Sapporo Medical Journal
Volume 55, Issue 09, September 2021

How to Interpret Normality Distribution in a Hormonal Ratio: X/Y Versus Y/X

Samih Abed Odhaib[1], Haider A[1], Abbas A[1]

College of Medicine, University of Basrah[1]

Abstract— **Background:** Hormonal ratios have analytical, interpretational importance. Our objective is to assess the best statistical model to interpret normality in a hormonal ratio. **Methods:** This descriptive review relied on hormonal data (testosterone T and estradiol E2) and (T/E2 and E2/T) from 1198 premenopausal women, through a systematic approach for testing normality, skewness, kurtosis, Kolmogorov–Smirnov (K-S), Shapiro-Wilk (S-W) which assume normal distribution.We tested the graphical methods for testing normality. Log-transformation, and correlation analyses, before and after adjustment for a proposed composite factor was done.ANCOVA and univariate linear regression tested the dependent variables and the outcome. **Results:** The data lacked normality graphically (markedly right skewed and positively leptokurtic), and on K-S and S-W. The log-transformation decreases (not eliminate) the normality deviation, which can be proven graphically but not in K-S and S-W, due to the outliers. Correlation analyses gave unsatisfactory results during adjustment for the composite factor. Parametric tests make more rigorous normality assumptions than that of the non-parametric tests, and could be applied for testing. ANCOVA used a more complex interplay to describe normality dispersion during adjustment, to provide an explainable robust normality assumption. Additional descriptive results were obtained from the linear regression analysis, similar to correlational analysis with different significance level. **Conclusion:** There is no privilege of one test over another to evaluate normality. We can use combined approach to reach near-normality especially in large sample data. ANCOVA and linear regression appeared to be more descriptive and genuine in preserving normality concept.

Keywords: Analysis of Covariance; Kolmogorov–Smirnov; Linear Regression; Normal Distribution; Shapiro-Wilk.

Introduction

Any hormonal ratio is calculated by dividing a numerator (X) by a denominator (Y), which represent the raw values hormones. The ratio distribution depends on the mathematical position of the hormone, i.e., X/Y or Y/X).[1]

One of the most popular reciprocal relationships is that of testosterone (T) and estradiol (E2) which represents the indirect expression of the aromatase enzyme activity that impacts levels of sex hormones and vascular, endothelial, and bone health.[2-4] Typically, there is a balance between T and E2 with a T/E2 ratio = 10.[5]

Hormonal ratios have analytical and interpretational importance. The main concerns are the representativeness of the ratio, its normality distribution, and statistical models for interpretations.[1,6]

When the normality assumption is violated, it is typically referred to as heteroscedasticity or heterogeneity of variance. It leads to incorrect standard error (SE), which can cause inflated type I error rate or decreased statistical power. It is imperative to manage heteroscedasticity effectively because failure to address this may affect substantive conclusions.[7,8]

The study aims to assess the best statistical approach to interpret the normality distribution in a hormonal ratio, using T and E2 as numerators and denominators interchangeably.

Methods

We performed a descriptive study to examine the distributions and behavior of (T), (E2), and their ratios, i.e., (T/E2) and (E2/T) after different adjustments in different statistical models. Exemplary T

and E2 actual data came from 1198 premenopausal women who attended Faiha Specialized Diabetes Endocrine and Metabolism Center (FDEMC)- Basrah- Southern Iraq, from January 2019- January 2021, with different reproductive complaints.

Various nuisance factors (outliers, normality violation, and heteroscedasticity), are encountered with big sample size data, and change the central tendency of the data. It is important to assess their robustness to mitigate their influence.

Heteroscedasticity occurs with normality assumption when population variances differ. In contrast, homoscedasticity is the prerequisite assumption to warrant the accuracy of SEs and asymptotic covariances of parameters in general linear models (GLM). When dealing with regression, homoscedasticity indicates that the DV's variance does not depend on the value of the independent variable.[7,8]

First, we perform normality (Gaussian) distribution model testing to interpret the different distributions of the hormonal ratios which were non-normally distributed. Our approach included:

General Description of the Distribution

We described the distribution of raw T, E2, and T/E2 and E2/T ratios after unifying the units (ng/dL to pg/mL). Then, we define the skewness (symmetry of the distribution) of each variable individually and in the ratio, and in which direction, i.e., positive (right-skewed distribution) or negative direction (left-skewed distribution). We defined kurtosis (peakedness of the distribution) of each variable individually and in the ratio, and in which direction, i.e., positive leptokurtic (high narrow peak and fatter tails), or negative platykurtic (lower broader peak and thinner tails). Then we plot the ratios against each other.

