Risk estimation of gestational diabetes and diabetes mellitus of type -2 because of PCOD through Mathematical and Artificial Intelligence models

DOI : 10.36909/jer.ICCEMME.15553

[a, b]Vijay K. Yadav, [a]Nilam*

[a]Department of Applied Mathematics, Delhi Technological University,

Delhi 110042, INDIA.

[b]G. L. Bajaj Institute of Technology and Management, Greater Noida 201306, INDIA

*Email: drnilam@dce.ac.in; Corresponding Author.

ABSTRACT

Pre-existence of PCOD (polycystic ovarian disease) cause the severity of diabetes during pregnancy as gestational diabetes (GD) and post-pregnancy diabetes mellitus of type -2 (DMT-2). Early detection of PCOD may help manage the severity of diabetes mellitus in pregnancy and postnatal. This analysis conveyed to understand the pervasiveness of PCOD and its complication with diabetes mellitus and body mass index (BMI). A contextual and statistical study of the data extracted from kaggle.com in 541 patients (180 with PCOD and 361 without PCOD) of southern India has been done. The random forest (RF) technique of Artificial Intelligence (AI) model has been used to analyze the correlations among parameters. In the body mass index, 42% of 180 PCOD patients have ≥ 27 kg/m2 body mass, as waist-hip ratios are in the range of 0.80 – 1.00. With pre-existence PCOD, 35% of women are pregnant. It has observed that 84% pregnant women have the risk of developing gestational diabetes, and few women have the chance to develop diabetes mellitus of type-2. The results were analyzed by RF technique of AI through Karl Pearson’s coefficient of correlation. The patients struggling with PCOD, facing high BMI and high waist-hip ratio (0.80 – 1.00) risk of gestational diabetes,

thereof early detection and diagnosis of PCOD will reduce the risk of development of GD and DMT-2.

Key words: artificial intelligence, correlation coefficient, diabetes mellitus, polycystic ovarian disease, regression.

INTRODUCTION

One out of 5 Indian women (20%) is suffering from PCOD and is a very high risk to develop the complication of diabetes mellitus (The Hindu, 26 September 2019). There are many other worse effects of PCOD such as unwanted hair growth, skin darkness and unbalanced BMI etc. If the body mass index raises ≥ 27 kg/m^2, it is alarming and needs to balance the lifestyle through a balanced diet plan and physical exercises. Women with PCOD are always having a higher risk of becoming insulin resistance and give rise to DMT-2 (A. Dunaif et al., 1992, G. Gennarelli et al., 2000 & L.C. Papunen et al., 2000). It has been observed the GD is associated with the adverse effects during and postnatal situation and leads to the hyperglycemic problem and gravid undergoes various metabolic changes (B.E. Metzger et al., 2008, T.A. Buchanan et al., 2007 & L. A. Barbour et al., 2007). Because of type-2 diabetes mellitus the complication and risk factors of PCOD, GD and macrosomia have discussed for the Lebanon women. For the Netherlands severity of PCOD and their adverse effects of metabolic disorders were explained (G.S. Michella et al., 2019 & E.A. Rotterdam et al., 2004).

We go through the various research articles similar to the present work. It was observed that no such types of studies are available for PCOD patients of the Indian subcontinent, only few analysis are available, but they are based upon the clinical research data such as G. Shivaprakash et al., 2013, have discussed the association of polycystic ovarian syndrome with acanthosis nigricansis as a marker of risk of type-2 diabetes mellitus. As per our cognizance neither any mathematical analysis nor artificial intelligence models of machine learning for parametric correlations are available for the Indian PCOD patients.

As per existing research, we have noticed that the risk factor of gestational diabetes (GD) and diabetes mellitus of the type-2 (DMT-2) are high in preexisting PCOD patients. The present study is to develop an artificial intelligence technology through machine learning approach by random forest (RF) classifier, and analyze the impact & complications in the Indian women because of PCOD.

MATERIALS AND METHODS

Materials

Our study is incorporated of 541 patients out of which 180 patients having PCOD problem and 361 are without PCOD (Figure 1). The data is from the Kerala state of southern India extracted from www.kaggle.com.

