smoking was a significant risk factor.

STUDY FUNDING/COMPETING INTEREST(S): The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grants U10 HD27049, U10 HD38992, U10HD055925, U10 HD39005, U10 HD33172, U10 HD38998, U10 HD055936, U10 HD055942, and U10 HD055944; and U54-HD29834. Heilongjiang University of Chinese Medicine Grants 051277 and B201005. R.S.L. reports receiving consulting fees from Euroscreen, AstraZeneca, Clarus Therapeutics, and Takeda, and grant support from Ferring, Astra Zeneca, and Toba. K.R.H. reports receiving grant support from Roche Diagnostics and Ferring Pharmascience. G.C. reports receiving Honorarium and grant support from Abbvie Pharmaceuticals and Bayer Pharmaceuticals. M.P.D. holds equity from Advanced Reproductive Care Inc. and DS Biotech, receives fees from Advanced Reproductive Care Inc., Actamax, Auxogyn, ZSX Medical, Halt Medical, and Neomed, and receives grant support from Boehringer-Ingelheim, Abbott, and BioSante, Ferring Pharmaceuticals, and EMD Serono. H.Z. receives research support from the Chinese 1000-scholar plan. Others report no disclosures other than NIH grant support.

TRIAL REGISTRATION NUMBER: PPCOS-I and -II were respectively registered at Clinicaltrials.gov: NCT00719186 and NCT00719186.

Key words: mathematical modeling / calibration / prediction / receiver operating characteristic / polycystic ovaries / pregnancy / conception / live birth

Introduction

Polycystic ovary syndrome (PCOS) is characterized by oligomenorrhea and ovulatory dysfunction, hyperandrogenism and polycystic ovary morphology (Taylor *et al.*, 1997; Rotterdam Group, 2003; Azziz *et al.*, 2006). It affects 6.5–10% of reproductive age women, and is the most common cause of anovulatory infertility (Knochenhauer *et al.*, 1998; Diamanti-Kandarakis *et al.*, 1999; Asunción *et al.*, 2000). Approximately 40% of women with PCOS present with a clinical complaint of infertility (Teede *et al.*, 2010), and 90–95% of anovulatory women who visit infertility clinics have PCOS (Sirmans and Pate, 2013). Despite its high prevalence, the optimal treatment of infertile women with PCOS is still surrounded by much controversy (Thessaloniki Group, 2008). The recommended first-line treatment for ovulation induction has long been clomiphene citrate (CC), but recently we determined that letrozole was markedly more effective than CC in inducing ovulation and live births in infertile women with PCOS (Legro *et al.*, 2014a,b).

Randomized, clinical trials can point out an overall superior therapy, but there are often individual patient factors that may play a role in the success of any specific treatment. There are many treatment choices for an infertile woman with PCOS; the selection of which one is best for any individual patient remains a clinical dilemma. In fact, not all women with PCOS will ovulate in response to CC or letrozole, and of those that do only half will achieve a pregnancy within up to five cycles (Legro *et al.*, 2014a,b). Given these challenges, there is significant interest in identifying predictors that may help physicians identify appropriate infertility treatments for women with PCOS and to counsel them regarding their prognosis. We sought to determine which patient attributes

were most likely to predict a successful outcome in the context of Reproductive Medicine Network-conducted clinical trials. It was reported previously that baseline free androgen index, baseline proinsulin, and duration of attempting conception all influenced the outcomes of ovulation, conception, pregnancy, and live birth (Rausch *et al.*, 2009).

In this study, we use data from our recently completed investigation on Pregnancy in Polycystic Ovary Syndrome (PPCOS-II) that compared letrozole to CC for infertility in women with PCOS. We attempt to confirm previously reported significant predictors based on an earlier Pregnancy in Polycystic Ovary Syndrome (PPCOS-I) study data (Rausch *et al.*, 2009), as well as construct new prediction models that may better fit the recent data in predicting successful pregnancies.

Materials and Methods

Study design

This is a secondary analysis of the PPCOS-I and -II trial data. Both trials were multi-center, randomized and double-blind. PPCOS-I compared CC, metformin and their combination in 626 women with PCOS. In PPCOS-II, 750 infertile women with PCOS were randomized to letrozole or CC for up to five cycles of ovulation induction (Legro *et al.*, 2014a,b). PCOS was defined by modified Rotterdam criteria and detailed inclusion and exclusion criteria have been reported previously (Legro *et al.*, 2007, 2014a,b). The institutional review board at each study site approved the protocol and all subjects (males/females) gave written informed consent. The study was monitored by a Data and Safety and Monitoring Board.

Demographics and a full infertility and medical history were obtained using standardized forms from all participants. Blood pressure, acne assessment,

hirsutism assessment and transvaginal ultrasound examinations were all performed at baseline. Fasting blood was obtained for hormonal assays. Samples were batched and analyzed at the Ligand Assay & Analysis Core Laboratory at the University of Virginia. The free androgen index (FAI) was calculated from the formula: (total testosterone in nmol/l / SHBG in nmol/l) \times 100 (Sodergard et al., 1982). Glucose levels were determined on a glucose analyzer using the glucose oxidase method. Laboratory tests including complete blood count, liver and renal function tests were performed at baseline and end of study as safety labs at each site.

In the PPCOS-II trial (Legro et al., 2014a,b), among the 376 women randomized to letrozole, 331 ovulated, 154 conceived, 117 achieved clinical pregnancy and 103 delivered at least one live born baby. In contrast, 288 of the 374 women randomized to CC ovulated; 103 conceived, 81 achieved clinical pregnancy, and 72 delivered a baby. Letrozole was superior to CC in all pregnancy outcomes. BMI was negatively associated with all pregnancy outcomes. In the PPCOS-I trial (Legro et al., 2007), live birth was observed in 47 of 209 women randomized to CC, 15 of 208 women randomized to metformin and 56 of 209 women who were randomized to combination therapy. CC and the combination of CC with metformin were superior to metformin alone for all outcomes. Live birth was significantly more likely for women with a BMI below 30 kg/m² compared with women with a BMI over 30 kg/m².

Data analysis

First, we evaluated the four prediction models previously built in the PPCOS-I population to predict success of ovulation, conception, pregnancy and live birth (Rausch et al., 2009) in the PPCOS-II population. The criteria for defining ovulation and conception have changed. For PPCOS-I, ovulation was defined as having a midluteal serum progesterone level >5 ng/ml, and conception was defined as any positive serum human chorionic gonadotrophin level. For PPCOS-II (Legro et al., 2014a,b), ovulation was defined as having a midluteal serum progesterone level >3 ng/ml and conception was defined as having a serum level of human chorionic gonadotrophin >10 mIU per milliliter. For both studies, clinical pregnancy was defined as an intrauterine pregnancy with fetal heart motion as detected by transvaginal ultrasound examination; live birth was defined as the delivery of a viable infant. Logistic regression was used for the analysis of these binary outcomes.