Theory-Driven Normality Testing

We used Kolmogorov–Smirnov (K-S) and Shapiro-Wilk (S-W) normality tests, which assume a normal distribution.Therefore, if the test is significant (p<0.05), the sample data distribution is non-normal.[9,10] The K-S statistics were corrected by (Lilliefors Significance Correction). The Lilliefors (LF) test is the preferred modified version of the K-S test, in which the probability of type 1 error is less because it adjusts the original K-S test for the population size.[11]

According to the Central Limit Theorem (CLT), if the sample data are approximately normal, its distribution will be normal. The two tests undertaken together in SPSS provide contrasting results about the null hypothesis. The K-S test calculates the maximum distance between the cumulative frequency curve of the analyzed data set and the theoretical normal frequency curve. S-W test is computationally more complex, but it is more powerful for detecting departures from normality than LF-corrected K-S and other frequently used tests.[12,13]

Normal Quantile-Quantile (Q-Q) Plots

We used the normal (Q-Q) plots for the graphical display of the normality distribution. The normal (Q-Q) plot is the most commonly used and effective diagnostic tool for checking data normality. This scatterplot compares our actual data with a theoretical normally distributed data set. The normally distributed data points will closely align with the diagonal reference line. Otherwise, in non-normal distribution, the points will stray from the diagonal line.[12]

Data Transformation and Adjustment

The log-transformation of the data was done, with resultant (new) distributions. Age, body weight, and height were the proposed covariates for our distribution. We used the (Dimension Reduction) by Principal Component Analysis (PCA) to create a composite covariate (Composite Body Shape Factor) with different component matrixloadings.

PCA resolves co-linearity and reduces the dimensions for high-dimensional data analysis. PCA transforms any set of potentially correlated variables to linearly uncorrelated variables.[14]

Adjustment of the DVs for this composite factor eliminates the influence of each covariate on the outcome as if all the cohort had the same age, body weight, and height. The dimensionality of the transformed data is reduced to achieve similar model fitting.[15]

Then we used the parametric (Pearson's r) and non-parametric (Spearman's rho, Rs) partial and bivariate correlation analyses to test the DVs after adjustment for the composite factor.

Analysis of Covariance (ANCOVA)

ANCOVA was developed to reduce error variance by combining the advantages of two highly acclaimed procedures of analysis of variance (ANOVA) and multiple linear regression.[16]

ANCOVA is a versatile and advanced regression technique. This method can compute and analyze any relationship between numerator and denominator even when the ratio is useless or in conditions when we need to standardize the numerator to the denominator.[17]The classic method makes several restrictive assumptions: (1) the regression lines are parallel; (2) for each regression line there is homoscedasticity; (3) the variance of the DV is the same for both groups; (4) normality; and (5) a straight regression line provides an adequate approximation of the true association.[8]

We identified T, E2, and their reciprocal ratios as DVs to implement ANCOVA. Outcomes like hirsutism, female pattern hair loss, and menstrual irregularity were considered as the fixed factors. In this study, we used hirsutism only for simplification. We adjusted the DVs for the composite body shape factor to eliminate the covariates' effect.

We applied the univariate linear regression analysis- full factorial model- to test the data during adjustment for the composite factor.

Then we used Levene's test to evaluate the error distribution. Levene's test is a unique application of the GLM, where the errors are correlated with known weights. The test is implemented by locating mean- or median-based residuals from ANOVA, developing absolute values of them, and reanalyzing the absolute values using the same original ANOVA tool.[18]

Finally, the estimated marginal means were evaluated. This mean is a hypothetical mean which is unaffected and unbiased by the covariates.

Linear Regression Analysis

Thirdly, we examined the ordinary linear regression platform in SPSS to assess the cumulative parametric change in Pearson's Coefficient of the parametric DVs after adjusting for outcomes like hirsutism and composite factor. We had similar results to the partial correlation analysis.Simple linear regression is a mathematical technique in which there is a single outcome or DV and a single predictor or independent variable with a linear or straight-line function.[19]

Results

Normality (Gaussian) Distribution Model Testing

Table 1 and Figure 1 described the distributional properties of the raw T, E2, and corresponding ratios. The T, E2, T/E2, and E2/T lacked normality as evident by the visual inspection of Figure 1 (Panels A-D), with marked positive right-skewed distribution and positive leptokurtic properties. Different interpretations of the distribution came from both the K-S and the S-W normality tests, which provided additional evidence of deviation from the normality, with $p<0.001$.

Figure 2 described the (Normal Q-Q Plots), which provide additional visual analysis of the data distribution along a diagonal reference line. The (Panels A – D) describe the data distribution of T, E2, T/E2 ratio, and E2/T ratio, respectively. The distribution is not a normal distribution, which confirmed the initial finding in Table 1 and Figure 1 (Panels A – D).

The log-transformation of the ratios decreases the deviation from the normality of the ratio, i.e., minimal skewness and kurtosis. Still, both K-S and S-W reject the null hypothesis and prove that even

with such log-transformation, there was a statistically significant deviation from a normal distribution ($p < 0.001$) in both tests (Table 1).

In Figure 1 (Panels E and F), the data of the log-transformed ratios appeared at the first instance as normally distributed. Still, a careful assessment of the characteristics of the normal (Q-Q) plots in Figure 2 (Panels E and F) showed that the outliers in the extremes of both log-transformed ratios prevent the normal distribution because they did not align along the diagonal reference line.

The relationship between the two reciprocal ratios was non-linear, preserving the position of the outliers in both ratios, i.e., patients with the highest T/E2 ratio ultimately had the lowest E2/T ratio and vice versa. The same effect was evident even after log-transformation of both ratios (Panel C and D).