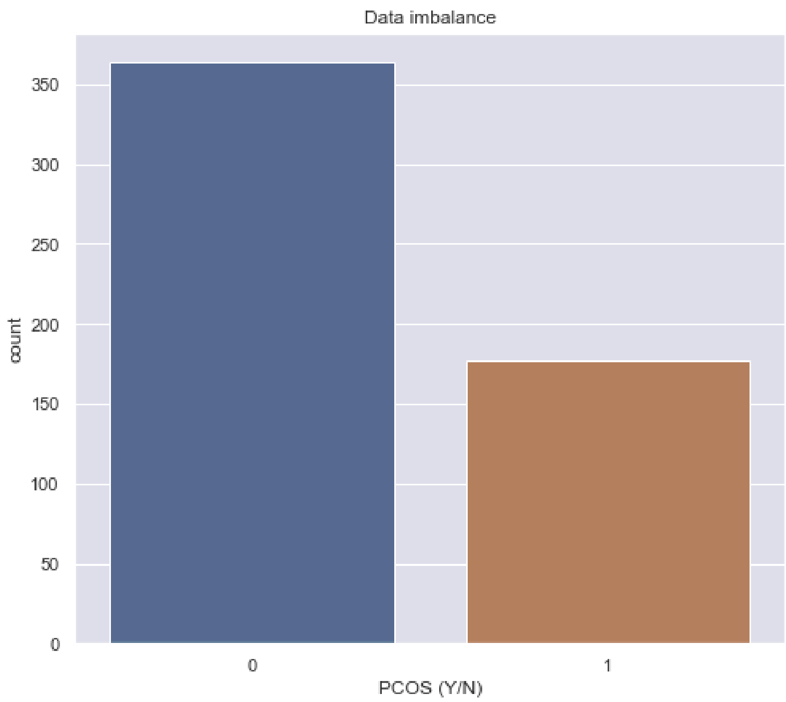

Figure 1 Patient with and without PCOD; (‘0’- without PCOD, and ‘1’- with PCOD)

This dataset of PCOD contains 45 parameters such as BMI (kg/m^2), beta hcg (miU/mL), age, weight, height, FSH (mIU/mL), pregnant(Y/N), waist to hip ratio, and FSH/LH etc.

First five values of incorporated parameters which have a significant impact to imbalance the metabolic situation and leads to GD and DMT-2 are presented in Table 1.

Table 1 Incorporated parameters of PCOD for the present study

PCOD (Y/N)	BMI (kg/m^2)	Age (year)	Pregnant (Y/N)	I-beta HCG (mIU/mL)	Waist/hip (ratio)	Hb (g/dl)	Follicle No.	FSH/LH
0	19.30	28	0	1.99	0.83	10.48	3	2.16
0	24.92	36	1	60.80	0.84	11.70	5	6.17
1	25.27	33	1	494.08	0.90	11.80	15	6.29
0	29.67	37	0	1.99	0.85	12.00	2	3.41
0	20.06	25	1	801.45	0.81	10.00	4	4.42

Methods

To mark a significant classification among the parameters, the random forest (RF) algorithms of AI are used. RF algorithms are also known as congregation of decision tree (DT) algorithms, it generates the clouds of various DT, and for every parameter, it starts from top nodes of the DTs and splits the dataset into their possible importance values (L. Breiman, 2001, A. Liaw et al., 2002 & V. Svetnik et al., 2015). Now the RF technique used the futuristic target related with the futures of the dataset of PCOD then each DTs gives his classification output. At the end RF algorithms, all the classification combines. It provides the final correlation among the parameters' feature, and the average of the regressions of the TDs considered ultimate output. The RF classifier used in Python 3 through jupyter notebook 6.0.3 of Anaconda which is open platform for the data related programming.