Since our first goal was to replicate the four published models, we considered the same set of baseline clinical and laboratory predictors as in Rausch et al. (2009). These comprised treatment, age, weight, body mass index (BMI), hirsutism score, race, waist measurement, waist/hip ratio, ethnic group, duration of attempting conception, pregnancy history, prior loss of pregnancy, prior parity, history of smoking, baseline total testosterone, baseline free androgen index (FAI), baseline glucose, baseline insulin, baseline proinsulin, baseline SHBG and baseline white blood cell count. We categorized BMI (<30 , $30-34$ and ≥ 35 kg/m²), age (≤ 34 , >34 years), proinsulin (<23 , ≥ 23 pmol/l), FAI (<10 , ≥ 10), hirsutism (<8 , $8-15$, ≥ 16) and duration of attempting conception (<1.5 , ≥ 1.5 years). For history of smoking, as an additional step, we believe it is more appropriate and informative to consider current, quit and never smoking.

We then used backward deletion to rebuild four logistic regression models for ovulation, conception, pregnancy and live birth using the PPCOS-II data. As in Rausch et al. (2009), both age and treatment variables were retained in the final models regardless of their statistical significance. For each covariate selected from the initial backward deletion procedure, interaction effects between treatment (CC or letrozole) and the selected covariates were then assessed, and those interactions that were significant at the 0.10 level or lower were included in the final models. For each outcome, the final models are presented in a table with odds ratios and the corresponding 90% confidence intervals for all covariates following the tables in Rausch et al. (2009). The interaction effect of the treatment with a covariate was

displayed by presenting the odds ratios for the treatment effect given the value of the covariate.

To compare the performance of the prediction models built from PPCOS-I and II data on the PPCOS-II data, we estimated the probability of live birth for each woman in the PPCOS-II study first using the PPCOS-I model and then using the PPCOS-II model. Then, we took the average of the estimated probabilities of live birth within the groups defined by the predictors in the PPCOS-I models. Figure 1 uses different colors to represent the ranges of the estimated probabilities and their differences.

To assess the predictive power of various models and the consistency of the predictors across PPCOS-I and II data, we constructed receiver operating characteristic (ROC) curves and calculated the areas under them (AUC). In this way, we assessed the PPCOS-I-generated models used in Rausch et al. (2009) applied to the PPCOS-II data and the PPCOS-II-generated models applied to the PPCOS-I data.

To assess the predictive power of our models in terms of both the precision and reproducibility, we adopted an intuitive calibration approach used in related studies (Veltman-Verhulst et al., 2012), performed the Hosmer–Lemeshow test (Hosmer and Lemeshow, 1980) for goodness-of-fit of the logistic regression models, and Pearson's chi-square test for independent validation. We stratified our samples into quintiles (i.e. very low, low, intermediate, high and very high success rates) according to the predicted model, obtained the average of the estimated success probabilities from the participants within each strata, and calibrated those averages with the observed percentages of success from the corresponding strata.

All analyses were performed with SAS software, version 9.2 (SAS Institute, Cary, NC, USA) or R (www.r-project.org). Trial Registration: ClinicalTrials.gov number: NCT00068861 (PPCOS-I) and NCT00719186 (PPCOS-II).

Results

Baseline characteristics

Table 1 summarizes baseline demographic and clinical variables, and biomarkers. There were no significant differences in baseline characteristics between the two treatment arms in the PPCOS-II data (Legro et al., 2014a,b); the same was true of the PPCOS-I sample. Thus, we did not present the treatment-specific data in Table 1. Instead, we examined potential overall differences between PPCOS-I and II participants without considering the treatment arms. PPCOS-I women were younger ($P = 0.001$) than those in PPCOS-II. Although BMI was not significantly different between the two samples, age (28.1 versus 28.9, $P = 0.001$), waist circumference (102.5 versus 105.9, $P = 0.002$) and waist to hip ratio (0.86 versus 0.90, $P < 0.001$) were significantly lower in PPCOS-I participants than PPCOS-II participants. Hirsutism score was also lower in women in PPCOS-I than those in PPCOS-II (14.4 versus 17.0, $P < 0.001$). Moreover, group distributions of race/ethnicity ($P < 0.001$) differed between the two studies, such that PPCOS-II women were more likely to be White (78.7%) and less likely to be African-American (13.3%) or Hispanic or Latino (17.1%), in comparison to PPCOS-I (White: 69.8%; African-American: 17.5%; Hispanic or Latino: 26.2%) (Table 1). The means for all hormonal markers except SHBG (Table 1) were considerably and significantly ($P \leq 0.01$) higher in PPCOS-I participants than those in PPCOS-II participants.

Predictive value of individual characteristics

As presented in Table 2, despite a lack of difference in ovulation rate by age group ($P > 0.05$), younger women had better outcomes in terms of conception (odds ratio [OR] = 2.06, Wald 90% confidence intervals

