

Decision trees for identifying predictors of treatment effectiveness in clinical trials and its application to ovulation in a study of women with polycystic ovary syndrome

Heping Zhang^{1,*}, Richard S. Legro², Jeffrey Zhang³, Leon Zhang³, Xiang Chen¹, Hao Huang¹, Peter R. Casson⁴, William D. Schlaff⁵, Michael P. Diamond⁶, Stephen A. Krawetz⁶, Christos Coutifaris⁷, Robert G. Brzyski⁸, Gregory M. Christman⁹, Nanette Santoro⁵, and Esther Eisenberg¹⁰, for the Reproductive Medicine Network

¹Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT 06520-8034, USA ²Department of Obstetrics and Gynecology, Pennsylvania State University, Hershey, PA, USA ³Guilford High School, Guilford, CT, USA ⁴Department of Medicine and Obstetrics and Gynecology, University of Vermont, Burlington, VT, USA ⁵Department of Obstetrics and Gynecology, University of Colorado, Denver, USA ⁶Departments of Obstetrics and Gynecology, Wayne State University, Detroit, MI, USA ⁷University of Pennsylvania, Philadelphia, PA, USA ⁸Department of Obstetrics and Gynecology, The University of Texas Health Science Center at San Antonio, San Antonio, TX, USA ⁹Department of Obstetrics and Gynecology, Reproductive Sciences Program, University of Michigan, Ann Arbor, MI, USA ¹⁰Reproductive Sciences Branch, National Institute of Child Health and Human Development, Bethesda, MD, USA

*Correspondence address. Tel: +1-203-785-5185; Fax: +1-203-785-5145; E-mail: heping.zhang@yale.edu

Submitted on March 8, 2010; resubmitted on June 20, 2010; accepted on June 24, 2010

BACKGROUND: Double-blind, randomized clinical trials are the preferred approach to demonstrating the effectiveness of one treatment against another. The comparison is, however, made on the average group effects. While patients and clinicians have always struggled to understand why patients respond differently to the same treatment, and while much hope has been held for the nascent field of predictive biomarkers (e.g. genetic markers), there is still much utility in exploring whether it is possible to estimate treatment efficacy based on demographic and baseline variables.

METHODS: The pregnancy in polycystic ovary syndrome (PPCOS) study was a prospective, multi-center, randomized clinical trial comparing three ovulation induction regimens: clomiphene citrate (CC), metformin and the combination of the two. There were 446 women who ovulated in response to the treatments among the entire 626 participants. In this report, we focus on the 418 women who received CC (alone or combined with metformin) to determine if readily available baseline physical characteristics and/or easily obtainable baseline measures could be used to distinguish treatment effectiveness in stimulating ovulation. We used a recursive partitioning technique and developed a node-splitting rule to build decision tree models that reflected within-node and within-treatment responses.

RESULTS: Overall, the combination of CC plus metformin resulted in an increased incidence of ovulation compared with CC alone. This is particularly so in women with relatively larger left ovarian volumes (≥ 19.5 cubic cm), and a left ovarian volume < 19.5 cubic cm was related to treatment outcomes for all subsequent nodes. Women who were older, who had higher baseline insulin, higher waist-to-hip circumference ratio or higher sex hormone-binding globulin levels had better ovulatory rates with CC alone than with the combination of CC plus metformin.

CONCLUSIONS: Polycystic ovary syndrome (PCOS) is a phenotypically diverse condition. Both baseline laboratory and clinical parameters can predict the ovulatory response in women with PCOS undergoing ovulation induction. Without *a priori* hypotheses with regard to any predictors, the observation regarding left ovary volume is novel and worthy of further investigation and validation.

Key words: PCOS / ovulation induction / decision trees / treatment effectiveness

Introduction

Randomized clinical trials have become the gold standard to compare therapeutic benefits of treatments (Hulley *et al.*, 2007). When interpreting the findings of these trials, the conclusions are based on the overall effects in treatment groups. Given the effort and resources required to conduct these studies, it is important to ask whether anything more can be learned from these trials. For example, can baseline information be extracted that will predict the effectiveness of a treatment or characterize the heterogeneity of treatment effects in subgroups of patients, and ultimately in individual patients? To answer these questions, exploratory analyses are usually needed. Some refer to this type of analysis as 'subgroup analyses' and caution against its use due to the potential to overstate findings and produce misleading results (e.g. Wang *et al.*, 2007). Irrespective, exploratory analysis provides a useful method for hypothesis generation.

In this article, we describe a technique of recursive partitioning to identify groups of patients within which the treatment effects are most different, hence making it easier for clinicians to choose more effective treatments based on patient characteristics. We used data from the pregnancy in polycystic ovary syndrome (PPCOS) study to develop and evaluate our proposed method (Legro *et al.*, 2006).

Polycystic ovary syndrome (PCOS) is the most common cause of female infertility (Hull, 1987), affecting ~7% of women of reproductive age (Azziz *et al.*, 2004; Legro *et al.*, 2007). It has been reported that women with PCOS suffer from anovulation (Hull, 1987), increased early pregnancy loss (Homburg *et al.*, 1988), and later pregnancy complications (Vanky *et al.*, 2004; Boomsma *et al.*, 2006,2008).

The etiology of PCOS is not well understood, and the diagnostic criteria and the treatment for the syndrome vary in practice (Ehrmann, 2005). In one of the largest randomized controlled clinical trials of ovulation induction in women with PCOS, investigators in the reproductive medicine network evaluated the effectiveness of three treatment arms—clomiphene citrate (CC) plus placebo, metformin plus placebo and the combination of CC and metformin—in 626 infertile women with PCOS for up to 6 months (Legro *et al.*, 2007). The trial confirmed the superiority of CC to metformin with respect to both ovulation and live birth, and also revealed that the rate of ovulation (but not of live birth) was significantly higher in the women treated with the combination of CC and metformin than in either single-agent treatment group, although there was no significant difference in the rate of live birth (Legro *et al.*, 2007).

It is important to recognize that the treatment effectiveness in this trial was assessed only at the group level. Given the diverse clinical characteristics of infertile women with PCOS, it is important to understand whether there are baseline characteristics that can predict the ovulatory response to different treatments. Identifying these variables and their discriminatory levels would improve the choice of infertility treatment in an evidence-based manner for women with PCOS.

In a follow-up study using data from the PPCOS trial, we analyzed which baseline characteristics were associated with the highest chance of achieving a successful pregnancy and live birth (Rausch *et al.*, 2009). We found that baseline free androgen index (FAI), baseline proinsulin level, interaction of the treatment arm with the body mass index (BMI) and the duration of attempting conception were significant predictors of ovulation, conception, pregnancy and live birth, and determined the

relative risk of these end-points based on clinical features or laboratory findings (Rausch *et al.*, 2009). This information on relative risks is useful in counseling and planning infertility treatment; however, it does not directly answer the question as to which treatment is potentially best for an individual woman with PCOS. For example, there could be patients who responded poorly to all treatment regimens, but in whom one of the treatment regimens can be shown to be superior on a relative scale. To this end, we propose a method to identify patient groups within which treatment effectiveness is most different, and demonstrate how our method would be applied to infertility treatment.