We created the (Composite Body Shape Factor) to adjust our data to the possible or the proposed covariates like age, body weight, and height together. Table 2 used both parametric and non-parametric partial and bivariate correlation analyses, respectively, after adjustment for the composite factor, which provide us with different levels of significance. Although the results were satisfactorily descriptive, neither parametric nor non-parametric correlations after adjustment for the composite factor were adequate to yield acceptable results because both Pearson's and Spearman's Coefficients were below 0.3, which is the minimally accepted level for correlation.

Analysis of Covariate (ANCOVA):

In Levene's test for the quality of variance, if the $p>0.05$ for any DV, this meant that the variance was not statistically significant, and vice versa. The significance levels of both log-transformed ratios approach around the significance and its effect on distribution cannot be neglected.

In Table 4, we tried to answer whether the outcome had been significantly affected by the changes in the DVs during adjustment for the composite factor. The answer will be yes if the $p \leq 0.05$ and no if the $p>0.05$. After adjustment, statistically significant relationships before adjustment and high observed power levels retained this characteristic property. All the estimated marginal means levels underwent some changes during the ANCOVA at different degrees.

Table (3) described the different levels of DVs in women with a (fixed) outcome like hirsutism. To measure the effect of any factor or a covariate, we used ANCOVA through the univariate linear regression model analysis before and after adjustment for the composite factor (Table 4).

Linear Regression

The linear regression analysis gave identical results to that of the bivariate correlational analysis in Pearson's Coefficient levels, although with different levels of significance, i.e., 1-tailed level of significance. Some of the relationships lost their significance after further adjustment for the proposed covariates (Table 5).

Discussion

The interpretation of the interplay of T and E2 in a ratio is matter of debate. It is familiar to use the values of reciprocal ratios with different distribution in different endocrine conditions; that is why we tried to provide a roadmap review to address non-normality and heteroscedasticity in the hormonal ratios.

Mathematically, there is no simple way to predict the mean and SD of a ratio (X/Y) given the mean and SD of its inverse (Y/X).[20]The relation between the variables in a given ratio is complex or paradoxical, i.e., if the relationship between the numerator and the denominator is absent, then the calculation of the ratio itself will generate a relationship between the ratio and the denominator. In that case, a ratio will be operative only when the relationship demonstrates a straight linear that intersects the origin.[17]

The ratio (X/Y) illustrates the changes in the numerator (X) rather than the denominator (Y), with the assumption of linearity and normality in distribution; otherwise, the ratio will misrepresent the

genuine relationship between them and will render it meaningless, with unreliable or invalid interpretation.[12,17]However, a non-linear relationship does not necessarily entail no association between the variables.[21]

Although estimated parameters remain unbiased and consistent with heteroscedasticity, the estimated covariance matrix among the parameter estimates will be incorrect, leading to a low statistical power or inflated type I error rates.[7] Methods that assume homoscedasticity may incorrectly estimate the SE when there is real heteroscedasticity. As we consider more and more complicated designs, heteroscedasticity becomes an increasing concern.[8,15]

To assess any random sample size for normal distribution, we can use graphical methods by the visual inspection (Q-Q-plots, histograms, and boxplots), descriptive numerical methods(skewness and kurtosis), and formal theory-driven normality tests.[12,13]

However, choosing the correct study designs and proper statistical models is often challenging to assume the data normality and whether there is homo- or heteroscedasticity.[15] The approach may fail because of the lack of enough power to detect violation in normality, whatever the sample size was.[8]

In this article, we tested different statistical methods using real hormonal data from a large sample (1198 women), to have the minimal dispersion of the data. The assumption of normality is inessential when the sample size is large enough for the CLT to be at work.[22] In large sample size settings, linear regression models are reasonably robust to normality violations, and hence arbitrary bias-inducing outcome transformations are usually unnecessary. Instead, researchers should focus on detecting model miss-specifications such as outliers values, high leverage, heteroscedasticity, correlated errors, non-linearity, and interactions, which may bias results irrespective of sample size.[23]

The distributions of the original variables T, E2, T/E2, and E2/T were already slightly right-skewed and leptokurtic, which significantly deviates from normality. This distribution pattern is typical for ratio measures[24,25] due to (blow-up phenomenon) since ratio values increase exponentially as the denominator decreased, leading to outliers in the right tail of the distribution.[26] The minimal value of a hormonal ratio cannot be zero or negative values. In a normal distribution, skewness and kurtosis are equal to zero.[12]

The most common normality tests in statistical software are the S-W test, K-S test, Anderson-Darling (A-D) test, and LF test. Each test had its conditions for normality testing.[12] We used two normality testing, the S-W test and the K-S test with LF significance correction. These tests differ in the complexity of the normality, skewness, and kurtosis.[27]

With our large population sample size, we overcame the minimal requirements of these normality tests, which were 10 for the K-S test, and 50 for the S-W test, which ultimately robust the power of these two tests.[12]In K-S and other normality tests, slight deviations from normality tend to reach significance with large samples[1] such as the present.

With small sample size (n=10 - 50) , normality tests lack power to detect departures from normality.[12,13] Similarly, S-W outperforms K-S for different distributions with kurtosis >3 (leptokurtic distributions),[12] as in Table 1.