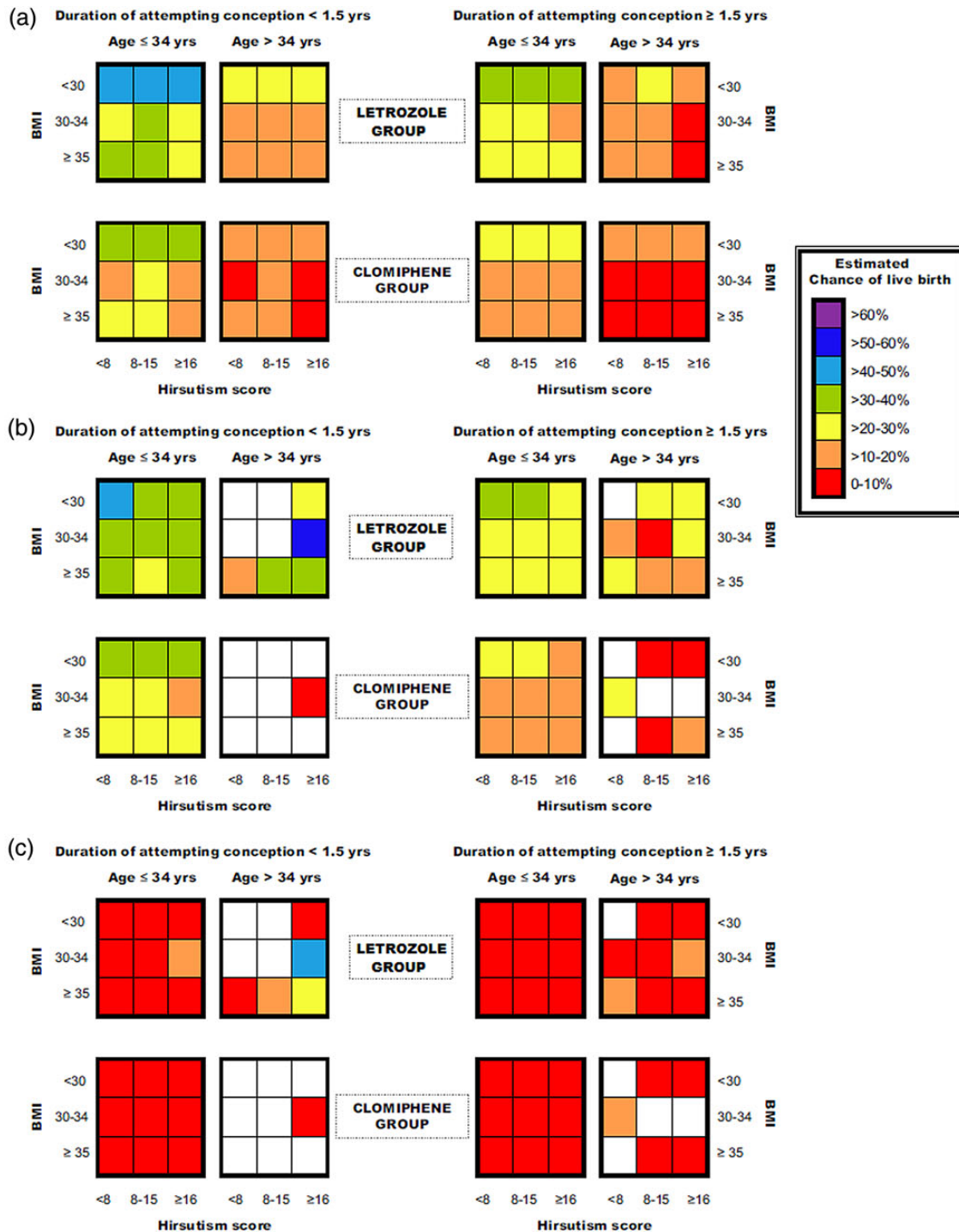


Figure 1 Group averaged estimates of probability in the PPCOS-II subgroups determined by treatment, duration of attempting conception, age, BMI and hirsutism score: **(a)** Based on the PPCOS-I model; **(b)** Based on the PPCOS-II model; and **(c)** The difference of the estimates between (a) and (b). Colors represent the levels of probability (a and b) or differences in probability (c) and the empty subgroups were represented with white cells. The blanks in (b) and (c) indicate that there were no PPCOS-II participants in those subgroups, and the estimates are not available.

[CI] = 1.20–3.53), pregnancy (OR = 1.92, CI = 1.06–3.50) and live birth (OR = 2.51, CI = 1.26–5.00). History of a prior pregnancy loss was not associated with the likelihood of ovulation, but was significantly associated with conception rate (OR = 1.53, CI = 1.11–2.12). This

latter finding was unique to the PPCOS-II data. Proinsulin level was not predictive of any of the outcomes, although there was a trend toward better outcomes with lower proinsulin levels. FAI was predictive of ovulation (OR = 1.64, CI = 1.14–2.36) and live birth (OR = 1.60,

Table I Clinical parameters and baseline biomarkers used as covariates in prediction models.

Parameter	PPCOS-II (N = 750)	PPCOS-I (N = 626)	P-value*
Age	28.9 ± 4.3	28.1 ± 4.0	0.001
Weight (kg)	94.8 ± 26.3	94.3 ± 24.7	0.715
BMI (kg/m ²)	35.1 ± 9.3	35.2 ± 8.7	0.849
<30	245/750 (32.7)	179/625 (28.6)	0.057
30–34	127/750 (16.9)	135/625 (21.6)	
≥35	378/750 (50.4)	311/625 (49.8)	
Waist circumference (cm)	105.9 ± 20.4	102.5 ± 19.6	0.002
Waist/hip ratio	0.902 ± 0.103	0.864 ± 0.091	<0.001
Hirsutism score	17.0 ± 8.5	14.4 ± 7.9	<0.001
<8	97/750 (12.9)	121/626 (19.3)	<0.001
8–15	250/750 (33.3)	262/626 (41.9)	
≥16	403/750 (53.7)	243/626 (38.8)	
Race			<0.001
White	590/750 (78.7)	435/623 (69.8)	
Black or African-American	100/750 (13.3)	109/623 (17.5)	
Asian	24/750 (3.2)	17/623 (2.7)	
American Indian or Alaska Native	7/750 (0.9)	72/623 (11.6)	
Native Hawaiian or Pacific Islander	2/750 (0.3)	1/623 (0.2)	
Mixed race	27/750 (3.6)		
Ethnic group			<0.001
Not Hispanic or Latino	622/750 (82.9)	462/626 (73.8)	
Hispanic or Latino	128/750 (17.1)	164/626 (26.2)	
Length of attempting conception (months)	41.7 ± 37.8	40.4 ± 35.8	0.503
Prior pregnancy	273/750 (36.4)	210/626 (33.6)	0.269
Prior pregnancy loss	174/750 (23.2)	138/626 (22.0)	0.610
Prior live birth	148/750 (19.7)	113/626 (18.1)	0.428
History of smoking	317/750 (42.3)	247/626 (39.5)	0.291
Total testosterone (ng/dl)	55.0 ± 28.8	62.0 ± 28.6	<0.001
Free androgen index	7.8 ± 5.9	9.5 ± 6.7	<0.001
Glucose (mg/dl)	86.0 ± 12.6	89.0 ± 17.4	<0.001
Insulin (μU/ml)	19.3 ± 27.0	23.0 ± 26.6	0.011
Proinsulin (pmol/l)	18.0 ± 14.4	24.9 ± 25.8	<0.001
SHBG (nmol/l)	33.9 ± 23.1	29.7 ± 18.1	<0.001
White blood cells (10 ³ /μl)	7.3 ± 1.9	7.3 ± 2.0	0.660

Data are presented as mean ± SD or frequencies (%). There are some missing values.
SHBG, sex hormone-binding globulin.

*Comparison for all patients between PPCOS-I and PPCOS-II studies.

CI = 1.08–2.39) but not conception or pregnancy, although a trend was observed. Hirsutism score was not selected in the initial model for ovulation from the PPCOS-I data and hence was not re-evaluated in the PPCOS-II data, but for the other three outcomes, the middle group had a modest increase in success, which was significant for conception (OR = 1.46, CI = 1.08–1.97) and pregnancy (OR = 1.40, CI = 1.02–1.92), and nearly significant for live birth (OR = 1.30, CI = 0.93–1.81). A shorter duration of attempting conception had little relationship to the ovulation rate, but was significantly positively associated with the other outcomes (conception: OR = 1.35,

CI = 1.003–1.82; pregnancy: OR = 1.50, CI = 1.10–2.05; live birth: OR = 1.39, CI = 1.01–1.92).

Building prediction models on baseline characteristics

Next, we used a backward deletion procedure to build prediction models for the four outcomes from the PPCOS-II data (Table III). As pointed out earlier, age was included in our models regardless of its significance. Unlike what we did in Table II, we considered age and BMI on a

Table II Odds ratios, and 90% Wald confidence limits obtained using PPCOS-I models to determine predictive factors for success applied to PPCOS-II data.