Materials and Methods

Study design

The PPCOS study was a prospective, multi-center, randomized clinical trial which was sponsored by Eunice Kennedy Shriver National Institute of Child Health and Human Development (Legro *et al.*, 2007). From November 2002 to December 2004, 626 infertile women with PCOS were randomized to one of three treatment arms: CC plus placebo, metformin plus placebo or combination CC plus metformin. The institutional review boards at all participating institutions approved the protocol, and all subjects gave written informed consent. In this report, we excluded 208 women who did not receive CC, as both their ovulation and live birth rates were significantly inferior to that of the CC treatment groups (Legro *et al.*, 2007), and for clarity, the presentation of the analytic method is simpler for the comparison of two treatments than three treatments.

All participants were diagnosed with PCOS, which was defined as oligomenorrhea (history of no more than eight spontaneous menses per year) and hyperandrogenemia (elevated testosterone level documented within the previous year in an outpatient setting on the basis of local laboratory results, with a predetermined cutoff level set by the principal investigator at each site). Exclusion criteria included hyperprolactinemia, congenital adrenal hyperplasia, thyroid disease, other causes of amenorrhea such as premature ovarian failure and clinically suspected Cushing's syndrome or androgen-secreting neoplasm. Other causes of infertility were excluded by documentation of a normal uterine cavity and at least one patent fallopian tube, and each woman's current partner had a semen concentration of at least 2×10^7 /ml. All subjects were in good health with no major medical disorders (Legro *et al.*, 2007).

Baseline laboratory testing was performed after an overnight fast, and all specimens were analyzed in a core laboratory using established assays. Subjects were treated for up to six cycles, or 30 weeks. Metformin was administered as 500 mg tablets, and increased to four daily (2000 mg); CC was administered as 50 mg tablets, one to three daily for 5 days starting day 5 of the menstrual cycle, with the dosage incrementally increased dependent on the ovulatory response. Progesterone was measured weekly, and ovulation was diagnosed by any weekly serum progesterone level of 5 ng/ml (15.9 nmol/l) or greater. Study medication(s) were discontinued at the time of a positive pregnancy test. Pregnant subjects were followed until fetal viability was documented on ultrasound evaluation. The primary outcome of the trial was the live birth rate. Secondary outcomes included the rate of pregnancy loss, singleton birth and ovulation. Data were analyzed according to the intention-to-treat principle. Detailed information on the power analysis, study design, main statistical methods, baseline characteristics, study medications and outcomes have been published previously (Myers *et al.*, 2005; Legro *et al.*, 2006, 2007; Cataldo *et al.*, 2008, Rausch *et al.*, 2009). We have also previously examined ovulation by treatment group incorporating genetic polymorphisms

(Legro et al., 2008), although we did not include genetic polymorphism in the present analysis.

Although the primary outcome of the PCOS trial was the live birth rate, we examined the ovulation rate in this report. We believe ovulation itself is an important outcome. When an anovulatory patient with PCOS presents to a clinician, the initial question is what treatment should be prescribed to enable this patient to ovulate. If there are clinical characteristics that are associated with better ovulatory response to a specific treatment, this would help with clinical decision making. This, however, does not diminish the importance of live birth as the ultimate outcome.

Statistical method

To identify groups in which the treatment effectiveness was most divergent, we extended the recursive partitioning technique (Breiman et al., 1984; Zhang and Singer, 1999). A chief motivation for recursive partitioning was to develop an expert system for disease diagnosis and treatment (Breiman et al., 1984; Zhang and Singer, 1999; Chen et al., 2010). Herein, we briefly describe the method. Suppose we have a number of study participants, namely the 418 PCOS patients who received CC. We chose to eliminate the metformin-only treatment group as both their ovulation and live birth rates were significantly inferior to that of the CC treatment groups (Legro et al., 2007). We also noted a similar fecundity per ovulation and per patient who ovulated in both CC-containing groups, such that ovulation is a validated marker for live birth in these treatment groups in our trial (Legro et al., 2007).

The data include a clinical outcome such as the success or failure of ovulation, treatment regimens and predictors including age, left and right ovarian volumes, ovulation cycle, BMI, hirsutism score, race, waist, waist/hip ratio, ethnic group, duration of infertility, pregnancy history, prior loss of pregnancy, history of smoking, baseline total testosterone, baseline glucose, baseline insulin, baseline proinsulin and baseline sex hormone-binding globulin (SHBG). The instructions and policy for accessing data are available at <http://c2s2.yale.edu/rmn>. For logistic regression, it may not be wise to include many variables in a model due to potentially missing data and collinearity. However, the decision tree has an embedded variable selection procedure, and we can consider any number of putative predictors.

As illustrated by Fig. 1, the goal of the recursive partitioning is to construct a classification tree that defines subgroups in our study cohorts within which we can observe differences in the effectiveness of the treatment regimens. Recall that, although ovulation is the clinical outcome of interest, we do not look for subgroups with high or low success rates of ovulation; instead, we look for subgroups in which the treatment regimens have differential success rates in ovulation. This approach is aimed at empowering clinician treatment decisions in different subgroups of patients. When different treatment regimens, even highly effective ones, are similar in efficacy then there is little clinical motivation to identify or exploit those differences in treatment decision making.

The process begins with the entire sample, denoted by node 1 in Fig. 1. This is usually referred to as the root node. In our data, node 1 consists of 418 women with PCOS. This node is split into two daughter nodes, labeled as 2 and 3. This split is such that nodes 2 and 3 are 'preferable' to node 1 according to a specific criterion. Here, our preference is based on our goal to have the effects of the two treatments be as different as possible to allow for easy distinction between them.

We make use of all available predictors and all possible splits from the predictors. Unlike parametric statistical methods such as logistic regression, recursive partitioning does not impose a known form of the relationship between the response and predictors. Instead, we use each predictor to create a statement that has a 'yes' or 'no' answer. For example, using the age of a woman, we may ask 'is she younger than 25

or not?' If the answer is yes, she is assigned to node 2, and otherwise to node 3. Whether we use 25 as the age cutoff depends on whether it is 'preferable' to 24, 26 or other ages in our data. Furthermore, whether we choose the age variable to split the root depends on whether it creates a more preferable split than other predictors, such as BMI and race.

As displayed in Fig. 1, 'left ovarian volume' is the best splitting variable and the ideal cutoff is 19.5 cm³. It is important to note that node 2 does not necessarily have better treatment response than node 3 or vice versa. Rather, nodes 2 and 3 are determined by finding the cutpoint at which the two treatments demonstrate the largest difference in the rate of ovulation. Precisely, the divergence of the two treatment effects can be mathematically characterized by the Kullback–Leibler divergence (Kullback and Leibler, 1951). In this study, we chose a simpler measure, i.e. the squared difference in the within-node ovulation rates for the two treatments.

