

Project Synopsis
on
**PCOcare: PCOS Detection and Prediction
using Machine Learning Algorithm**

Submitted as a part of course curriculum for

Bachelor of Technology
in
Computer Science



Submitted by

Kirti Jayant (2000290120087)
Aditi Singh (210029012001)
Pooja Kumari (2000290120107)

Under the Supervision of

Mr. Pawan Kumar Pal
Professor Department of Computer Science

KIET Group of Institutions, Ghaziabad
Department of Computer Science
Dr. A.P.J. Abdul Kalam Technical University
2022-2023

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students

Name: Kirti Jayant/ Aditi Singh/ Pooja Kumari

Roll No.: 2000290120087/ 210029012001/ 2000290120107

Date:

CERTIFICATE

This is to certify that Project Report entitled “**PCOcare: PCOS Detection and Prediction using Machine Learning**” which is submitted by **Kirti Jayant/ Aditi Singh/ Pooja Kumari** in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date:

Supervisor Signature

Mr. Pawan Kumar Pal
Professor Department
of Computer Science

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to **Mr. Pawan Kumar Pal (Professor)**, Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his/her constant support and guidance throughout the course of our work. **His** sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his/her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Shrivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:

Name : Kirti Jayant/ Aditi Singh/ Pooja Kumari

Roll No: 2000290120087/ 2100290129001/ 2000290120107

Date:

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years. The hormonal imbalance leads to a delayed or even absent menstrual cycle. Women with PCOS majorly suffer from excessive weight gain, facial hair growth, acne, hair loss, skin darkening and irregular periods leading to infertility in rare cases. The existing methodologies and treatments are insufficient for early-stage detection and prediction. To deal with this problem, we propose a system which can help in early detection and prediction of PCOS treatment from an optimal and minimal set of parameters. To detect whether a woman is suffering from PCOS, 5 different machine learning classifiers like Random Forest, SVM, Logistic Regression, Gaussian Naïve Bayes, K Neighbours have been used. Out of the 41 features from the dataset, top 30 features were identified using CHI SQUARE method and used in the feature vector. We also compared the results of each classifier and it has been observed that the accuracy of the Random Forest Classifier is the highest and the most reliable.

INTRODUCTION

These days, machine learning, a field of study that gives computers to learn without being explicitly programmed, is playing a key role in the healthcare sector. Machine learning can deal with obscenely huge datasets, convert analysed data into clinical insights and help in the diagnosis of various ailments. Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years. PCOS occurs as a result of hormonal imbalances. In this disorder, the ovaries develop small collections of fluids called follicles (cysts) and fail to release eggs, which is why women suffering from PCOS tend to have complications in conceiving [Zhang, 2018]. A lot of women have PCOS, but do not get diagnosed with it at an earlier stage.

WHAT IS PCOS:

Polycystic ovary syndrome (PCOS) is a problem with hormones that happens during the reproductive years. If you have PCOS, you may not have periods very often. Or you may have periods that last many days. You may also have too much of a hormone called androgen in your body. With PCOS, many small sacs of fluid develop along the outer edge of the ovary. These are called cysts. The small fluid-filled cysts contain immature eggs. These are called follicles. The follicles fail to regularly release eggs. The exact cause of PCOS is unknown. Early diagnosis and treatment along with weight loss may lower the risk of long-term complications such as type 2 diabetes and heart disease.

PROBLEM STATEMENT:

Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years. The hormonal imbalance leads to a delayed or even absent menstrual cycle. Women with PCOS majorly suffer from excessive weight gain, facial hair growth, acne, hair loss, skin darkening and irregular periods leading to infertility in rare cases. The existing methodologies and treatments are insufficient for early-stage detection and prediction. To deal with this problem, we propose a system which can help in early detection and prediction of PCOS treatment from an optimal and minimal set of parameters. To detect whether a woman is suffering from PCOS, 5 different machine learning classifiers like Random Forest, SVM, Logistic Regression, Gaussian Naïve Bayes, K Neighbours have been used. Out of the 41 features from the dataset, top 30 features were identified using CHI SQUARE method and used in the feature vector. We also compared the results of each classifier and it has been observed that the accuracy of the Random Forest Classifier is the highest and the most reliable.

OBJECTIVE:

Polycystic ovary syndrome (PCOS) is one of the most common reproductive, endocrine, and metabolic disorders in premenopausal women. Clinically, PCOS is mainly caused by androgen excess and ovarian dysfunction, manifested by anovulatory menstrual cycles, infertility, and hirsutism.