Although the graphical presentation in Figures 1 and 2 (Panels 1-4) illustrated the lack of normal distribution, it is not advised to depend solely on them to conclude the distribution of the data. The interpretation should include the combined results of graphical methods, formal normality testing, and coefficients of skewness and kurtosis in different sample sizes.[9,12,28,29]

We performed three types of data transformation, algorithmic, parametric, and non-parametric. The data transformation is the most often recommended remedy to change the heteroscedastic distribution in non-normal data towards normality.[22,30,31] Log-transformation in regression stabilizes the variance over target levels, turning it homoscedastic, and turns the measurement domain from arithmetic to proportional.[32] In general, parametric transformations manage non-normality as informative, while non-parametric or algorithmic transformations manage it as a nuisance.[22]

By transforming a DV in a ratio, the goal is to use ordinary least-squares (OLS) estimators without worrying about the sensitivity of the skewness.[22,30]

These misspecifications are addressed after transformation, even with the preserved linear model in OLS, but in a rescaled pattern.[22]Also, the transformation changes the original hypothesis being tested and the original construct that it measures,[33] and alters the fundamental relationships among variables, and subsequently changes the errors distribution.[32]

Even after the transformation, we may fail to have a normality distribution, as illustrated in Table 1 and Figure 2 (Panels E and F). However, they demonstrated less skewness and kurtosis during the algorithmic transformation of both ratios in reciprocal forms.[8]

The outliers were another problem because this simple approach was unsatisfactory for outliers management. The data transformation might decrease or increase the outliers but not mitigate the effect of the outliers on the final interpretation.[8]

Researchers tried to transform the log-transformed parameter back to its original scale.[34] The mean estimated from log-transformed data can be transformed back to its original scale. In the log, the mean becomes geometric instead of arithmetic in the original data.[35]

We can transform the data using parametric and non-parametric testing, with and without adjustment for a composite factor. The choice between parametric versus non-parametric is generally predicated on the number and rigor of the assumptions (requirements) regarding the underlying study population.[36]

Parametric tests make more rigorous, powerful, and robust assumptions of normal distribution of the underlying population than that of the non-parametric tests and could be applied for testing.[12,36,37] The decision to apply a parametric versus a non-parametric test is complex, difficult, and controversial and depends on the original data itself.[9]

It is a common problem for any data set to find outliers that may affect the outcome. This problem will be augmented if there was a co-linearity between variables. We cannot simply manipulate these outliers because this will lead to unpredictable changes in the estimated coefficients or changes in sign or SE inflation, leading to poor statistical power.[8,15,38,39]

Figures 1-3 illustrated the outliers in the different situations during transformation. The outliers were constant in positions in the (long and thin) left tail of the distribution. The graphical presentation was illustrative for the outliers determination. The use of the mean and variance are known to be unsatisfactory.[38]

When Pearson's and Spearman's rho correlations used a correct estimate of the SE, they perform well in type 1 errors when there is no association. Independence implies homoscedasticity. While when a homoscedastic method rejects, it is reasonable to conclude that there is an association, but inferring the nature of the association, these methods can perform poorly.[8]

Another way to deal with the non-normally distributed data is by applying ANCOVA, which includes continuous and categorical variables as independent variables. It is a mixture model that combines the advantages of ANOVA and multiple linear regression model.[40,16]It is generally recognized that ANCOVA may considerably reduce the number of subjects required than an ANOVA design to attain the required precision and power.[16] Still, it inappropriately generates a higher bias.[15]

The inclusion of a covariate, which is highly correlated with response, can remove or reduce a considerable portion of errors. The explanation ability, significance of factors, mitigated heteroscedasticity is better with ANCOVA.[7,40,41]

ANCOVA assumes the normality, where the type 1 error is controlled even when the model is misspecified, which may lead to a biased outcome. In ANCOVA, it is difficult to accurately assess and evaluate outcome effects given the normative presence of covariate measurement error.[40-42] To solve this limitation, we have three (inadequate) options. First, assume normality, pretend nonexistence of the problem, use the OLS regression, and find unbiased results, which is not logical.

Second, apply blind correction methods, which may give unreliable results. And third, Use another statistical model, not use ANCOVA at all.[42]

Hess et al. advised using linear regression to manage this problem to a certain extent. Regression and correlation can be easily performed with most statistical and spreadsheet software, and C.I.s and statistical hypothesis tests can be calculated to describe the results.[43]

The main limitation of the regression model is its ability to accommodate an analysis between two unrelated variables in ratio, even when this ratio is useless. Yet, we cannot guarantee that the relation between the numerator and denominator is linear and pass the origin; otherwise, the interpretation of the ratio is useless.[17] Also, we did not use bootstrapping in its simple and wild form, which assumes a less restrictive assumption of the sample representation.[22]

Conclusions

When conducting statistical analysis for any hormonal ratio of any form, the researcher must adopt more than one statistical method to robust the normality assumption of the distribution. There is no privilege of one test over another to achieve normality. However, ANCOVA and linear regression appeared to be more descriptive and more genuine in preserving the normality concept.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not for- profit sectors.