The Pearson product-moment correlation coefficient has been used for establishing the correlation among the parameters as described below (C. Jennifer et al., 2015). This is very strong tool to correlate the parameters, as they are strongly correlated, weakly correlated, and or no correlation for the values of $r \in (0,1], r \in (0,-1], and\ r = 0$ respectivly.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} \qquad (1)$$

Where; r = Pearson’s correlation coefficient,

$\bar{x}$ = the mean of the dataset x,

$\bar{y}$ = the mean of the dataset y,

n = number of values of parameter available in the dataset,

$\sigma_x = \sqrt{\frac{\sum(x-\bar{x})^2}{(n-1)}}$ is known as the standard deviation (S.D.) of x, and

$\sigma_y = \sqrt{\frac{\sum(y-\bar{y})^2}{(n-1)}}$ is known as the standard deviation (S.D.) of y.

The correlations between each parameter within and with other parameters have delineated using a graphical representation of different colour code known as a heat map (Figure 2). The heat map helps to visualizing the association of parameters within and with other parameters. The values of correlation coefficients are presenting in Table 3.

RESULT AND DISCUSSION

A total of 180 patients with PCOD (out of 541) have considered for the study, in which 63 out of 180 women are pregnant. The following table shows the ranges of the parameter which have incorporated in the study. There were no remarkable differences in the age, BMI, and waist-hip ratios of patients with/without PCOD. The ranges of incorporated parameters in study are shown in the Table 2.

A significant difference was seen in pregnant women with preexisting PCOD in the form of gestational diabetes (77%, 53 out of 63) in comparison with those who have not diagnosed with PCOD.

The coefficient of correlation among the parameters is lies in the range $-1 \leq r \leq 1$, if values of the coefficient are lies $-1 \leq r < 0$, then the parameters are conversely associated, $r = 0$ means that no correlation shows in parameters and for $0 < r \leq 1$ showing strong correlations (P. Schober et al., 2018). From Table 3, it is clearly visible that PCOD has a positive association with every parameter except FSH/LH, so in PCOD patient metabolic disorders have been noticed surly. As metabolic disorders caused the complication during pregnancy and it is also lead to the postnatal ferocity of diabetes mellitus type-2.