TREATMENT OF PCOS:

The overall aims of treatment are **to induce ovulation for women desiring conception, to reduce androgen levels, to reduce body weight and to reduce long-term health risks of diabetes mellitus and cardiovascular disease.**

SCOPE:

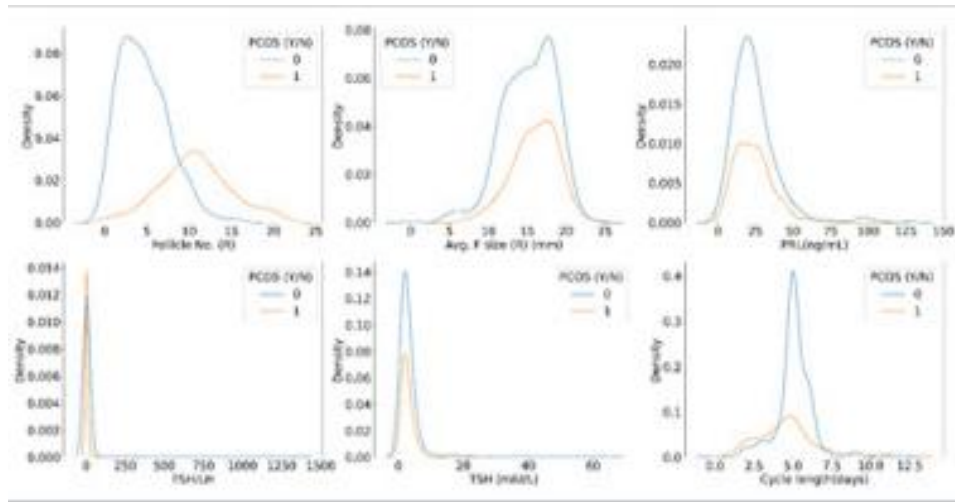
Polycystic ovary syndrome (PCOS) is a heterogeneous disorder characterized by hyperandrogenism and chronic anovulation. Depending on diagnostic criteria, **6% to 20% of reproductive aged women are affected.** Symptoms of PCOS arise during the early pubertal years.

LITERATURE REVIEW

1. PCOS Detection and Prediction using Machine Learning Algorithm:

Emerging technologies are reshaping mankind in a lot of ways. These days, machine learning, a field of study that gives computers to learn without being explicitly programmed, is playing a key role in the healthcare sector. Machine learning can deal with obscenely huge datasets, convert analysed data into clinical insights and help in the diagnosis of various ailments. Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years. PCOS occurs as a result of hormonal imbalances. In this disorder, the ovaries develop small collections of fluids called follicles (cysts) and fail to release eggs, which is why women suffering from PCOS tend to have complications in conceiving [Zhang, 2018]. A lot of women have PCOS, but do not get diagnosed with it at an earlier stage.

The formulation of a good machine learning model is an important aspect of project design. Having the correct patient data is very important because one cannot afford mistakes while devising healthcare services. We have used multiple machine learning models to check which model gives us the most accurate results. To support our claims and results obtained, use of plots and evaluation metrics has been made. A basic workflow diagram to explain the proposed system is given in Figure 1. The following sections will give a detailed insight into the system.



2. PCOS Detect:

(PCOS), also known as polycystic ovarian syndrome, is hormonal endocrine disorder among women of reproductive age. Over five million women worldwide in their reproductive age are suffering from PCOS. The most common symptoms of this disorder may include missed periods, irregular periods, or very light periods, it affects in a way that ovaries become large or may contain many cysts, it can also cause excess body hair, including the chest, stomach, and hirsutism, can cause weight gain, especially around the abdomen, Acne or oily skin. The exact pathophysiology of PCOS is not yet known. This heterogenous disorder is characterized by the ovaries mainly. PCOS is a multifactorial and polygenic condition. Machine Learning is capable of "learning" features from very large amount through clinical practice to diagnose this disorder. This paper put forwards a solution to this problem which helps in early detection and prediction of PCOS treatment from an optimal and minimal set of parameters