Conflict of interest

The authors have no conflict of interest to declare

Ethical approval

The ethical committee of FDEMC provided ethical approval for the study (E2/19/2019) which was given on December 2018.

Authors contribution

SAO conceived and designed the study. SAO and HAA conducted the study. SAO and HAA provided research statistics and collected and organized data. AAM supervised the study process and provided logistic support. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

References

[1] Sollberger S, Ehlert U. How to use and interpret hormone ratios. Psychoneuroendocrinology 2016 Jan;63:385-97. doi: 10.1016/j.psyneuen.2015.09.031.

[2] Amato MC, Verghi M, Nucera M, Galluzzo A, Giordano C. Low estradiol-to-testosterone ratio is associated with oligo-anovulatory cycles and atherogenic lipidic pattern in women with polycystic ovary syndrome. Gynecol Endocrinol. 2011;27(8):579-86. doi: 10.3109/09513590.2010.495797.

[3] Gong Y, Xiao H, Li C, Bai J, Cheng X, Jin M, Sun B, Lu Y, Shao Y, Tian H. Elevated T/E2 ratio is associated with an increased risk of cerebrovascular disease in elderly men. PLoS One. 2013;8(4):e61598. doi: 10.1371/journal.pone.0061598.

[4] van Koeverden ID, de Bakker M, Haitjema S, van der Laan SW, de Vries JPM, Hoefer IE, de Borst GJ, Pasterkamp G, den Ruijter HM. Testosterone to oestradiol ratio reflects systemic and plaque inflammation and predicts future cardiovascular events in men with severe atherosclerosis. Cardiovasc Res. 2019;115(2):453-462. doi: 10.1093/cvr/cvy188.

[5] Spector IP, Carey MP, Steinberg L. The sexual desire inventory: Development, factor structure, and evidence of reliability. J Sex Marital Ther. 1996;22(3):175-90. doi: 10.1080/00926239608414655.

[6] Maninger N, Wolkowitz OM, Reus VI, Epel ES, Mellon SH. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS).

Front Neuroendocrinol. 2009 Jan;30(1):65-91. doi: 10.1016/j.yfrne.2008.11.002. Epub 2008 Dec 3. PMID: 19063914; PMCID: PMC2725024.

[7] Rosopa PJ, Schaffer MM, Schroeder AN. Managing heteroscedasticity in general linear models. Psychol Methods. 2013 Sep;18(3):335-51. doi: 10.1037/a0032553.

[8] Wilcox RR, Rousselet GA. A Guide to Robust Statistical Methods in Neuroscience. Curr Protoc Neurosci. 2018 Jan;82:8.42.1-8.42.30. doi: 10.1002/cpns.41.

[9] Motulsky H (2010). Intuitive biostatistics: a nonmathematical guide to statistical thinking. New York: Oxford University Press.

[10] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika 1965, 52 (3-4):591611, https://doi.org/10.1093/biomet/52.3-4.591.

[11] Lilliefors HW. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc 1967; 62:318, 399-402. DOI: 10.1080/01621459.1967.10482916.

[12] Razali NM, Wah YB.Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests.Journal of Statistical Modeling and Analytics 2011;2(1):21-33.https://www.nrc.gov/docs/ML1714/ML17143A100.pdf

[13] Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab. 2012;10(2):486-9. doi: 10.5812/ijem.3505.

[14] Jolliffe IT (2002). Principal Component Analysis. 2th ed. New York, NY: Springer.

[15] Yan L, Sun Y, Boivin MR, Kwon PO, Li Y. Revealing facts and avoiding biases: A review of several common problems in statistical analyses of epidemiological data. Front Public Health. 2016 Oct;4:207. doi: 10.3389/fpubh.2016.00207.

[16] Shieh, G. Power analysis and sample size planning in ANCOVA Designs. Psychometrika 2020;85:101-120. https://doi.org/10.1007/s11336-019-09692-3

[17] Curran-Everett D. Explorations in statistics: the analysis of ratios and normalized data. Adv Physiol Educ. 2013 Sep;37(3):213-9. doi: 10.1152/advan.00053.2013.

[18] O'Neill ME, Mathews KL. Levene tests of homogeneity of variance for general block and treatment designs. Biometrics. 2002 Mar;58(1):216-24. doi: 10.1111/j.0006-341x.2002.00216.x.

[19] Marill KA. Advanced statistics: linear regression, part I: simple linear regression. Acad Emerg Med. 2004 Jan;11(1):87-93. PMID: 14709436.

[20] Weltje, GJ. (2012). Quantitative models of sediment generation and provenance: State of the art and future developments. Sedimentary Geology 2012; 280: 4-20. https://doi.org/10.1016/j.sedgeo.2012.03.010.

[21] Pandis N.Correlation and linear regression.Am J Orthod Dentofacial Orthop 2016;149:298-9 2016; 149(2): 298-299.https://doi.org/10.1016/j.ajodo.2015.11.010

[22] Pek J, Wong O, Wong ACM. How to address non-normality: A Taxonomy of approaches, reviewed, and illustrated. Front Psychol. 2018 Nov;9:2104. doi: 10.3389/fpsyg.2018.02104.