After we find the best possible split for node 1, we continue to try to split nodes 2 and 3 further using the same procedure as we did for node 1. By recursively employing this partitioning technique, we eventually end up with a classification tree such as the one depicted in Fig. 1. After an initial tree is built, a pruning step is usually applied (Breiman et al., 1984; Zhang and Singer, 1999) to remove daughter nodes that are deemed unwarranted. This leads to a final tree for decision making. In this report, because we are interested in the relative effectiveness of two treatments, a pair of nodes is regarded unwarranted if they identify the same treatment to be more effective than the other, because the same clinical decision would be made.

To avoid comparisons between subgroups that are too small to be clinically meaningful, we required at least 20 subjects in any node of the tree, and the partitioning process stopped when no further splits were possible. We also revised the pruning step based on our current need to compare the relative effectiveness of two treatments. After the partitioning process is terminated, we merged any pair of terminal nodes if the same treatment is preferred in both of them. The terminal nodes are the nodes that are not further split. We have implemented our method using JAVA programming, and the program is available upon request.

Results

Ovulation is the primary response variable of our data analysis, and it was defined as a serum progesterone level of 5 ng/ml or greater during a cycle. For our analysis, the response was defined as whether a woman ever had an ovulatory cycle; and 331 women, taking at least CC, did ovulate. Baseline characteristics of the 418 randomized patients are listed in Tables I and II (similar to what have been previously reported) (Myers et al., 2005; Legro et al., 2006, 2007; Rausch et al., 2009). The number of subjects was 209 in the CC group and 209 in the combination therapy group of CC and metformin. Tables I and II indicate no statistically significant differences (at the 0.05 significance level) between the treatment arms in the baseline characteristics, except the BMI (P -value = 0.035).

While it is known that the combined treatment was more effective at achieving ovulation in the overall study population, Figs 1 and 2 suggest that CC alone is a competitive or better treatment for many women. For 153 women in four nodes of Fig. 1 (#8, #12, #14, #16), CC appeared to be more effective. Those women tended to be older and/or have higher baseline insulin level (nodes 8 and 14). They also tended to have higher waist-to-hip circumference ratio (node 12) or higher SHBG levels (node 16).

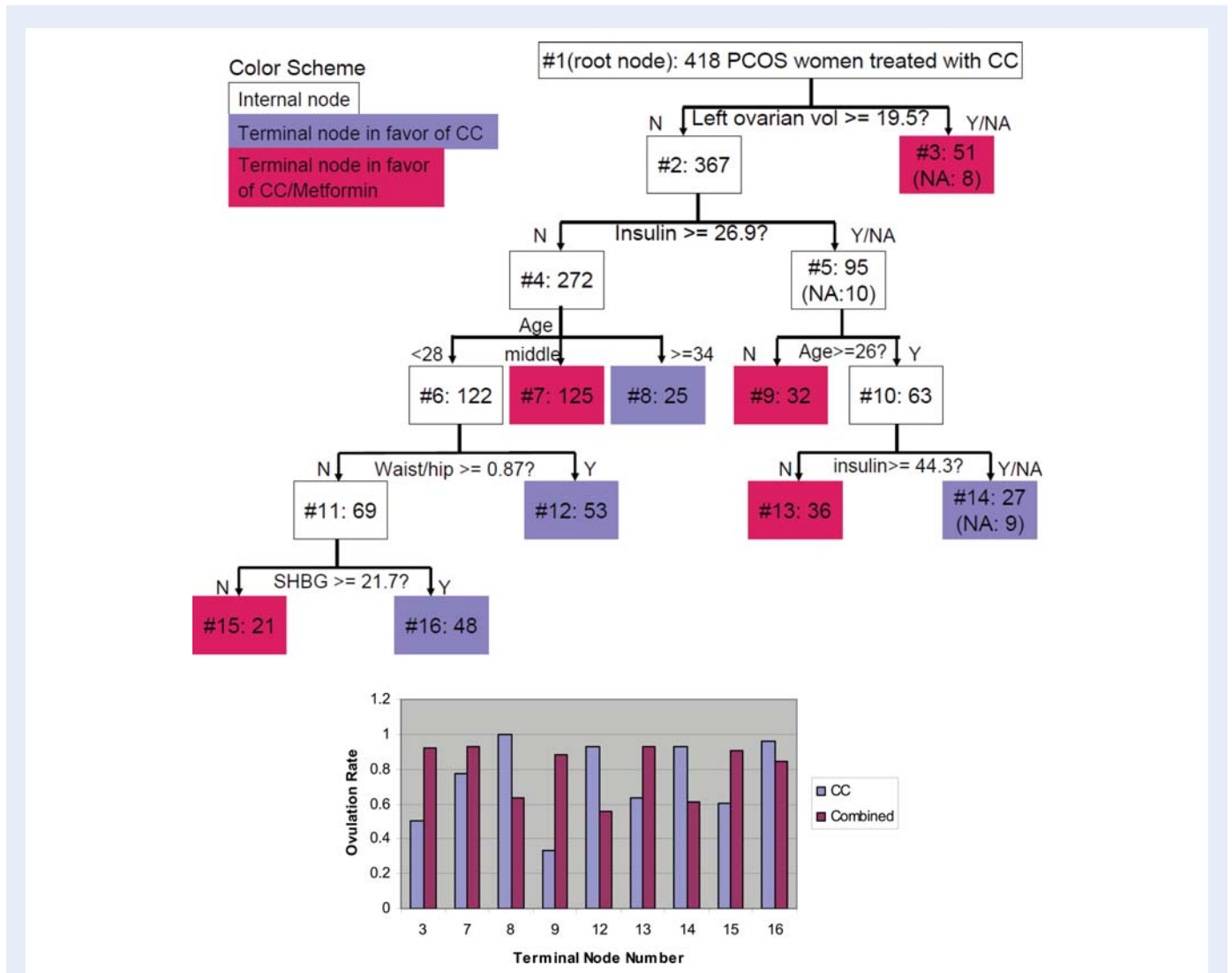


Figure 1 A Classification Tree (top). The top box, labeled #1, is the root node containing the entire study sample of 418 PCOS women. It is partitioned into two nodes 2 and 3 in the next layer. The number of subjects is also displayed next to the node number inside each node. The splitting role is beneath the internal node. The value along an arrow is the cutoff of the splitting variable. When the splitting variable has missing values, the number of affected subjects is presented in the resulting node. In the bottom panel, the vertical bars (purple for CC and red for combination) display the success rates of ovulation within each terminal node, and the terminal node number is printed on the x-axis. Terminal nodes 3, 7, 9, 13 and 15 in the tree are colored red to indicate that the combined treatment is preferred, and other terminal nodes are colored purple indicating that CC is more successful.