Technology is changing every outlook of our lives making remarkable transformations in the healthcare industry, nowadays technology and humans are working hand in hand. For example, robots performing surgeries once seemed a fiction but now they are performing critical and complex surgeries in hospitals. Machine learning is a subclass of artificial intelligence, it helps the system learn, identify patterns of datasets, make logical decisions and performing digital analysis on digital information

including words, numbers, images and clicks. Machine Learning applications mainly include image recognition, data prediction, Medical Diagnosis – Health Care and Clinical Care, etc. In this world of technology many advancements are taking place for detection of PCOS and Machine Learning algorithms are one of them

3. Assessment of Cardiovascular Risk and Prevention of Cardiovascular Disease in Women with the Polycystic Ovary Syndrome:

Polycystic ovary syndrome (PCOS) is a common endocrinopathy affecting 6–10% of reproductive-aged women and manifested by hyperandrogenism, ovulatory dysfunction, and polycystic ovaries in its complete phenotype. Although evidence for cardiovascular events in women who were affected by PCOS during fertile age is limited, available data suggest more frequent cardiovascular disease (CVD) in classic PCOS. In young women with PCOS, multiple risk factors for CVD, including metabolic syndrome (MBS), type 2 diabetes mellitus (T2DM), dyslipidemia, abdominal obesity, and hypertension may be found, and prevention of future cardiovascular adverse effects is needed. With increased adiposity in two thirds of American PCOS women, the degree to which obesity and PCOS interact to promote premature atherosclerosis and increase cardiovascular mortality is a worldwide concern.

Androgen Excess and Polycystic Ovary Syndrome (AE-PCOS) Society appointed a panel to review all published evidence assessing CVD risk in PCOS *vs.* non-PCOS women and to recommend PCOS-related guidelines for CVD prevention. An important consideration was the broader definition of PCOS [*i.e.* Rotterdam 2003 or AE-PCOS 2006 *vs.* the more restrictive NIH 1990 criteria. The former non-NIH criteria include more women and increase heterogeneity of PCOS phenotypic expression.

4. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome:

The disorder that eventually would be known as the polycystic ovary (or ovarian) syndrome (PCOS) was initially described by Stein and Leventhal in 1935 (2). However, the findings of polycystic (or cystic oophoritis or sclerocystic) ovaries dates back at least a century before that (3–5). Despite the difficulty in ascertaining the prevalence of this disorder among women there are convincing data today to suggest that it affects between 6% and 8% of women worldwide, using the National Institutes of Health (NIH) 1990 criteria (6–10), such that it can be considered one of the most common disorders of humans, and the single most common endocrine abnormality of women of reproductive age. There is little disagreement that PCOS should be considered a syndrome, that is, a collection of signs and features, where no single test is diagnostic. In essence, the whole (or global assessment) is greater than the sum of the individual parts (or features). However, establishing a clear and contemporaneous definition for what this syndrome is has important clinical and investigational implications. Clinically, diagnosing a woman as having PCOS implies an increased risk for infertility, dysfunctional bleeding, endometrial carcinoma, obesity, type 2 diabetes mellitus (DM), dyslipidemia, hypertension, and

possibly cardiovascular disease (CVD). Furthermore, it has important familial implications, principally, but not exclusively, for her sisters and daughters. Finally, a diagnosis of PCOS may mandate life-long treatments (e.g., the use of insulin sensitizers), and may negatively affect her ability to access healthcare coverage, principally in capitalistic markets. Consequently, the diagnosis of PCOS should not be assigned lightly, and diagnostic criteria should be based on robust data.

Historic usage in medical practice and/or literature: historic usage may be best reflected in the definitions presented in contemporaneous texts. However, it can be effectively argued that historic usage has limited value in yielding a contemporaneous definition of a disorder or syndrome, except to provide a reference point for the development of an updated definition.

The Pathogenesis of Polycystic Ovary Syndrome (PCOS): The Hypothesis of PCOS

Polycystic ovary syndrome (PCOS) is the most common endocrine disorder in reproductive aged women, with a prevalence between 5% and 15%, depending on the diagnostic criteria applied . PCOS was first described by Stein and Leventhal as a syndrome of oligo-amenorrhea and polycystic ovaries that was variably accompanied by hirsutism, acne, and obesity. Demonstration of polycystic ovaries became required for PCOS diagnosis, which required gynecologic expertise, yet polycystic ovaries were found to be variably associated with the signs and symptoms that characterize the disorder.