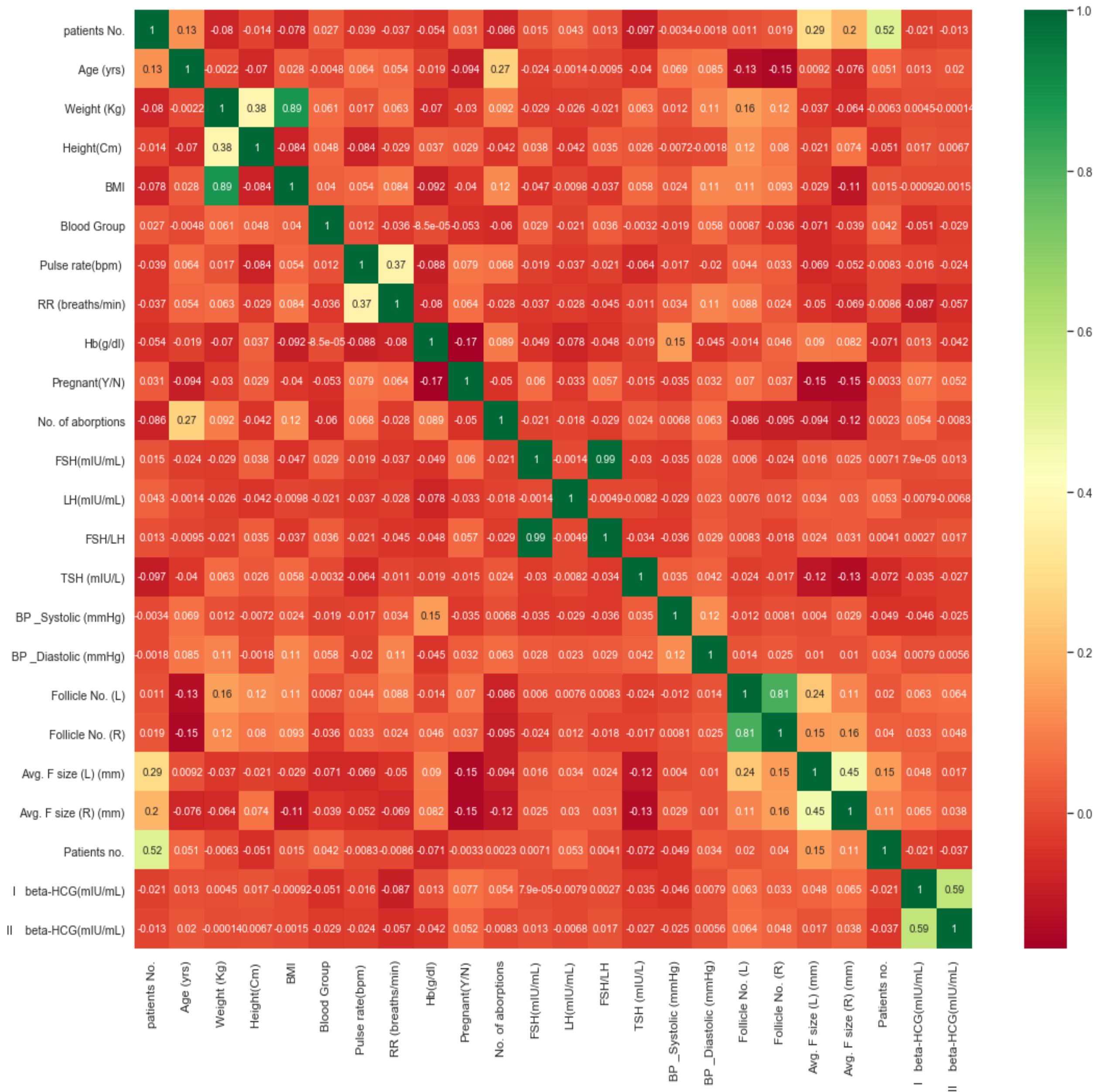

Figure 2 Heat map of dataset visualizing the correlation among the parameters

The box plots are the scenic view of the distribution of the data for various parameters (D. F. Williamson et al., 1989).

The box plots are indicating that how the data of parameters have been distributed alongside various statistical parameters such as mean, median, and quartiles (Figure 3). The central line of box plot indicating the median of the parameter own and if the cable is in middle of the box, then numerous amounts of data are distributing symmetrically about mean & median as endpoints are the quartile values.

Table 2 The range of the parameters

	With PCOD	Without PCOD
Age (years)	35 ± 12	35 ± 10
BMI (kg/m^2)	30 ± 8	30 ± 7
Waist / hip (ratio)	0.8 ± 0.2	0.7 ± 0.2
GD (%)	84	11

Table 3 Values of Pearson correlation coefficient within and with parameters

	PCOD (Y/N)	BMI (kg/m^2)	Age (year)	I-beta HCG (mIU/mL)	Hb (g/dl)	Follicle No.	FSH/LH
PCOD(Y/N)	1.00	0.135	0.128	0.017	0.068	0.592	- 0.024
BMI(kg/m^2)	0.135	1.00	0.027	- 0.001	- 0.09	0.113	- 0.037
Age (year)	- 0.17	0.027	1.00	0.013	- 0.02	- 0.129	- 0.010
I-beta HCG (mIU/mL)	0.017	- 0.001	0.013	1.00	0.013	0.062	0.002
Hb (g/dl)	0.068	- 0.092	- 0.019	0.013	1.00	- 0.013	- 0.047
Follicle No.	0.592	0.113	-0.129	0.062	- 0.013	1.00	0.133
FSH/LH	- 0.024	- 0.037	- 0.010	0.002	- 0.047	0.133	1.00

It was observed in the present work that 75% among the PCOD patients had BMI of greater than 26.40 (kg/m^2). The greater BMI were more susceptible to develop the gestational and diabetes mellitus of type-2 (WHO Expert Consultation, 2004). From the analysis of the data through RF algorithm among the PCOD victims it has been identified that in 50% patients the FSH/LH values are at 2.30, and 75% have waist-hip ratio equal to 0.93 which are at the higher side of the obesity. The values of follicle no. in the left (L) and right (R) ovaries were in the range of 5.4 ± 3.2 and 5.9 ± 4.0, respectively. According to a study has done in the Brazil, states that the obese person has greater chance to develop the metabolic disorder which leads to complications of GD and DMT-2 (L. M. B. Araújo et al., 2002).