Figure 2 is a revision of Fig. 1, specifically after examining the further splits of node 2. The age-based split for node 5 in Fig. 1 is eliminated. We present Fig. 2 emphasizing that it is possible and sometimes reasonable to revise a computer-generated tree in order to simplify the interpretation or apply existing knowledge, although it is possible that this could introduce human confounding. In the present analysis, Fig. 2 is somewhat simpler than Fig. 1, while the overall conclusion is similar.

As an example of interpreting the trees, Fig. 2 supports the notion that CC plus metformin led to a higher incidence of ovulation in women with a left ovarian volume of $>19.5 \text{ cm}^3$. Further, if a woman's left ovarian volume is 18 cm^3 and baseline insulin is

$30 \mu\text{U/ml}$, then her recommended treatment is also CC plus metformin. However, if a woman's left ovarian volume is 18 cm^3 and baseline insulin is $46 \mu\text{U/ml}$, then her recommended treatment is CC alone.

Although this report focuses on ovulation, we are interested in how the preferential treatment for ovulation correlates with the preferential treatment for live birth. The tree in Fig. 5 was grown using the same method as in Fig. 1, except that we replaced ovulation with live birth. As expected, the splitting variables and values are not all the same, but we can also observe that insulin level and hip-to-waist ratio are important for both live birth and ovulation. For insulin, the cutoff values are notably higher in Fig. 1 than those in Fig. 5.

Table I Clinical parameters used as covariates in prediction models.

Parameter	CC	Combined	P-value
Age ¹	27.9 ± 4.0	28.3 ± 4.0	0.280
Cycles to earliest ovulation ¹	2.2 ± 1.7	1.9 ± 1.5	0.057
BMI (kg/m ²) ¹	36.0 ± 8.9	34.2 ± 8.4	0.035
Waist/hip ratio ¹	0.87 ± 0.10	0.86 ± 0.10	0.120
Hirsutism score ¹	14.7 ± 8.2	14.4 ± 7.4	0.690
<8	21.1%	19.6%	0.870
8–16	40.2%	39.2%	
>16	38.8%	41.1%	
Race			0.762
White	70.7%	71.2%	
Black or African American	17.8%	15.4%	
Asian	2.4%	3.4%	
American Indian or Alaska Native	10.1%	11.5%	
Native Hawaiian or Pacific Islander	0.5%	0.0%	
Ethnic group			0.734
Not hispanic or latino	74.6%	76.1%	
Hispanic or latino	25.4%	23.9%	
Length of attempting conception (months) ¹	41.4 ± 39.4	40.7 ± 36.0	0.854
Prior pregnancy history	36.8%	32.1%	0.303
Live birth	22.5%	26.8%	0.307
History of smoking	42.1%	37.8%	0.369
History of alcohol use	66.0%	62.7%	0.475
Prior exposure to metformin only	6.7%	11.5%	0.089
Prior exposure to clomiphene only	32.1%	25.4%	0.130
Prior exposure to combination therapy	18.7%	22.0%	0.395

¹Mean ± SD.**Table II Baseline laboratory parameters used as covariates in prediction models.**

Parameter	CC	Combined	P-value
Total testosterone (ng/dl) ¹	61.3 ± 32.0	63.1 ± 28.4	0.557
Left ovarian volume (cm ³) ¹	10.9 ± 6.9	11.2 ± 6.2	0.606
Glucose (mg/dl) ¹	89.2 ± 16.5	88.9 ± 18.6	0.882
Insulin (μU/ml) ¹	22.6 ± 20.7	22.4 ± 30.0	0.959
SHBG (nmol/l) ^{1,2}	29.8 ± 18.7	31.8 ± 20.3	0.305

¹Mean ± SD.²A homeostasis model assessment of insulin resistance (HOMA-IR) was calculated according to the following formula: (insulin × glucose)/405.

Discussion

The method that we proposed and used to build classification trees to predict treatment-specific successful outcomes is new and data-driven. In this study, we present a proof-of-principle using the data from the PCOS study reported previously by Legro et al. (2007). While the properties of the method require further attention (besides the concerns associated in general with subgroup analyses), the tree-based methods are well established in biomedical research by Zhang and Singer (1999), and the results of our study suggest that this method can produce useful information for clinical consultation and decision making. Furthermore, we are interested in the relative effectiveness of two treatments on ovulation. We cannot use logistic regression to directly investigate this issue, because each patient received only one treatment and the comparison of treatments cannot be done at an individual level, and must be done at the group level. Furthermore, the groups that are of interest to us are not a priori defined. These considerations render our method ideally suited to the clinical problem of selecting the appropriate ovulation induction agent for the appropriate patient. Thus, we propose an important concept to examine the usefulness of the data from a clinical trial and a novel method to achieve the goal.

Some of the factors used in our decision trees are related to those identified in our previous analysis (Rausch et al., 2009), which had the objective to predict the outcome of successful ovulation, pregnancy and live birth in the PPCOS trial. We observed that the baseline FAI, baseline proinsulin level, BMI and duration of infertility were all associated with ovulation and live birth. However, our present objective differed from our previous studies (Legro et al., 2008; Rausch et al., 2009), in that we wanted to evaluate the likelihood of successful ovulation specific to the treatment based on the characteristics of an individual woman. Left ovarian volume, insulin, age, waist-hip ratio and SHBG all served as significant partitions predicting a differential response to CC alone versus CC plus metformin in our model.

Other clinical trials have identified similar predictive factors for ovulation in agreement with our data. In a study of 182 women with normogonadotropic oligomenorrheic infertility, a category that overlaps with our diagnostic criteria of PCOS, Imani et al. (2000) developed a prediction model for ovulation success with CC that included the following: (i) FAI (FAI = testosterone/SHBG ratio and inversely proportional to success); (ii) cycle history (oligomenorrhea or amenorrhea, with oligomenorrhea more favorable); (iii) leptin level (which was superior to BMI in the model and inversely associated with success) and (iv) mean ovarian volume (positively associated with success). In a larger cohort (Imani et al., 1999), the same authors identified a nomogram to determine the chance of ovulation with CC therapy based on the age of the patient, FAI, BMI and whether the patient was oligomenorrheic or amenorrheic, with effects in the same direction.