We then found that most hyperandrogenic women (two-thirds of those with oligo-amenorrhea, 30% of eumenorrheic ones) had this type of androgenic ovarian dysfunction and that this was independent of serum LH elevation or PCOM in about half of cases. This abnormality was termed functional ovarian hyperandrogenism (FOH), because the steroidogenic disorder is gonadotropin

dependent (ie, any treatment that suppresses gonadotropin production suppresses androgen production), and there is not a requisite anatomic basis for the disorder.

Finally, the hyperandrogenism of PCOS improves during middle age, which is sometimes accompanied by normalization of menstrual regularity. These changes seem related to the fall in follicle number during the premenopausal transition, which is accompanied by falling serum inhibin-B and rising FSH levels that maintain estradiol secretion . Although hyperandrogenism may remit during menopause, lifelong metabolic dysfunction persists and may increase postmenopausal cardiovascular disease risk. Criteria for the diagnosis of postmenopausal PCOS remain to be defined.

Polycystic Ovary Syndrome (PCOS): Arguably the Most Common Endocrinopathy Is Associated with Significant Morbidity in Women:

Women's health is about the prevention, screening, diagnosis, and treatment of disorders that are unique to women. Polycystic ovary syndrome (PCOS) is extremely prevalent and probably constitutes the most frequently encountered endocrinopathy in women of reproductive age. Primary care providers do not commonly appreciate that the syndrome is associated with significant morbidity in terms of both reproductive and nonreproductive events. Having the disorder may significantly impact the quality of life of women during the reproductive years, and it contributes to morbidity and mortality by the time of menopause. A cohort of women with PCOS who were followed for many years after wedge resection revealed several important findings by the time they reached the age of menopause. Their symptoms of PCOS had persisted over this time, they had a later menopause, and they had experienced a higher hysterectomy rate. Most importantly, there was a high prevalence of diabetes (16%) and hypertension (40%).

A uniform definition of PCOS does not exist, in large part because of its diverse and heterogeneous nature. It is clear to us, however, that the disorder is an endocrinopathy, and that it should be referred to as PCOS, a syndrome, rather than a disease. At a meeting held at the National Institutes of Health 10

years ago, there was no consensus but a general agreement that hyperandrogenism and chronic anovulation are the principal facets of the syndrome and that once other disorders (CAH, tumors) were ruled out, the diagnosis of PCOS may be presumed. In the literature, this general definition is quoted as the “NIH Consensus Statement.” Indeed, this was not a consensus conference, and there was no consensus.

6- Endocrine Disruptors and Polycystic Ovary Syndrome (PCOS): Elevated Serum Levels of Bisphenol A in Women with PCOS.

The female gonad appears to be a particularly sensitive target of BPA disruption, this indicated by evidence of interference with ovarian steroidogenesis, folliculogenesis, and ovarian morphology. The underlying mechanisms of BPA that impact upon ovarian function appear to be bidirectional. Specifically, *in vitro* studies have provided evidence that exposure of rat ovarian theca-interstitial cells to BPA results in elevated testosterone synthesis. Androgens interfere with BPA clearance in the liver leading to increased serum levels of BPA. Moreover, BPA alters androgen metabolism in the liver and, acting as a potent sex hormone-binding globulin (SHBG) binder, displaces androgens resulting in increased levels of serum free androgens.

Hyperandrogenaemia, insulin resistance, and chronic anovulation are the cardinal features of polycystic ovary syndrome (PCOS), the most common endocrinopathy of women of reproductive age. Insulin resistance is found in the majority of obese women with PCOS and in a significant proportion (30%) of lean women with the syndrome. Moreover, the prevalence of carbohydrate metabolism disorders, such as impaired glucose tolerance or frank diabetes mellitus, is significantly increased in women with PCOS compared with body mass index (BMI)-matched peers. Although the etiology of the syndrome remains enigmatic, the potential influence of environmental factors on PCOS development has recently been explored. BPA may contribute to the pathogenesis of the syndrome, because elevated BPA levels have been reported in women with ovulatory dysfunction compared with regularly ovulating women. Furthermore, sex is significantly associated with BPA levels given that serum BPA concentrations are significantly higher in men than in women. Recently, it was shown that exposure of neonatal rats to BPA is linked with PCOS-like syndrome and dysregulation of insulin secretion/glucose metabolism.

Early Endocrine, Metabolic, and Sonographic Characteristics of Polycystic Ovary Syndrome (PCOS): Comparison between Nonobese and Obese Adolescents.