Effect	Ovulation	Conception (achieved pregnancy)	Pregnancy (clinical pregnancy)	Live birth
Baseline BMI ≥ 35 (kg/m ²)				
Clomiphene (reference)	1.0	1.0	1.0	1.0
Letrozole	2.42 (1.58, 3.70)	2.83 (1.90, 4.21)	2.47 (1.60, 3.80)	2.21 (1.40, 3.48)
Baseline BMI 30–34 (kg/m ²)				
Clomiphene (reference)	1.0	1.0	1.0	1.0
Letrozole	2.45 (1.15, 5.20)	1.12 (0.59, 2.10)	1.18 (0.59, 2.38)	1.13 (0.52, 2.46)
Baseline BMI <30 (kg/m ²)				
Clomiphene (reference)	1.0	1.0	1.0	1.0
Letrozole	2.44 (1.06, 5.61)	1.50 (0.99, 2.30)	1.29 (0.83, 1.99)	1.39 (0.89, 2.17)
Treatment				
Clomiphene (reference)	1.0	1.0	1.0	1.0
Letrozole	2.42 (1.70, 3.44)	1.89 (1.44, 2.48)	1.64 (1.23, 2.19)	1.59 (1.18, 2.15)
BMI (kg/m ²)				
≥ 35 (reference)	1.0	1.0	1.0	1.0
30–34	1.12 (0.71, 1.77)	1.00 (0.67, 1.49)	0.91 (0.59, 1.41)	0.83 (0.52, 1.33)
<30	2.57 (1.59, 4.13)	1.61 (1.16, 2.24)	1.45 (1.02, 2.06)	1.63 (1.14, 2.34)
Age (years)				
>34 (reference)	1.0	1.0	1.0	1.0
≤ 34	0.89 (0.48, 1.65)	2.06 (1.21, 3.53)	1.92 (1.06, 3.50)	2.51 (1.26, 5.00)
History of prior loss	1.45 (0.94, 2.26)	1.53 (1.11, 2.12)	N.A.	N.A.
Baseline proinsulin (pmol/l)				
≥ 23 (reference)	1.0	1.0	1.0	1.0
<23	1.40 (0.95, 2.07)	1.40 (0.97, 2.02)	1.21 (0.81, 1.79)	1.15 (0.76, 1.74)
Baseline free androgen index				
≥ 10 (reference)	1.0	1.0	1.0	1.0
<10	1.64 (1.14, 2.36)	1.23 (0.88, 1.73)	1.41 (0.97, 2.05)	1.60 (1.08, 2.39)
Hirsutism score				
≥ 16 (reference)	N.A.	1.0	1.0	1.0
8–15		1.46 (1.08, 1.97)	1.40 (1.02, 1.92)	1.30 (0.93, 1.81)
<8		1.08 (0.70, 1.66)	1.06 (0.68, 1.67)	1.12 (0.70, 1.79)
Duration of attempting conception				
≥ 1.5 years (reference)	1.0	1.0	1.0	1.0
<1.5 years	1.16 (0.78, 1.74)	1.35 (1.003, 1.82)	1.50 (1.10, 2.05)	1.39 (1.01, 1.92)

N.A. refers to the variables not selected in the PPCOS-I model.

continuous scale. Older age had an overall negative relationship to pregnancy (OR = 0.96, CI = 0.93–1.00) and live birth (OR = 0.94, CI = 0.91–0.98). BMI was negatively associated with ovulation (OR = 0.97, CI = 0.95–0.99) and conception (OR = 0.98, CI = 0.96–0.996), but did not remain for pregnancy and live birth. Hispanic or Latino ethnicity remained in the model for a lower rate of ovulation (OR = 0.48, CI = 0.32–0.71), but was removed for the other three outcomes. Smoking remained in the model for a lower ovulation rate (OR = 0.70, CI = 0.50–0.99), but not for the other three outcomes. History of prior pregnancy increased the likelihood of pregnancy (OR = 1.45, CI = 1.06–1.99) and live birth (OR = 1.49, CI = 1.08–2.06), but was removed for the prediction models of ovulation and conception. History of prior

loss remained only for conception (OR = 1.57, CI = 1.13–2.18), but not for the other three outcomes.

Biomarkers

Again, as presented in Table III, testosterone was negatively associated with conception (OR = 0.99, CI = 0.99–0.996), pregnancy (OR = 0.99, CI = 0.98–0.99) and live birth (OR = 0.99, CI = 0.98–0.99), but it was removed from the model for ovulation. Attempting conception longer reduced the odds of conception (OR = 0.99, CI = 0.99–0.995), pregnancy (OR = 0.99, CI = 0.98–0.99) and live birth (OR = 0.99, CI = 0.99–0.996), but it was also removed from the model for

Table III Odds ratios (90% confidence intervals) from PPCOS-II models to determine predictive factors for success.

Effect	Ovulation	Conception (achieved pregnancy)	Pregnancy (clinical pregnancy)	Live birth
Treatment				
Clomiphene (reference)	1.0	1.0	1.0	1.0
Letrozole	2.40 (1.69, 3.42)	1.90 (1.45, 2.51)	1.62 (1.21, 2.18)	1.55 (1.14, 2.10)
Age (years)	1.02 (0.98, 1.06)	0.98 (0.95, 1.02)	0.96 (0.93, 0.999)	0.94 (0.91, 0.98)
Baseline BMI (kg/m ²)	0.97 (0.95, 0.99)	0.98 (0.96, 0.996)	N.A.	N.A.
Ethnic group				
Not Hispanic or Latino (reference)	1.0	N.A.	N.A.	N.A.
Hispanic or Latino	0.48 (0.32, 0.71)	N.A.	N.A.	N.A.
History of smoking				
Never smoked (reference)	1.0	N.A.	N.A.	N.A.
Current or former smoker	0.70 (0.50, 0.99)	N.A.	N.A.	N.A.
History of prior pregnancy	N.A.	N.A.	1.45 (1.06, 1.99)	1.49 (1.08, 2.06)
History of prior loss	N.A.	1.57 (1.13, 2.18)	N.A.	N.A.
Baseline testosterone (ng/dl)	N.A.	0.99 (0.99, 0.996)	0.99 (0.98, 0.99)	0.99 (0.98, 0.99)
Baseline proinsulin (pmol/l)	0.99 (0.97, 0.997)	N.A.	N.A.	N.A.
Baseline free androgen index	0.95 (0.93, 0.98)	N.A.	N.A.	N.A.
Baseline glucose (mg/dl)	0.98 (0.97, 0.998)	0.99 (0.98, 0.999)	N.A.	N.A.
Baseline SHBG (nmol/l)	N.A.	1.01 (1.002, 1.02)	1.01 (1.01, 1.02)	1.02 (1.01, 1.02)
Duration of attempting conception	N.A.	0.99 (0.99, 0.995)	0.99 (0.98, 0.99)	0.99 (0.99, 0.996)

Backwards selection was performed.

N.A. indicating that the specific covariate in the row was not included in the prediction model of the outcome in the column.