Palomba et al. (2009) found that the baseline BMI, Ferriman–Gallwey score, serum testosterone and androstenedione, FAI and fasting insulin level were statistically significantly lower in those women who ovulated in response to CC. In addition, SHBG and the fasting glucose to insulin ratio were statistically significantly higher in women who ovulated, compared with those who did not ovulate, in response to CC. In our sample, while higher SHBG was associated with slight superiority of CC alone, a high baseline insulin

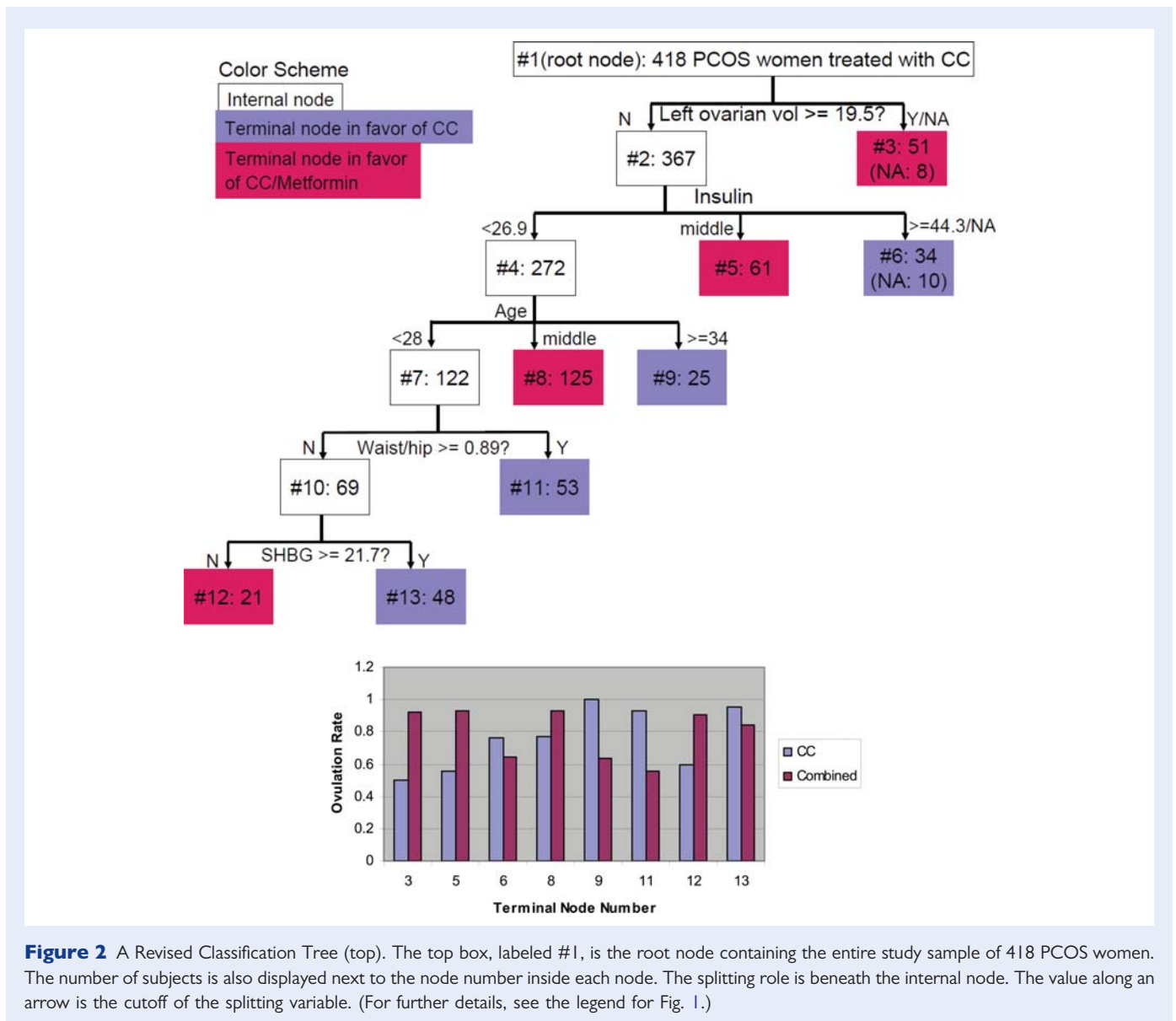


Figure 2 A Revised Classification Tree (top). The top box, labeled #1, is the root node containing the entire study sample of 418 PCOS women. The number of subjects is also displayed next to the node number inside each node. The splitting role is beneath the internal node. The value along an arrow is the cutoff of the splitting variable. (For further details, see the legend for Fig. 1.)

level was somewhat surprisingly also associated with a superiority of CC over combined CC plus metformin. In contrast to Palomba *et al.*, higher waist-hip ratio, which is expected to covary with the BMI, was associated with greater effectiveness of CC in our sample. However, in the Palomba study, CC was compared with placebo, in contrast to comparing CC with CC plus metformin. Also, the difference in the BMI between the patients in the Palomba *et al.* study and this study may contribute to the different results.

Nonetheless, one would have expected that women with the strongest evidence of metabolic dysregulation (lower SHBG, higher fasting insulin, larger waist-hip ratio) would have had a higher rate of success with combined CC plus metformin treatment in our study; yet this was not uniformly the case.

Our findings may also reflect that the addition of metformin to the ovulation induction regimen is not sufficient to adequately reverse the metabolic derangement when it is severe, as is the case with higher fasting insulin levels, lower SHBG and larger waist-hip ratio.

Our analysis suggests that simple parameters that can be easily obtained in the routine evaluation of women with PCOS such as left ovarian volume and baseline insulin level appear to be of great value in helping to choose the more effective treatment option. Personal characteristics that are routinely obtained such as height and weight are also useful for the same purpose. Although BMI is not explicitly used in the trees, the waist-to-hip circumference ratio was found to play an important role in choosing between CC alone and CC plus metformin. It has been reported in the literature that obese PCOS patients, especially those with centripetal obesity have more severe clinical features including worse metabolic parameters, hyperandrogenemia and menstrual abnormalities than do normal weight PCOS women and that obesity also negatively impacts stimulation in ovulation induction cycles, necessitating higher doses with longer periods of stimulation (Dickey *et al.*, 1997; Balen *et al.*, 2006; Pasquali *et al.*, 2006). It is somewhat counterintuitive that women with the highest waist-hip ratio, who would

also be expected to have the highest BMI, had a greater likelihood of ovulation when treated with CC alone, as opposed to CC plus metformin. On the one hand, it may be that the fat patterning in women in the PPCOS trial was such that waist-hip ratio was not as unfavorably impacted as suspected and thus the BMI and the waist-hip ratio did not covary as expected. On the other hand, it may be necessary to examine the clinical features of metabolic dysregulation more carefully in women with PCOS to discern the usefulness of metformin therapy.