POLYCYSTIC OVARY SYNDROME (PCOS) is the most common endocrinopathy in women, present in 5–7% of women of reproductive age. PCOS is characterized by hyperandrogenism and chronic anovulation, and its morbidity may include hyperinsulinemia, insulin resistance, early onset of type 2 diabetes mellitus (DM), dyslipidemia, cardiovascular disease, and infertility. Women with PCOS demonstrate marked clinical heterogeneity; the commonly associated features of hirsutism, acne, polycystic-appearing ovaries, obesity, and acanthosis nigricans are neither uniform nor universal.

The etiology of PCOS remains unclear. The syndrome is often perimenarcheal in onset, and similarities between the physiological changes of puberty and the pathological features of PCOS have been noted, such as the hyperpulsatile gonadotropin secretion, increased ovarian and adrenal steroidogenesis, menstrual irregularity, reduced levels of SHBG and IGFBP-1, hyperinsulinemia, and insulin resistance that develop in both conditions. Decreased levels of SHBG, hyperinsulinemia, insulin resistance, and unfavorable lipid profiles have also been demonstrated in prepubertal girls with premature adrenarche (PA) and pubertal girls with a history of PA, a condition that may herald the later development of anovulation and functional ovarian hyperandrogenism, including PCOS. Nonetheless, PCOS has not been as extensively investigated in the adolescent population. Studies have demonstrated disturbances in insulin sensitivity and insulin secretion early in the course of PCOS and indicate that similar to their adult counterparts, both lean and obese adolescent girls with PCOS are at increased risk for impaired glucose tolerance and DM.

Development of a Health-Related Quality-of-Life Questionnaire (PCOSQ) for Women with Polycystic Ovary Syndrome (PCOS).

POLYCYSTIC ovary syndrome (PCOS) is the most common endocrine disorder among women of reproductive age in the developed world, affecting 5–10% of this population. The disorder exhibits a variety of symptoms including oligomenorrhea, hirsutism, and obesity, not all of which are necessarily present in any one woman. Women with PCOS may complain about irregular menstrual periods and/or heavy menstrual bleeding, infertility, excessive growth of coarse facial and body hair, obesity, oiliness of the skin, seborrhoea, and cystic acne. The impact of these symptoms on a woman's quality of life may be profound and can result in psychological distress that threatens her feminine identity. The condition may therefore result in altered self-perception, a dysfunctional family dynamic, and problems at work.

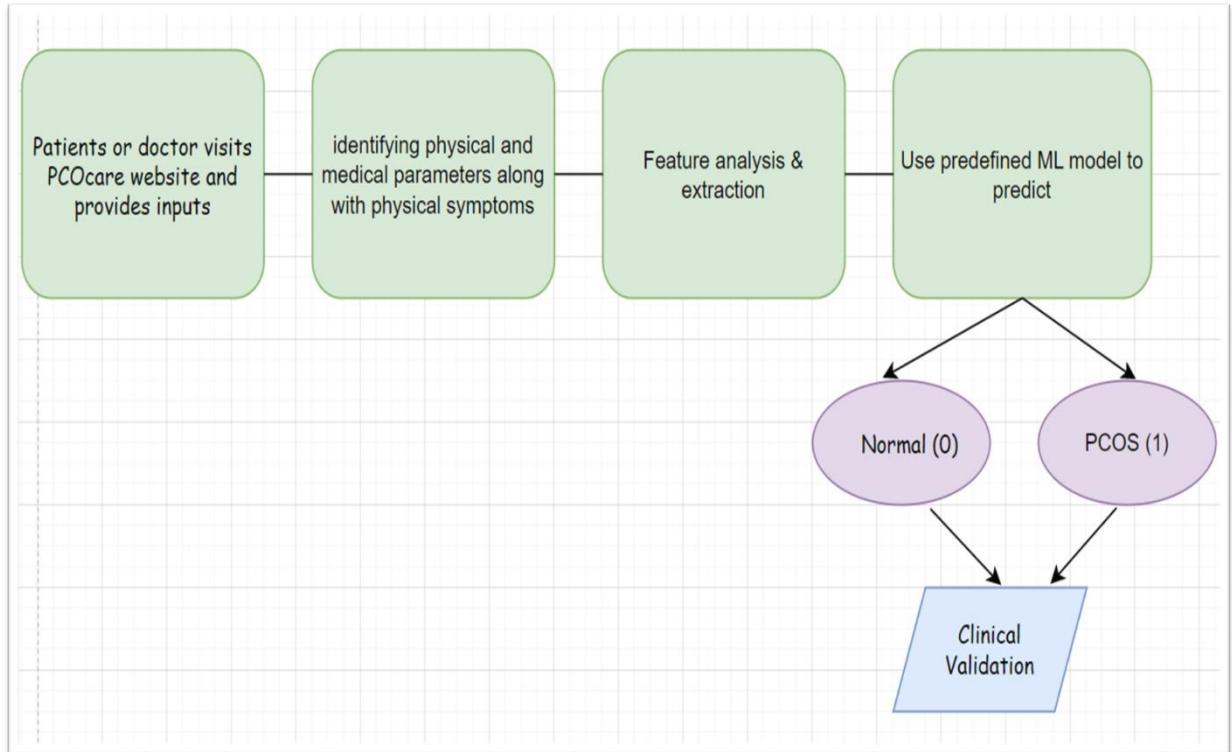
The therapy of PCOS is usually focused on ameliorating its symptoms. Effective treatment can reduce the burden of these symptoms as well as the associated psychological distress and thus improve health-related quality of life (HRQL). Although generic instruments for measuring quality of life are available, they are not designed to measure the range of health-related problems experienced by women with PCOS or to detect the changes in these problems induced by effective interventions. Accordingly, we developed the first health status measure that examines disease-related dysfunction in PCOS women for use in clinical trials and natural history studies.