Therefore these types of similar studies further validate that the patient suffering from PCOD with higher BMI and obesity are at higher risk of GD and DMT-2.

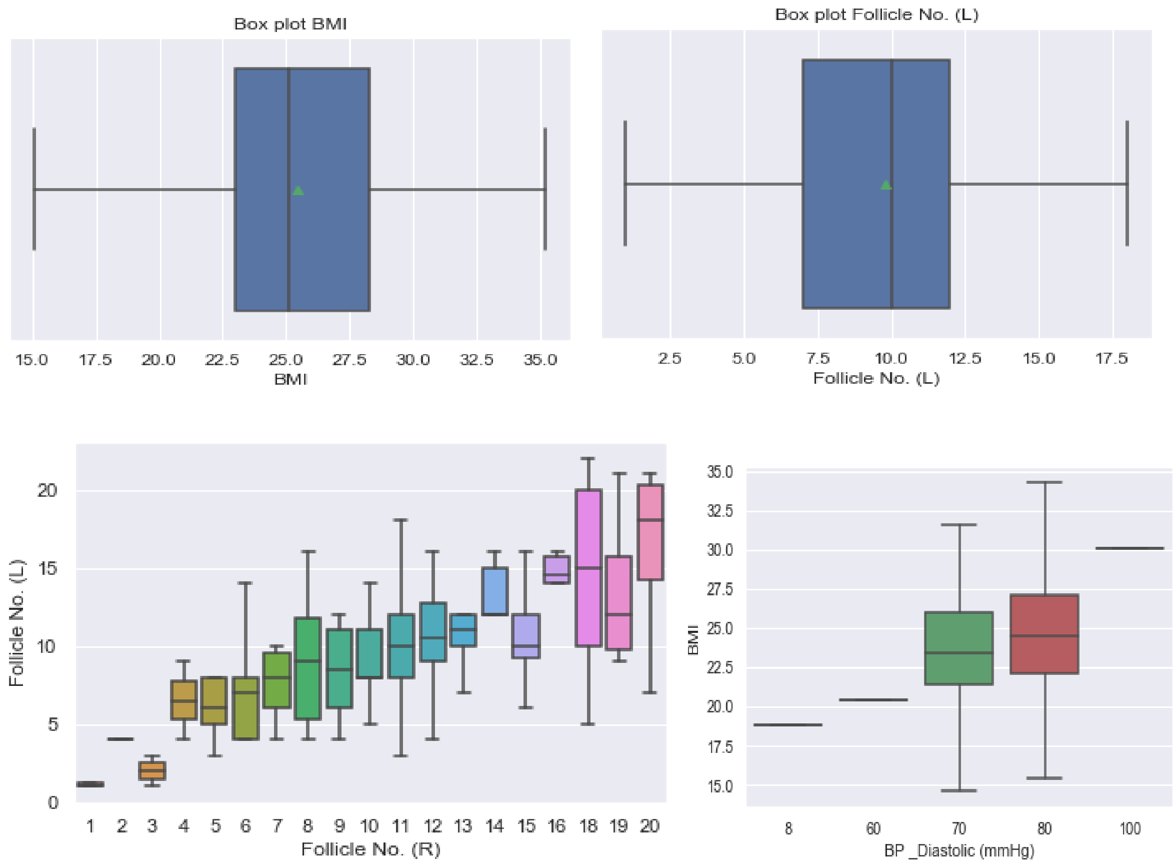

Figure 3 Box plots of the parameters: BMI, Follicle No. (L), Follicle No. (L) Vs Follicle No. (R) and Blood Pressure (mmHg)

CONCLUSION

PCOD propose a fascinating perspective for changing the way of living life. Among the PCOD patients who have obesity with BMI greater than 27 kg/m^2, and with the ratio of waist-hip between 0.8 ± 1.0 are at higher risk of diabetes mellitus of type -2. Those PCOD patients who are pregnant and at the higher side of the obesity (ratio of waist to hip is equal or more than to 0.8) are also in the risk zone of developing gestational diabetes. Hence perceptible symptom inspection marker such as PCOD can recognize the patients with a higher risk of DMT-2 and GD. This study would help energize the discussions about changes in the daily lifestyle with some modification at the primary care and diagnosis.