The primary cause of PCOS is unknown. However, it is not surprising that an ovarian parameter occupies the root node (node 1), though few would have *a priori* chosen left ovarian volume as that parameter. Combined CC plus metformin is clearly preferable in women with relatively larger left ovarian volumes. The recommended treatment in women with relatively smaller left ovarian volumes (<19.5 cm³) depends on other factors. Imani et al. (2000) reported that an increasing mean ovarian volume was associated with ovulatory success, but they did not report whether one ovary had superior

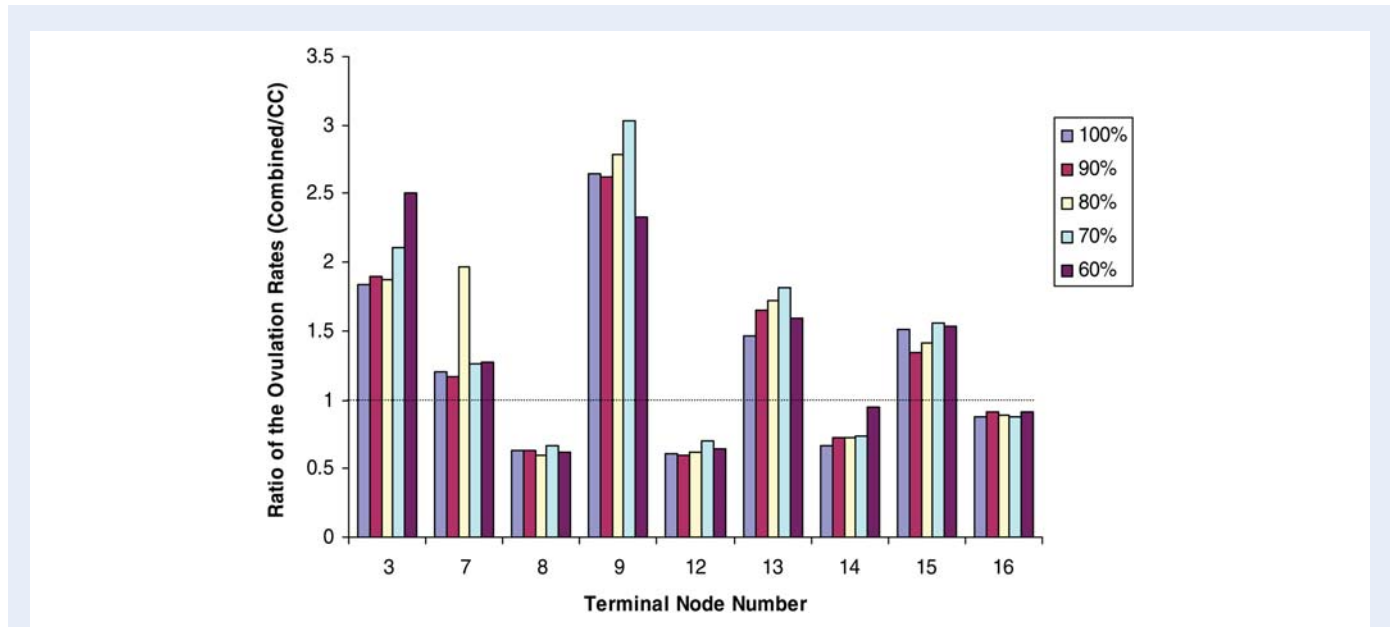


Figure 3 Internal Validation of the Ratio of the Ovulation Rates between the combined and CC treatment groups. The ratios are presented by the terminal nodes in Fig. 1. We took a random 60, 70, 80, 90 and 100% of the samples to examine the trend in the within-node preferential treatment.

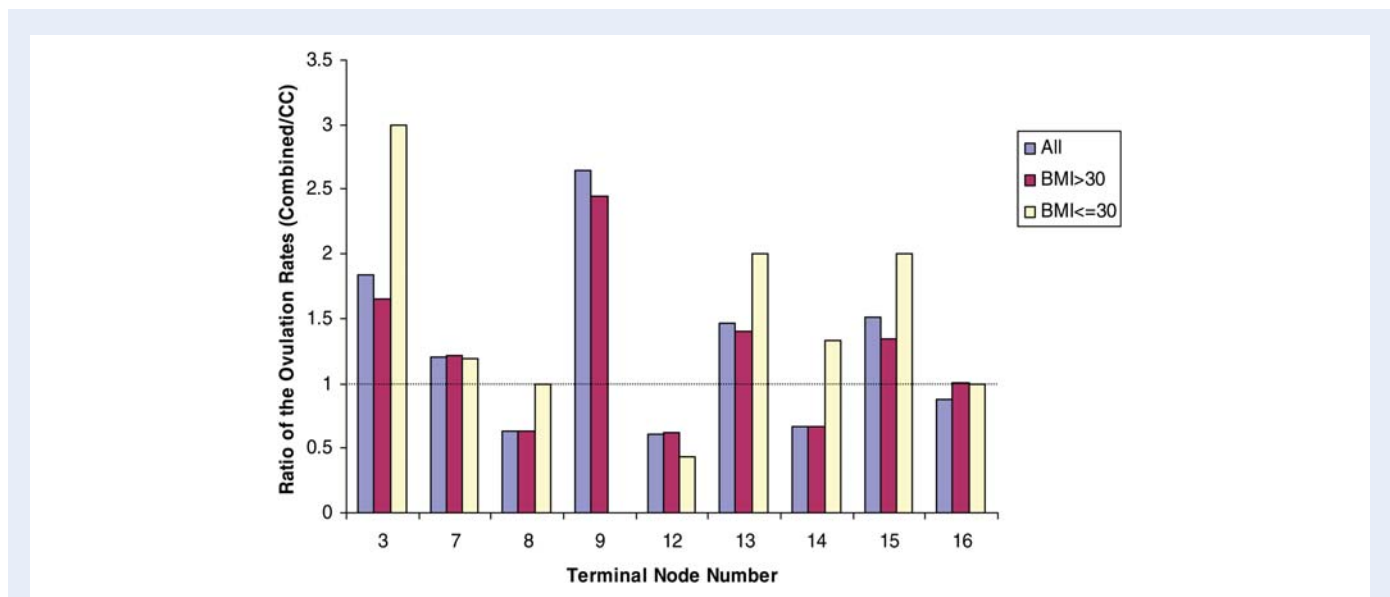


Figure 4 Internal Validation of the Ratio of the Ovulation Rates between the combined and CC treatment groups with respect to the BMI groups. The ratios are presented by the terminal nodes in Fig. 1. No data were available to estimate the ratio for the lower BMI group in terminal node 9.

predictive power over the contralateral ovary. We previously reported that the left ovary in our study cohort was significantly smaller than the right ovary (by ~10%) (Legro *et al.*, 2006). Atiomo *et al.* (2000) have also shown that the left ovary on ultrasound shows greater sensitivity for the diagnosis of PCOS than the right, though it was based on the follicle count and distribution more than the volume.

We caution that our analysis, and the left ovarian volume finding in particular, is best used for hypothesis generation. As a preliminary step for internal validation, in Figs 3 and 4, we examined the trends in the ratio of the ovulation rates between combined and CC treatment groups in the terminal nodes of the tree in Fig. 1. In Fig. 1, the rates of the ovulation in the terminal nodes (the bottom panel) were calculated from all available samples as presented in the terminal nodes (the top part). We examined whether the trend for the preferential treatment would be upheld if we use only a random percentage of the available samples. Specifically, we used 60, 70, 80, 90 and 100% of the samples to compute the ratio of the ovulation rates between

the two treatment groups. Clearly, the trend (i.e. the preferential outcome of one treatment over the other) is consistent with regard to the use of a random subset of the data.