[23] Schmidt AF, Finan C. Linear regression and the normality assumption. J Clin Epidemiol. 2018 Jun;98:146-151. doi: 10.1016/j.jclinepi.2017.12.006.

[24] Anderson DE, Lydic R. On the effect of using ratios in the analysis of variance. Biobehavioral Reviews 1977; 1(4):225-229.https://doi.org/10.1016/0147-7552(77)90024-9

[25] Atchley WR, Gaskins CT, Anderson D. Statistical properties of ratios. I. Empirical results. Systematic Biology 1976; 25(2):137-148. https://doi.org/10.2307/2412740

[26] Kane G, Meade N. Ratio analysis using rank transformation. Rev. Quant. Finance Account 1998; 10: 59-74. https://doi.org/10.1023/A:1008248130384

[27] Seier E. (2011) Normality tests: Power comparison. In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_421

[28] Das KR, Imon AHM. A brief review of tests for normality. American Journal of Theoretical and Applied Statistics 2016; 5(1): 5-12. doi: 10.11648/j.ajtas.20160501.12

[29] Kac M, Kiefer J, Wolfowitz J. On tests of normality and other tests of goodness of fit based on distance methods. Ann. Math. Statist. 1955; 26(2): 189-211. DOI: 10.1214/aoms/1177728538

[30] Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001 Jul;20(4):461-94. doi: 10.1016/s0167-6296(01)00086-8.
[31] Atkinson G, Batterham AM, Black MA, Cable NT, Hopkins ND, Dawson EA, Thijssen DH, Jones H, Tinken TM, Green DJ. Is the ratio of flow-mediated dilation and shear rate a statistically sound approach to normalization in cross-sectional studies on endothelial function? J Appl Physiol (1985). 2009 Dec;107(6):1893-9. doi: 10.1152/japplphysiol.00779.2009.
[32] Gao A, Fischer-Jenssen J, Wroblewski C, Martos P. Interpretation and implications of lognormal linear regression used for bacterial enumeration, J. AOAC Int 2020; 103(4):1105-1111, https://doi.org/10.1093/jaoacint/qsaa005
[33] Grayson D. Some myths and legends in quantitative psychology. Underst. stat. 2004, 3:2, 101-134, DOI: 10.1207/s15328031us0302_3
[34] St-Pierre AP, Shikon V, Schneider DC. Count data in biology-Data transformation or model reformation? Ecol Evol. 2018 Feb;8(6):3077-3085. doi: 10.1002/ece3.3807.
[35] Keene ON. The log transformation is special. Stat Med. 1995 Apr;14(8):811-9. doi: 10.1002/sim.4780140810.9
[36] Sheskin, D. J. (2000). Handbook of parametric and nonparametric statistical procedures (2nd ed.). Chapman and Hall/CRC.
[37] Speller V, Learmonth A, Harrison D. The search for evidence of effective health promotion. BMJ. 1997 Aug; 315(7104):361-3. doi: 10.1136/bmj.315.7104.361.
[38] Wilcox R (2017). Modern statistics for the social and behavioral sciences: A practical introduction. Second Edition. Boca Raton. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781315154480
[39] Wilcox R (2016). Introduction to robust estimation and hypothesis testing: A volume in statistical modeling and decision science. Fourth Edition. Elsevier Science, Academic Press.
[40] Kim HY. Statistical notes for clinical researchers: analysis of covariance (ANCOVA). Restor Dent Endod. 2018 Oct;43(4):e43. doi: 10.5395/rde.2018.43.e43.
[41] Bartlett JW. Robustness of ANCOVA in randomized trials with unequal randomization. Biometrics. 2020 Sep;76(3):1036-1038. doi: 10.1111/biom.13184.
[42] Culpepper SA, Aguinis H. Using analysis of covariance (ANCOVA) with fallible covariates. Psychol Methods. 2011 Jun;16(2):166-78. doi: 10.1037/a0023355.
[43] Hess AS, Hess JR. Linear regression and correlation. Transfusion 2017; 57: 9-11. https://doi.org/10.1111/trf.13928

Tables and Figures

Table (1): Descriptive and distributional properties of different hormonal variables in the study.

Variables	Mean ± SE	Range	Skewness ± SE[a]	Kurtosis ± SE	Normality Testing	
					Kolmogorov-Smirnov Statistic[b,c]	Shapiro-Wilk Statistic[b]
T	36.94 ± 0.60	0.29 - 147.20	1.21 ± 0.07	3.10 ± 0.14	0.065	0.932
E2	78.04 ± 1.47	4.10 - 349.00	2.36 ± 0.07	7.57 ± 0.14	0.190	0.787
T/E2	6.52 ± 0.16	0.23 - 44.23	2.19 ± 0.07	6.74 ± 0.14	0.138	0.803
E2/T	0.32 ± 0.01	0.02 - 4.38	4.98 ± 0.07	33.98 ± 0.14	0.238	0.543
Log T/E2	0.68 ± 0.01	-0.64- - 2.74	-0.401- ± 0.07	0.36 ± 0.14	0.041	0.989
Log E/2T	-0.68- ± 0.01	-2.74- - 1.19	0.30 ± 0.07	2.42 ± 0.141	0.045	0.974

Abbreviations: E2, Estradiol; SE, Standard Error; T, Testosterone

[a] The negative mark of skewness indicates the left skewness.