SHBG, sex hormone-binding globulin.

ovulation. SHBG was positively associated with conception (OR = 1.01, CI = 1.002–1.02), pregnancy (OR = 1.01, CI = 1.01–1.02) and live birth (OR = 1.02, CI = 1.01–1.02), and again it was removed from the model for ovulation. We should note that all ORs appear very close to 1 because of the small unit in the variable, and a one unit increase of these variables did not lead to a large magnitude change of the ORs. Proinsulin (OR = 0.99, CI = 0.97–0.997), and FAI (OR = 0.95, CI = 0.93–0.98), were negatively associated with ovulation, but removed from the models for the other three outcomes. Glucose remained for ovulation (OR = 0.98, CI = 0.97–0.998), and conception (OR = 0.99, CI = 0.98–0.999), but not for the other two outcomes.

Following Rausch et al. (2009), we divided the 750 PPCOS-II participants into 72 subgroups determined by treatment, duration of attempting conception, age, BMI and hirsutism score, of which 16 subgroups had no participants. Then, as displayed in Fig. 1 we estimated the probability of live birth in each subgroup. The estimates in Fig. 1a and b were derived from the PPCOS-I and II models, respectively. It can be seen clearly from the top-left panel of Fig. 1a that the nine subgroups who received letrozole, had attempted conception less than 1.5 years, and were 34 years of age or younger had the overall highest chance of success. Within these nine subgroups, the variations from BMI and hirsutism are relatively minor. On the other hand, the age effect was clear from the colors in the two top-left panels of Fig. 1a, and so was the effect of duration of attempting conception (e.g. the first and third panels in the top row). Although the PPCOS-I and II models selected different predictors, their estimated probabilities of live birth in the PPCOS-II participants

were similar as most of the differences between the estimates were below 10%. See Fig. 1c.

As a *post hoc* analysis, we examined whether extreme values in any of the biomarkers were indicative of negative pregnancy outcomes. Although PPCOS-II was not powered to adequately address this issue, Fig. 2 indicates that the chance of delivering live births is diminished with very high FAI (≥ 20), glucose (≥ 110 mg/dl), proinsulin (≥ 55 pmol/l), insulin (≥ 70 μ U/ml) or testosterone (≥ 130 ng/dl). Specifically, 2 out of 36 women who had FAI ≥ 20 , 1 out of 17 women who had glucose ≥ 110 mg/dl, 1 out of 17 women who had proinsulin ≥ 55 pmol/l, 1 out of 12 women who had insulin ≥ 70 μ U/ml, and 0 out of 15 women who had testosterone ≥ 130 ng/dl delivered live birth. The estimates of these ORs were less than one-third but unreliable due to the small number of events (0 or 1), and hence the details were omitted here.

ROC curves

For convenience, we refer to the models built from the PPCOS-I and II data as the PPCOS-I and II models, respectively. Figure 3 displays the ROC curves when the PPCOS I models were used to predict the outcomes in the PPCOS-I and II data and PPCOS-II models were used to predict the outcomes in the PPCOS-I and II data. When PPCOS-I models were used to predict the outcomes in the PPCOS-I data (model 1), AUCs were the largest for conception (0.74), pregnancy (0.76), and live birth (0.76), and the second largest for ovulation