Considering the potentially important role of the BMI, we further validated our result with respect to the ranges of the BMI by examining the trends in the ratio of the ovulation rates between combined and CC treatment groups in the terminal nodes of the tree in Fig. 1. As discussed above, it was reported that BMI is associated with ovulation in response to CC, but for the purpose of this work, namely comparing the ratio of the ovulation rates in the two treatments, Fig. 4 suggests that the BMI does not alter the relative effectiveness of CC versus CC plus metformin. The direction of preference is not changed in eight of the nine terminal nodes when we examined the lower ($BMI \leq 30$) or higher ($BMI > 30$) group alone or together.

There is some evidence in the literature for a role of left ovarian size in identifying treatment effectiveness. In an early animal study, Chávez *et al.* (1987) reported differences in the ovulation rate of the right or left ovary in unilaterally ovariectomized rats. Fukuda *et al.* (2000)

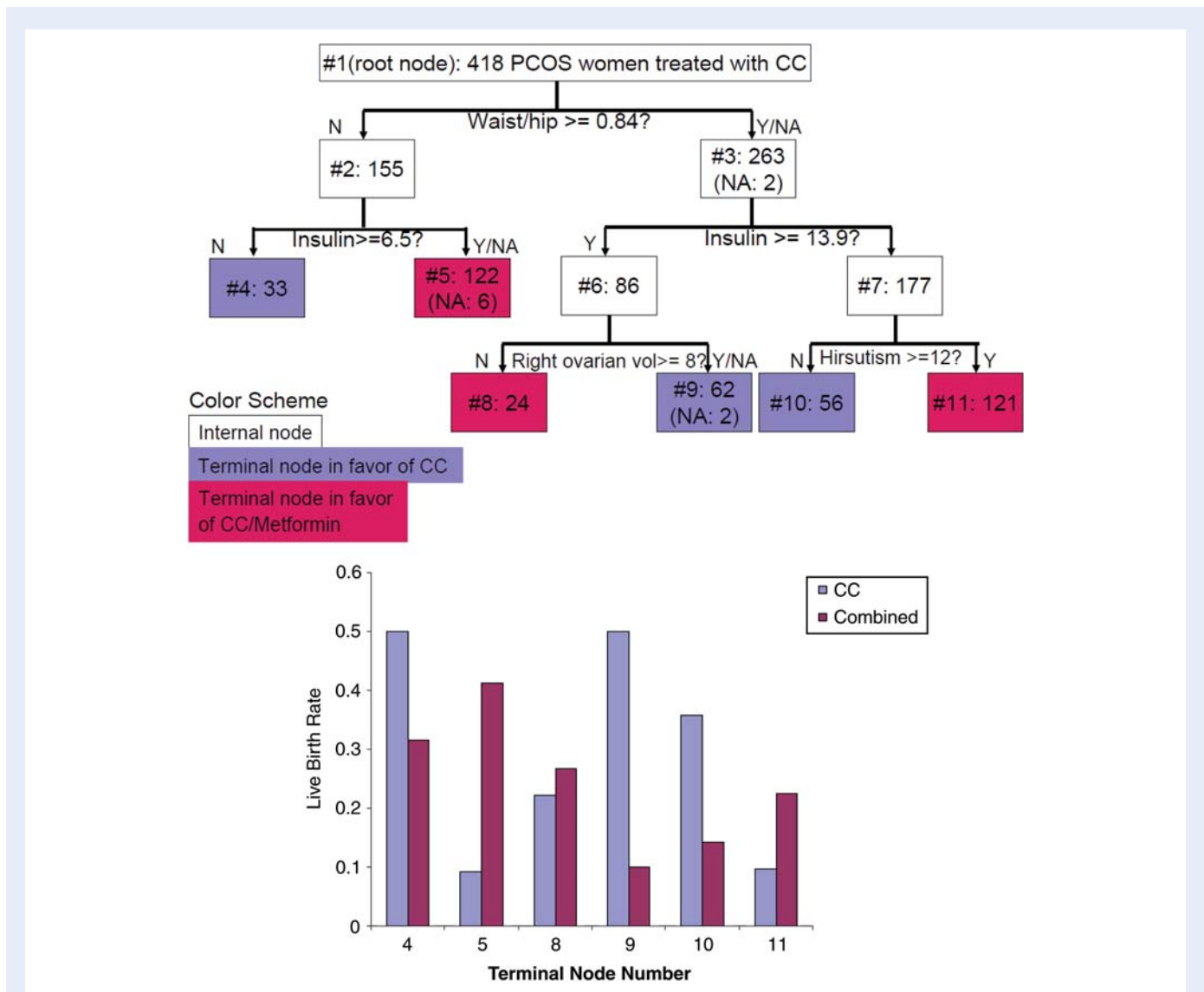


Figure 5 A Decision Tree for Relative Effectiveness of Treatments for Live Birth. (For further details, see the legend for Fig. 1.)

evaluated whether frequency of ovulation and fertility potential of oocytes from the two ovaries differed in regularly menstruating women and concluded that in both fertile and infertile women, the fertility potential of oocytes from the right ovary surpasses that of the left ovary. Furthermore, Järvelä *et al.* (2000) analyzed a database retrospectively that covered 477 cycles during which frozen/thawed embryo transfer had been carried out. They reported that the side of ovulation has a clinical impact and hypothesized that the side of ovulation is significant in terms of embryo implantation. Our findings and the existing reports together support the notion of a role for the side of ovulation. This hypothesis can be tested in prospective trials of ovulation induction and infertility in women with PCOS.

Although the primary interest of this report is on ovulation, we also constructed a tree for live birth. Although the baseline insulin level and the hip-to-waist ratio are common splitting values in Figs 1 and 5, the tree structures are different. One explanation for the discrepancy is that the study was done over a limited period of time and most subjects were treated for up to six cycles or 30 weeks. Thus, if a woman did not ovulate until she was treated with a dose of clomiphene of 150 mg × 5 d, then she would have had a total of three cycles of attempts/opportunity for conception. Given that higher doses of clomiphene could have caused the development of other clinical factors that interfered with conception and live birth (i.e. thickened cervical mucus), it is reasonable to study ovulation rates in this decision tree model. In addition, this model evaluates characteristics of the woman herself and correlates with ovulation, whereas conception and live birth are affected by factors external to the woman herself, e.g. husband's sperm count, timing of intercourse near ovulation, development of pregnancy complications, etc.

In summary, we built classification models to predict treatment-specific occurrence of ovulation in women with PCOS undergoing ovulation induction, and constructed a clinically intuitive and useful chart to contrast ovulation induction by two treatments: CC alone or CC plus metformin. In the current model, women with PCOS can be counseled on their likelihood for ovulation, and ultimately live birth, according to treatment selection by the use of the following baseline clinical parameters: left ovarian volume, insulin, age, baseline waist-to-hip ratio and SHBG. However, due to the data-driven and exploratory nature of the method and analysis, replications of our finding are necessary.

Authors' roles

In addition to the authors, other investigators of the National Cooperative Reproductive Medicine Network were as follows: University of Pennsylvania: K. Barnhart; Pennsylvania State University: J. Ober, S. Eyer, WC Dodson; Yale University: P. Patrizio, L. Sakai, H. Taylor, T. Thomas, S. Tsang and M. Zhang; Wayne State University: E. Puscheck, K. Collins and M. Yoscovits.