[b] p <0.001

[c] Corrected by Lilliefors Significance Correction

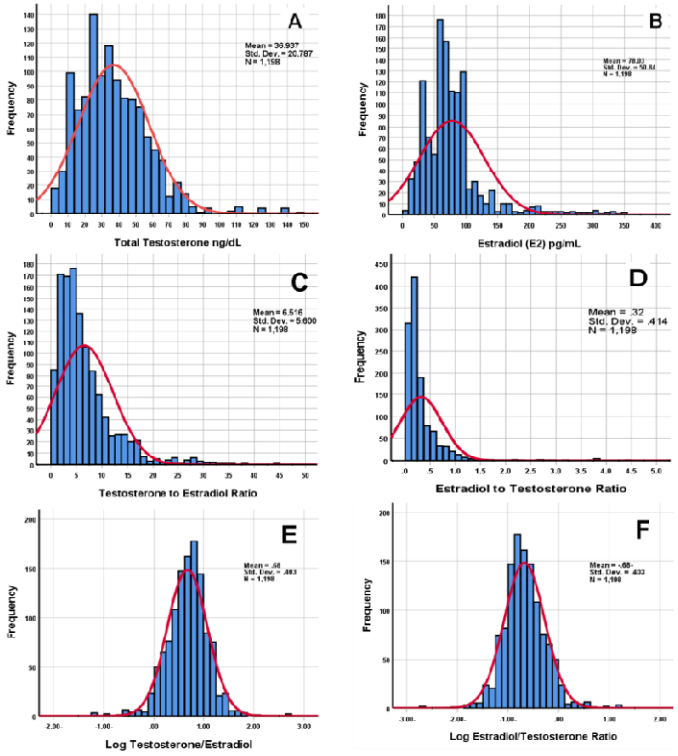

Figure 1: Frequency distribution of testosterone (Panel A), estradiol (Panel B), testosterone to estradiol ratio (Panel C), estradiol to testosterone ratio (Panel D), log-transformed testosterone to estradiol ratio (Panel E), and log-transformed estradiol to testosterone ratio (Panel F).

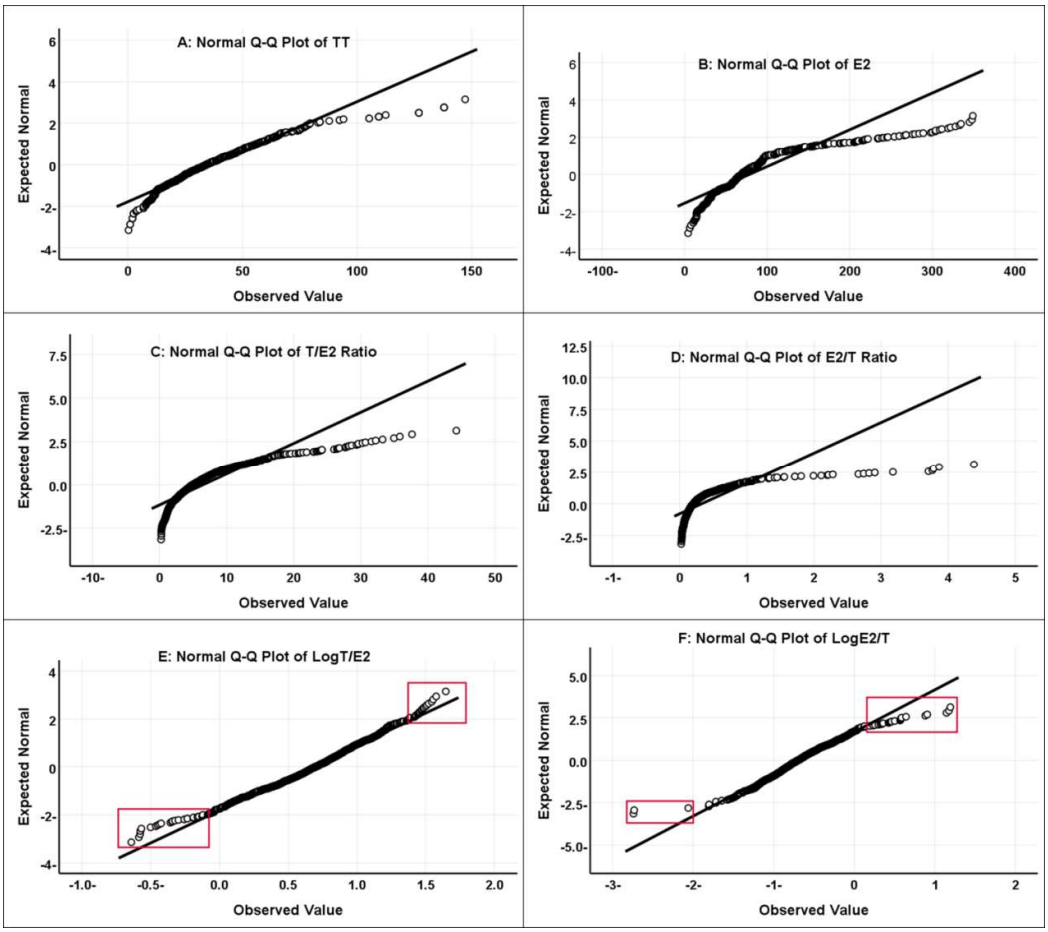

Figure 2: The normal (Q-Q) plots of the distribution of T (Panel A), E2 (Panel B), T/E2 ratio (Panel C), and E2/T ratio (Panel D). Panels E and F describe the distribution of the log-transformed ratio along the diagonal reference line. The outliers in both panels (in the red rectangles) affected the distribution to be non-normal distribution.