Diagnosis of polycystic ovary syndrome (PCOS): revisiting the threshold values of follicle count on level for the definition of polycystic ovaries.

Polycystic ovary syndrome (PCOS) is a common endocrine disorder, affecting up to 10% of women of reproductive age. Its prevalence varies according to the definition used and to the reference population.

The cardinal features of PCOS are hyperandrogenism (HA) and oligo-anovulation. The metabolic abnormalities often associated with this syndrome (obesity, insulin resistance, hyperinsulinemia and dyslipidemia) are not included in the definition of the syndrome because it is still unclear whether they are intrinsic to the disease or not. The current diagnostic classifications use HA, oligo-anovulation and polycystic ovarian morphology (PCOM) at ultrasound. Whether HA is a necessary criterion remains controversial. By allowing the diagnosis of PCOS with only two items out of the three (HA, oligo-anovulation and PCOM), the so-called Rotterdam classification includes patients without overt HA.

There is indeed an urgent need to revisit these markers, but setting thresholds to define PCOM is particular. According to their symptoms, the 240 patients included in this study were divided into three groups: group 1 ($n = 105$) including women without HA (clinical or biological) and with regular menses (non-PCOS group), group 2 ($n = 73$) including women with only HA or only oligo-anovulation (presumption of PCOS) and group 3 ($n = 62$) including women with HA and oligo-anovulation, i.e. patients with genuine PCOS.



REFERENCE:

1. Adams J, Polson DW and Franks S (1986)
) Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism.
Br Med J

293

,
355
-359.

2. Al-Inani H, Aboughar M, Mansour R and Serour G (2003)
) Meta-analysis of recombinant versus urinary-derived FSH: an update.
Hum Reprod

18

,
305
-313.

3. American Diabetes Association (1998)
) Consensus Development Conference on Insulin Resistance.
Diabetes Care

21

,
310
-314.

4. Anonymous (1999)
) Glucose tolerance and mortality: comparison of WHO and American Diabetes Association diagnostic criteria.
Lancet

354

,
617
-621.

5. Asuncion M, Calvo RM, San Millan JL, Sancho J, Avila S and Escobar-Morreale HF
(

2000

) A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain.

J Clin Endocrinol Metab

85

,

2434

-2438.

6. Azziz R, Hincapie LA, Knochenhauer ES, Dewailly D, Fox L and Boots LR (1999

) Screening for 21-hydroxylase deficient non-classic adrenal hyperplasia among hyperandrogenic women: a prospective study.

Fertil Steril

72

,

915

-925.

7. Azziz R, Ehrmann D, Legro RS, Whitcomb RW, Hanley R, Fereshetian AG, O'Keefe

M and Ghazzi MN (

2001

) PCOS/Troglitazone Study Group. Troglitazone improves ovulation and hirsutism in the polycystic ovary syndrome: a multicenter, double blind, placebo-controlled trial.

J Clin Endocrinol Metab

86

,

1626

-1632.