REFERENCES

The Hindu, 26 Sep. 2019. https://www.thehindu.com/sci-tech/health/one-in-five-indian-women-suffers-from-pcos/article29513588.ece.

Dunaif, A., Segal, K. R., Shelley, D. R., Green, G., Dobrjansky, A. and Licholai, T. 1992. Evidence for distinctive and intrinsic defects in insulin action in polycystic syndrome. Diabetes 41:1257–1266.

Gennarelli, G., Holte, J., Berglund, L., Berne, C., Massobrio, M. and Lithell, H. 2000. Prediction models for insulin resistance in the polycystic ovary syndrome. Hum Reprod 15: 2098–2102.

Morin-Papunen, L. C., Vauhkonen, I., Koivunen, R. M., Ruokonen, A. and Tapanainen, J. S. 2000. Insulin sensitivity, insulin secretion, and metabolic and hormonal parameters in healthy women and women with polycystic ovarian syndrome. Hum Reprod 15: 1266–1274.

Metzger, B. E., Lowe, L. P. and Dyer A. R. et al. 2008. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med; 358: 1991–2002.

Buchanan, T. A., Xiang, A., Kjos, S. L. and Watanabe, R. 2007. What is gestational diabetes? Diabetes Care; 30 (Suppl 2): S10511.

https://www.ncbi.nlm.nih.gov/pubmed/17596457.

Barbour, L. A., McCurdy, C. E., Hernandez, T. L., Kirwan, J. P., Catalano, P. M. and Friedman, J. E. 2007. Cellular mechanisms for insulin resistance in normal pregnancy and gestational diabetes. Diabetes Care; 30 (Suppl 2): S112–9.

https://www.ncbi.nlm.nih.gov/pubmed /17596458.

Michella, G. S., Zeina, M., Layal, A. F., Angelique, K. S. and Pierre, A. Z. 2019. Gestational Diabetes mellitus and macrosomia predispose to the diabetes in the Lebanese population. Journal of Clinical & Translational Endocrinology.16, 100185.

Rotterdam, E. A., SPCWG. 2004. Revised 2003 consensus on diagnostic criteria and longterm health risks related to polycystic ovary syndrome. Fertil Steril; 81(1):19–25.

Shivaprakash, G., Ashwin, K., Pallavi, S., Prabha, A., Rathanakar, U. P., Gopalakrishna, H. N. and Jagadish, R. P. 2013. Acanthosis Nigricansin PCOS and its Relation with Type 2 Diabetes Mellitus and Body Mass. Journal of Clinical and Diagnostic Research.Vol-7(2): 317-319.

https://www.kaggle.com/prasoonkottarathil/polycystic-ovary-syndrome-pcos.

Breiman, L. 2001. Random Forests, Mach. Learn. 45(1): 5–32.

Liaw, A. and Wiener, M. 2002. Classification and regression by random forest. R. News 2, 18–22.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P. and Feuston, B. P. 2015. Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inform. Comput. Sci. 43, 1947 – 1958. https://doi.org/10.1021/ci034160g.

Jennifer, C. 2015. Pearson's Product Moment Correlation: Sample Analysis. University of Hawaii at Mānoa School of Nursing.

Schober, P. and Boer, C. 2018. Correlation Coefficients: Appropriate Use and Interpretation, Anesthesia & Analgesia.Vol.126, no. 5.

Williamson, D. F., Parker, R. A. and Kendrick, J. S. 1989. The Box Plot: A Simple Visual Method to Interpret Data. Annals of Internal Medicine. Volume 110, Number 11.

WHO Expert Consultation, 2004. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet. 363: 157-63.

Araújo, LMB., Porto, M. V., Netto, E. M. and Ursich, M. J. 2002. Association of acanthosis nigricans with race and metabolic disturbances in obese women. Brazilian Journal of Medical and Biological Research. 35: 59-64.