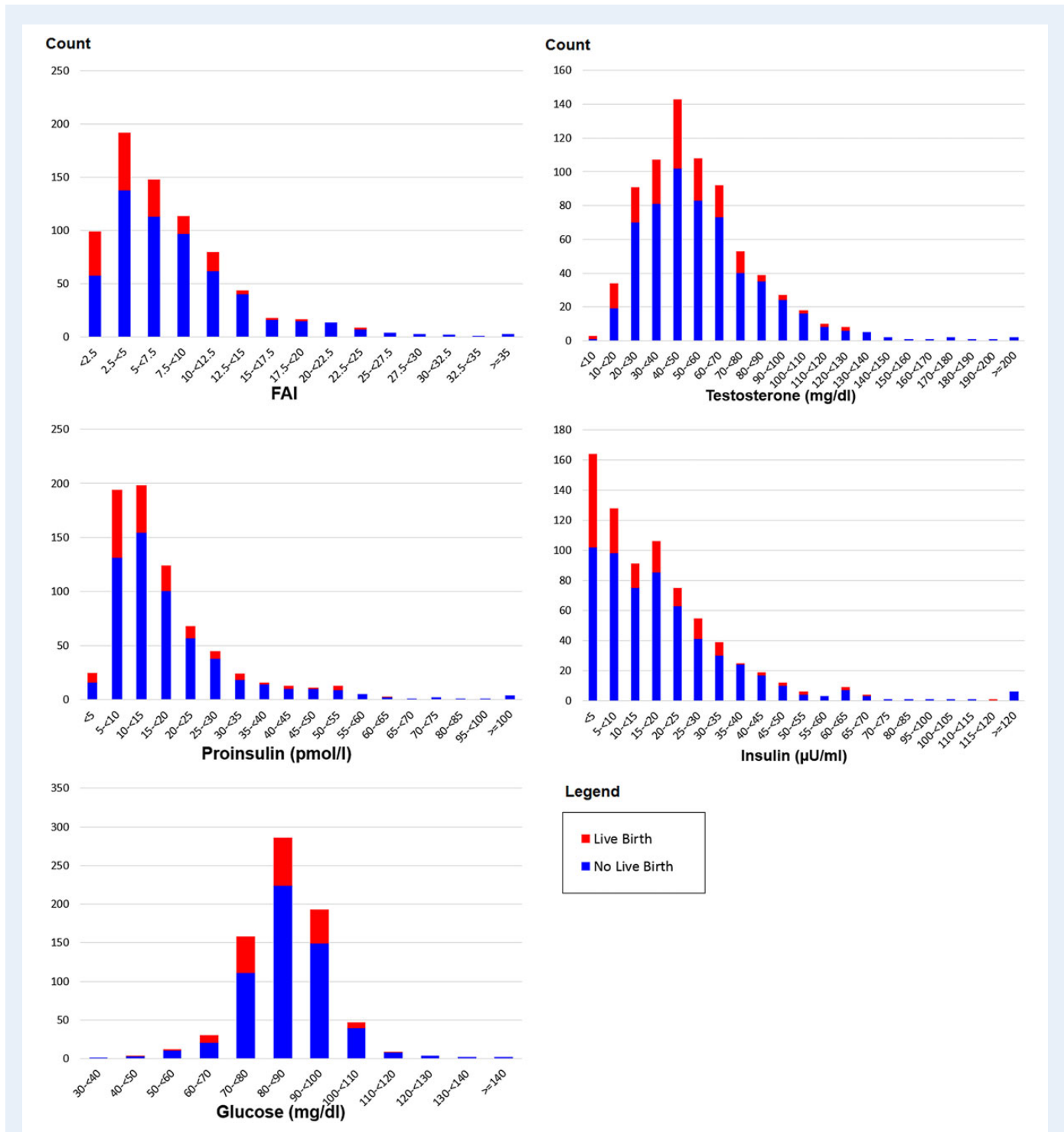
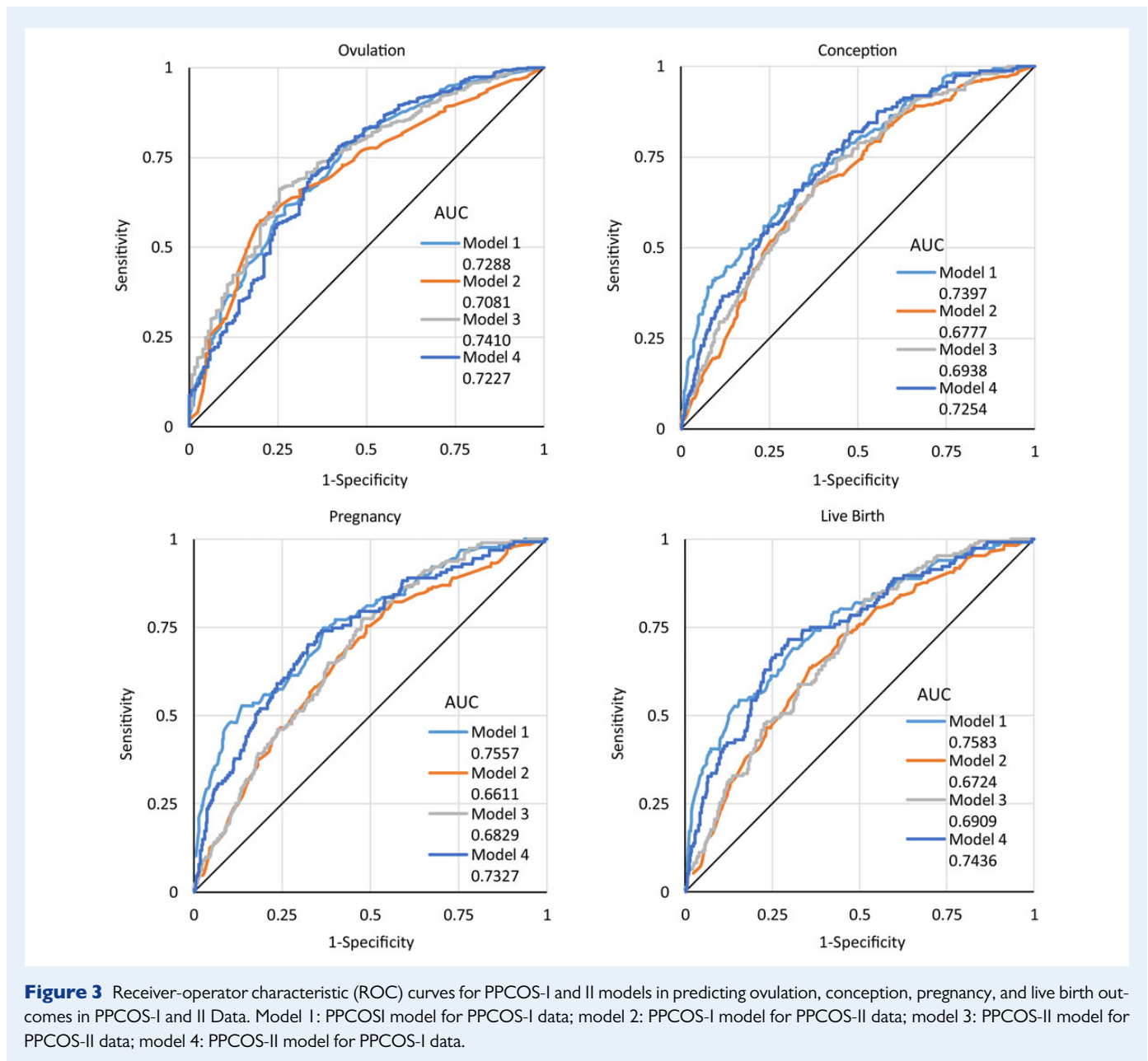


Figure 2 Histograms of biomarkers in PPCOS-II. Red and blue indicate live birth and no live birth, respectively.

(0.73). When the PPCOS-II models were used to predict the outcomes in the PPCOS-II data (model 3), the AUCs were the third largest for conception (0.69), pregnancy (0.68), and live birth (0.69), and the largest for ovulation (0.74). When the PPCOS-I models were used to predict the outcomes in the PPCOS-II data (model 2), the AUCs were the lowest for all outcomes (ovulation: 0.71, conception: 0.68, pregnancy: 0.66 and live birth: 0.67). When the PPCOS-II models were used to predict

the outcomes in the PPCOS-I data (model 4), the AUCs were very similar to those when the PPCOS-I models were used to predict the outcomes in the PPCOS-I data, suggesting that the PPCOS-II models were better than the PPCOS I models when they were applied to both the PPCOS-I and II data.

Considering the independence of the PPCOS-I and II data, we tested the differences of AUCs when the same model was applied to both the



PPCOS-I and II data. The AUC from the use of the PPCOS-I model for the PPCOS-I data is significantly larger than that from the use of the PPCOS-II model for the PPCOS-II data for predicting pregnancy (difference 0.073, $P = 0.025$, 95% CI: 0.009–0.137) and live birth (difference 0.067, $P = 0.045$, 95% CI: 0.002–0.133). Overall, the smaller AUCs obtained from the PPCOS-II data suggest that predicting outcomes in the PPCOS-II participants was more difficult than that in the PPCOS I participants.

Calibration analysis

Supplementary Figure S1 displays the plots for the PPCOS I data. Each panel is one of the four outcomes. Two lines (solid for the PPCOS I model and dashed for the PPCOS-II model) are the least squares fits to the five points (based on the division into quintiles described in

Materials and Methods) where the x-axis is the average of the estimated success probabilities and the y-axis the observed success rates based on the counts. Supplementary Figure S2 is analogous to Supplementary Figure S1, but displays the result for the PPCOS-II data. All regression lines have the intercepts very close to 0 (e.g. off by the second decimal point, P -value >0.05), and the slopes very close to 1 (e.g. off by the second decimal point, P -value <0.05). These plots suggest again the PPCOS I and II models are well calibrated between these two independent trials. Furthermore, the P -values obtained from the Hosmer Lemeshow test for goodness-of-fit for all logistic models were ~ 0.08 or above, suggesting reasonable fits of the data. The P -values obtained from the Pearson's chi-square test for the independent validation of all logistic models between PPCOS I and II data were greater than 0.2, confirming a reasonable match between the prediction models and the independently observed data.