Acknowledgements

We would like to acknowledge the exceptional efforts of Sui Tsang for her data management and programming.

Conflict of interest: Dr Zhang reports receiving consulting fees from Eisai, Inc. and Dr Coutifaris reports receiving consulting fees from

Nora Therapeutics, unrelated to the present study. Dr Diamond reports receiving consulting fees from EMD Serono, a manufacturer of Clomiphene Citrate. Other authors report no disclosures.

Funding

This work was supported in part by NIH/NICHD grants U10HD055925 (H.Z.), U10 HD038992 (R.L.), U10 HD038998 (W.S.), U10 HD027049 (C.C.), U10 HD039005 (M.D.), U10 HD055936 (G.C.), U10 HD055942 (R.B.) and H10 HD055944 (P.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or NIH.

References

- Atiomo WU, Pearson S, Shaw S, Prentice A, Dubbins P. Ultrasound criteria in the diagnosis of polycystic ovary syndrome (PCOS). *Ultrasound Med Biol* 2000;**26**:977–980.
- 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**:2745–2749.
- Balen AH, Platteau P, Andersen AN, Devroey P, Sorensen P, Helmggaard L, Arce JC. The influence of body weight on response to ovulation induction with gonadotrophins in 335 women with World Health Organization group II anovulatory infertility. *BJOG* 2006;**113**:1195–1202.
- Boomsma CM, Eijkemans MJ, Hughes EG, Visser GH, Fauser BC, Macklon NS. A meta-analysis of pregnancy outcomes in women with polycystic ovary syndrome. *Hum Reprod Update* 2006;**12**:673–683.
- Boomsma CM, Fauser BC, Macklon NS. Pregnancy complications in women with polycystic ovary syndrome. *Semin Reprod Med* 2008;**26**:72–84.
- Breiman L, Friedman J, Stone C, Olshen R. *Classification and Regression Trees*. New York: Chapman and Hall, 1984.
- Cataldo NA, Barnhart HX, Legro RS, Myers ER, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C *et al.* Extended-release metformin does not reduce the clomiphene citrate dose required to induce ovulation in polycystic ovary syndrome. *J Clin Endocrinol Metab* 2008;**93**:3124–3127.
- Chávez R, Cruz ME, Domínguez R. Differences in the ovulation rate of the right or left ovary in unilaterally ovariectomized rats: effect of ipsi- and contralateral vagus nerves on the remaining ovary. *J Endocrinol* 1987;**113**:397–401.
- Chen X, Wang M, Zhang HP. The use of classification trees for bioinformatics. *WIREs Data Min Knowl Discov* 2010 (in press).
- Dickey RP, Taylor SN, Curolle DN, Rye PH, Lu PY, Pyrzak R. Relationship of clomiphene dose and patient weight to successful treatment. *Hum Reprod* 1997;**12**:449–453.
- Ehrmann DA. Polycystic ovary syndrome. *N Engl J Med* 2005;**352**:1223–1236.
- Fukuda M, Fukuda K, Andersen CY, Byskov AG. Right-sided ovulation favours pregnancy more than left-sided ovulation. *Hum Reprod* 2000;**15**:1921–1926.
- Homburg R, Armar NA, Eshel A. Influence of serum luteinising hormone concentrations on ovulation, conception, and early pregnancy loss in polycystic ovary syndrome. *Br Med J* 1988;**297**:1024–1026.
- Hull MG. Epidemiology of infertility and polycystic ovarian disease: endocrinological and demographic studies. *Gynecol Endocrinol* 1987;**1**:235–245.

- Hulley SB, Cummings Steven R, Browner WS, Grady DG, Newman TB. *Designing Clinical Research: An Epidemiologic Approach*, 3rd edn. Philadelphia: Lippincott Williams and Wilkins, 2007.
- Imani B, Eijkemans MJ, te Velde ER, Habbema JD, Fauser BC. Predictors of chances to conceive in ovulatory patients during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrhoeic infertility. *J Clin Endocrinol Metab* 1999;**84**:1617–1622.
- Imani B, Eijkemans MJ, de Jong FH, Payne NN, Bouchard P, Giudice LC, Fauser BC. Free androgen index and leptin are the most prominent endocrine predictors of ovarian response during clomiphene citrate induction of ovulation in normogonadotropic oligoamenorrhoeic infertility. *J Clin Endocrinol Metab* 2000;**85**:676–682.
- Järvelä I, Nuojua-Huttunen S, Martikainen H. Ovulation side and cycle fecundity: a retrospective analysis of frozen/thawed embryo transfer cycles. *Hum Reprod* 2000;**15**:1247–1249.
- Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat* 1951;**22**:79–86.
- Legro RS, Myers ER, Barnhart HX, Carson SA, Diamond MP, Carr BR, Schlaff WD, Coutifaris C, McGovern PG, Cataldo NA et al. The Pregnancy in Polycystic Ovary Syndrome study: baseline characteristics of the randomized cohort including racial effects. *Fertil Steril* 2006;**86**:914–933.
- 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, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, Coutifaris C, McGovern PG, Cataldo NA et al. Ovulatory response to treatment of polycystic ovary syndrome is associated with a polymorphism in the STK11 gene. *J Clin Endocrinol Metab* 2008;**93**:792–800.
- Myers ER, Silva SG, Hafley G, Kunselman AR, Nestler JE, Legro RS. Estimating live birth rates after ovulation induction in polycystic ovary syndrome: sample size calculations for the pregnancy in polycystic ovary syndrome trial. *Contemp Clin Trials* 2005;**26**:271–280.
- Palomba S, Falbo A, Orio F, Rolino A, Zullo F. Efficacy predictors for metformin and clomiphene citrate treatment in anovulatory infertile patients with polycystic ovary syndrome. *Fertil Steril* 2009;**91**:2557–2567.
- Pasquali R, Gambineri A, Pagotto U. The impact of obesity on reproduction in women with polycystic ovary syndrome. *BJOG* 2006;**113**:1148–1159.
- Rausch ME, Legro RS, Barnhart HX, Schlaff WD, Carr BR, Diamond MP, Carson SA, Steinkampf MP, McGovern PG, Cataldo NA et al., for the Reproductive Medicine Network. Predictors of pregnancy in women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2009;**94**:3458–3466.
- Vanky E, Salvesen KA, Heimstad R, Fougner KJ, Romundstad P, Carlsen SM. Metformin reduces pregnancy complications without affecting androgen levels in pregnant polycystic ovary syndrome women: results of a randomized study. *Hum Reprod* 2004;**19**:1734–1740.
- Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—report of subgroup analysis in clinical trials. *N Engl J Med* 2007;**357**:2189–2194.
- Zhang H, Singer B. *Recursive Partitioning in the Health Sciences*. New York: Springer, 1999.