Discussion

In this study, we examined a number of clinical and biochemical predictors for fertility outcomes in women with PCOS who participated in two large, randomized clinical trials. The participants recruited for both of these trials came from similar geographical areas and met similar criteria. Simple demographic, physical measures or biochemical features that are readily available to the clinician appear to have modest power to predict pregnancy outcomes in women with PCOS based on the cumulative sample of 1376 women. A younger age, lower BMI, shorter duration of attempting to become pregnant, and overall hormonal levels indicative of less insulin resistance and hyperandrogenism were all found to be predictive of ovulation, conception, clinical pregnancy and/or live birth. Most of the predictors were similar for the PPCOS-I and II data, although in the PPCOS-II trial, an association between history of prior pregnancy loss and conception was observed. Unlike in the PPCOS-I trial, proinsulin levels were not predictive of pregnancy outcome. These minor discrepancies between the two studies may reflect inconsistent relationships between these latter predictive factors. Alternatively, they may represent a modest interaction between the treatments used in the two studies. Although there were patterns of potential interactions between treatment and BMI groups in both the PPCOS-I and II data, they did not reach the statistical significance ($P = 0.05$). The sample sizes of the resulting groups might limit the power of the PPCOS-I and II trials in detecting such interactions.

In both the PPCOS-I and II populations, female obesity can have a significantly negative impact on all outcomes as can be seen in Table III and Fig. 1 when very obese women ($BMI \geq 35$) were compared with non-obese women ($BMI < 30$). However, such an association may not be apparent always. For example, when BMI was treated as a continuous variable, it was only significantly associated with ovulation and conception (Table III). These inconsistencies suggest that the effects of BMI on pregnancy outcomes are complex in both its own association with the pregnancy outcomes and the existence of confounding factors. Nonetheless, the detrimental effects of obesity on menstrual function (Lake *et al.*, 1997), preterm birth, stillbirth (Chu *et al.*, 2007) and recurrent miscarriage (Metwally *et al.*, 2010) have all been well documented.

As presented in Table II, using the PPCOS-II data, we have largely confirmed the predictors that were found to be significant in the prior PPCOS-I models. Of note, duration of attempting conception is the most consistent predictor among all considered factors for the four pregnancy-related outcomes. Due to the known association among the predictors that we considered as well as the difference in treatments, the predictors that were selected from the PPCOS-II models were mostly different from those selected in the PPCOS-I models. This difference does not necessarily suggest inconsistencies in our data and findings, but rather reflects the complex relationship among the predictors and outcomes. In fact, the PPCOS-I and II models performed similarly in terms of prediction (Fig. 1). We have also observed differences in the biomarkers reported in Table I between the PPCOS-I and II data. We have used the same lab for those tests and confirmed no change in the assays. Though clinically relatively small, those differences can be statistically significant due to the large sample sizes in the two studies.

Smoking did not remain in any PPCOS-I models, but was a significant predictor for ovulation in the PPCOS-II model. This may partially be due to its known association with increased free testosterone and worsening insulin resistance (Cupisti *et al.*, 2010). In a recent study specifically

designed to address the accuracy of self-reported smoking and the effect on smoking behavior from the infertility treatment (Legro *et al.*, 2014c), hirsutism scores at baseline were found to be lower in the never smoker than in the past smokers. Total testosterone levels at baseline were also lower in the never smokers than in the current smokers. At end of study follow-up insulin levels and homeostatic index of insulin resistance increased in the current smokers compared with baseline and with non-smokers. Moreover, the effect of smoking may also have been difficult to detect, as only about 40% of women reported smoking, although we should note that self-report of smoking among women with infertility has been shown to be reliable (Cupisti *et al.*, 2010). Given these complexities with smoking, we believe we should examine current, quit and never smoking. Interestingly enough, although history of smoking characterized by ever smoking is not significant for live birth, current and quit smoking are significant risk ($OR = 0.31$, $CI = 0.16-0.63$) and protective predictors ($OR = 1.76$, $CI = 1.21-2.55$) of live birth, respectively. Adverse effects of smoking on fetal health have been well established, but only about half of the smokers quit smoking during pregnancy in our studies. These data should provide additional motivation to engage infertile, anovulatory women to quit smoking; in particular, given that smoking status is unlikely to change during infertility treatment, extra attention should be paid to smoking cessation in current or recent smokers who seek or who are receiving infertility treatment (Legro *et al.*, 2014c).

We have noted different criteria that were used to define ovulation and conception and differences in the baseline characteristics between the PPCOS-I and II participants. Nonetheless, we have largely confirmed the predictors that were identified in the PPCOS-I trial, and vice versa. The ROC curves in Fig. 3 suggest that the selected predictors had good and stable power to predict the outcomes, although there is clearly room for improvement in prediction of these complex outcomes. We should note that expecting the AUC be 80% or higher is likely to be unrealistic because of both unknown and uncertain etiologies of the pregnancy outcomes. Despite the modest AUCs, our findings have important clinical implications. First, some published studies used relatively small number of samples. For example, the article by Veltman-Verhulst *et al.* (2012) used only 108 women who were from the same medical center. We made use of 1376 couples in > 10 medical centers, and hence these findings are more representative of and generalizable to clinical practice. Second, our findings also suggest that it is very difficult, if not impossible, to have both high and consistent prediction power for pregnancy outcomes because there are likely many factors that have small effects which not only increase the risk but also the uncertainty. Lastly, the ROC curves are helpful to understand the discriminant power of our prediction models, which is modest as presented above. While a high discriminant power is indicative of the quality of our prediction models, a model with a high discriminant power (i.e. the ability to call a subject at high or low risk of a certain status) can be built without depending on accurate probability estimates of the outcomes. We are, however, more interested in the predictive power of our models in terms of both the precision and reproducibility. Our calibration analysis confirms the validity of the predictors that were identified before and here. Similar techniques have been used in predicting the probability of a live birth after *in vitro* fertilization (Leushuis *et al.*, 2009; Smith *et al.*, 2015).

In conclusion, duration of attempting conception and age were the most consistent predictors among all considered factors for the four

pregnancy-related outcomes, although longer duration of attempting conception may be interrelated to increasing age of the women. The mutual verification of the predictors between the PPCOS-I and II trials attests to the value of simple clinical and biochemical criteria for predicting pregnancy outcomes in women with PCOS, but a great deal of variability remains, limiting our prediction precision.

Supplementary data

Supplementary data are available at <http://humrep.oxfordjournals.org/>.

Acknowledgements

In addition to the authors, other members of the NICHD Reproductive Medicine Network were as follows: Penn State College of Medicine, Hershey: C. Bartlebaugh, W. Dodson, S. Estes, C. Gnatuk, R Ladda, J. Ober; University of Texas Health Science Center at San Antonio: C. Easton, A. Hernandez, M. Leija, D. Pierce, R Bryzski; Wayne State University: A. Awonuga, L. Cedo, A. Cline, K. Collins, S.A. Krawetz, E. Puscheck, M. Singh, M. Yoscovits; University of Pennsylvania: K. Barnhart, K. Lecks, L. Martino, R. Marunich, P. Snyder; University of Colorado: W.D. Schlaff, A. Comfort, M. Crow; University of Vermont: A. Hohmann, S. Mallette; University of Michigan: M. Ringbloom, J. Tang; University of Alabama Birmingham: S. Mason; Carolinas Medical Center: N. DiMaria; Virginia Commonwealth University: M. Rhea; Stanford University Medical Center: K. Turner; University of Virginia: D.J. Haisenleder; SUNY Upstate Medical University: J.C. Trussell; Yale University: D. DelBasso, Y. Li, R. Makuch, P. Patrizio, L. Sakai, L. Scahill, H. Taylor, T. Thomas, S. Tsang, M. Zhang; Eunice Kennedy Shriver National Institute of Child Health and Human Development: C. Lamar, L. DePaolo; Advisory Board: D. Guzick (Chair), A. Herring, J. Bruce Redmond, M. Thomas, P. Turek, J. Wactawski-Wende; Data and Safety Monitoring Committee: R. Rebar (Chair), P. Cato, V. Dukic, V. Lewis, P. Schlegel, F. Witter.

Authors' roles

H.K., R.S.L., M.P.D., N.S., E.E. and H.Z. were involved in study concept and design. S.J., H.K. and H.Z. analyzed data. H.K. and H.Z. drafted the manuscript, and H.Z. had a primary responsibility for final content. All authors were involved in acquisition of the data collection, interpreted the data, provided critical input to the manuscript, and approved the final manuscript.

Funding

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Grants U10 HD27049, U10 HD38992, U10HD055925, U10 HD39005, U10 HD33172, U10 HD38998, U10 HD055936, U10 HD055942, and U10 HD055944; and U54-HD29834. Heilongjiang University of Chinese Medicine Grants 051277 and B201005. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD.

Conflict of interest

R.S.L. reports receiving consulting fees from Euroscreen, Astra Zeneca, Clarus Therapeutics, and Takeda, and grant support from Ferring, Astra Zeneca, and Toba. K.R.H. reports receiving grant support from Roche Diagnostics and Ferring Pharmascience. G.C. reports receiving Honorarium and grant support from Abbvie Pharmaceuticals and Bayer Pharmaceuticals. M.P.D. holds equity from Advanced Reproductive Care Inc and DS Biotech, receives fees from Advanced Reproductive Care Inc, Actamax, Auxogyn, ZSX Medical, Halt Medical, and Neomed, and receives grant support from Boehringer-Ingelheim, Abbott, and Bio-Sante, Ferring Pharmaceuticals, and EMD Serono. H.Z. receives research support from the Chinese 1000-scholar plan. Others report no disclosures other than NIH grant support.

References

- Asunción M, Calvo RM, San Millán JL, Sancho J, Avila S, Escobar-Morreale HF. A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. *J Clin Endocrinol Metab* 2000; **85**:2434–2438.
- Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, Janssen OE, Legro RS, Norman RJ, Taylor AE et al. Positions statement: criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: an Androgen Excess Society guideline. *J Clin Endocrinol Metab* 2006; **91**:4237–4245.
- Chu SY, Kim SY, Lau J, Schmid CH, Dietz PM, Callaghan WM, Curtis KM. Maternal obesity and risk of stillbirth: a metaanalysis. *Am J Obstet Gynecol* 2007; **197**:223–228.
- Cupisti S, Häberle L, Dittrich R, Oppelt PG, Reissmann C, Kronawitter D, Beckmann MW, Mueller A. Smoking is associated with increased free testosterone and fasting insulin levels in women with polycystic ovary syndrome, resulting in aggravated insulin resistance. *Fertil Steril* 2010; **94**:673–677.
- Diamanti-Kandarakis E, Kouli CR, Bergiele AT, Filandra FA, Tsianateli TC, Spina GG, Zapanti ED, Bartzis MI. A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. *J Clin Endocrinol Metab* 1999; **84**:4006–4011.
- Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. *Commun Stat* 1980; **A10**:1043–1069.
- Knochenhauer ES, Key TJ, Kahsar-Miller M, Waggoner W, Boots LR, Azziz R. Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States: a prospective study. *J Clin Endocrinol Metab* 1998; **83**:3078–3082.
- Lake JK, Power C, Cole TJ. Women's reproductive health: the role of body mass index in early and adult life. *Int J Obes Relat Metab Disord* 1997; **21**:432–438.
- Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA et al. Clomiphene, metformin, or both for infertility in the polycystic ovary syndrome. *N Engl J Med* 2007; **356**:551–566.
- Legro RS, Brzyski RG, Diamond MP, Coutifaris C, Schlaff WD, Alvero R, Casson P, Christman GM, Huang H, Yan Q et al. The Pregnancy in Polycystic Ovary Syndrome II study: baseline characteristics and effects of obesity from a multicenter randomized clinical trial. *Fertil Steril* 2014a; **101**:258–269.e8.
- Legro RS, Brzyski RG, Diamond MP, Coutifaris C, Schlaff WD, Casson P, Christman GM, Huang H, Yan Q, Alvero R et al. Letrozole versus clomiphene for infertility in the polycystic ovary syndrome. *N Engl J Med* 2014b; **371**:119–129.

- Legro RS, Chen G, Kunselman AR, Schlaff WD, Diamond MP, Coutifaris C, Carson SA, Steinkampf MP, Carr BR, McGovern PG *et al.* Smoking in infertile women with polycystic ovary syndrome: baseline validation of self-report and effects on phenotype. *Hum Reprod* 2014c;**29**:2680–2686.
- Leushuis E, van der Steeg JW, Steures P, Bossuyt PM, Eijkemans MJ, van der Veen F, Mol BW, Hompes PG. Prediction models in reproductive medicine: a critical appraisal. *Hum Reprod Update* 2009;**15**:537–552.
- Metwally M, Saravelos SH, Ledger WL, Li TC. Body mass index and risk of miscarriage in women with recurrent miscarriage. *Fertil Steril* 2010;**94**:290–295.
- Rausch ME, Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, McGovern PG, Cataldo NA *et al.* Predictors of pregnancy in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2009;**94**:3458–3466.
- Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril* 2004;**81**:19–25.
- Sirmans SM, Pate KA. Epidemiology, diagnosis, and management of polycystic ovary syndrome. *Clin Epidemiol* 2013;**6**:1–13.
- Smith AD, Tilling K, Lawlor DA, Nelson SM. External validation and calibration of IVFpredict: a national prospective cohort study of 130,960 in vitro fertilisation cycles. *PLoS One* 2015;**10**:e0121357.
- Sodergard R, Backstrom T, Shanbhag V, Carstensen H. Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. *J Steroid Biochem* 1982;**16**:801–810.
- Taylor AE, McCourt B, Martin KA, Anderson EJ, Adams JM, Schoenfeld D, Hall JE. Determinants of abnormal gonadotropin secretion in clinically defined women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 1997;**82**:2248–2256.
- Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Med* 2010;**8**:41.
- Thessaloniki ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Consensus on infertility treatment related to polycystic ovary syndrome. *Fertil Steril* 2008;**89**:505–522.
- Veltman-Verhulst SM, Fauser BCJM, Eijkemans MJ. High singleton live birth rate confirmed after ovulation induction in women with anovulatory polycystic ovary syndrome: validation of a prediction model for clinical practice. *Fertil Steril* 2012;**98**:761–768.