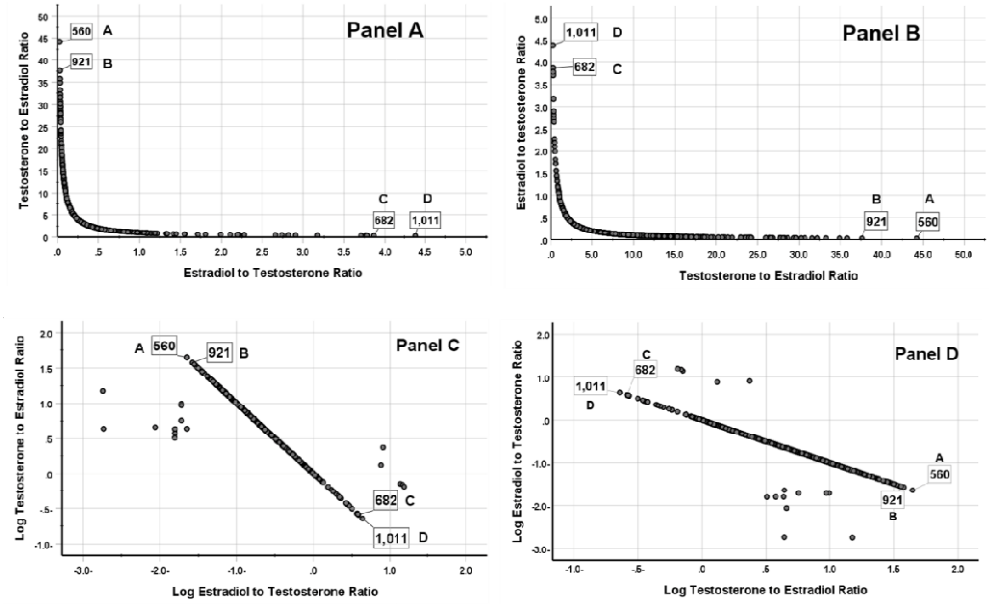

Figure 3: Plotting the ratios against each other. Panels A and B represent the non-linear distribution. The highest and the lowest values in both ratios had preserved their position in the opposite direction. Patients 560 and 921 had the highest T/E2 (Panel A),(44.231 and 37.625, respectively), and the lowest E2/T (Panel B) (0.02 and 0.03, respectively). Patients 682 and 1011 had the lowest T/E2 (Panel A) (0.228 and 0.259, respectively), respectively, but the highest E2/T (Panel B) (4.38 and 3.86, respectively). The rank within the two ratios was consistent but reversed.

Table (2): Parametric (Pearson's r) and non-parametric (Spearman's rho, Rs) partial and bivariate correlation analyses, respectively.[a]

Variables	Parametric correlation		Non-parametric correlation	
	Pearson's coefficient Adjusted for the Composite Body Shape Factor	2-tailed sig	Spearman's Coefficient Adjusted for the Composite Body Shape Factor	2-tailed sig
T	**0.001**	**0.981**	**0.004**	**0.900**
E2	**-0.022-**	**0.444**	**-0.062-**	**0.032**
T/E2	**0.055**	**0.057**	**0.028**	**0.338**
E2/T	**0.014**	**0.636**	**-0.028-**	**0.338**
Log T/E2	**0.023**	**0.428**	**0.028**	**0.338**
Log E/2T	**-0.016-**	**0.572**	**-0.027-**	**0.346**

Abbreviations: E2, Estradiol; SE, Standard Error; sig, significance; T, Testosterone

[a]The negative marks in Pearson's and Spearman's Coefficient define the direction not the value.

Table (3): Mean value of T, E2, and their ratios for the study group, with hirsutism as a possible outcome. The data were expressed as Mean ± SE.

Dependent variables	Total (n=1198)	Hirsutism[a] (n=834)	No Hirsutism (n=364)
T	**36.937 ± 0.601**	**37.786 ± 0.731**	**34.992 ± 1.043**
E2	**78.035 ± 1.469**	**77.930 ± 1.963**	**78.275 ± 1.778**
T/E2	**6.516 ± 0.162**	**6.945 ± 0.207**	**5.533 ± 0.237**
E2/T	**0.321 ± 0.012**	**0.311 ± 0.015**	**0.342 ± 0.021**
Log T/E2	**0.672 ± 0.011**	**0.696 ± 0.013**	**0.617 ± 0.018**
Log E/2T	**-0.678- ± 0.012**	**-0.700- ± 0.014**	**-0.627- ± 0.020**

Abbreviations: E2, Estradiol; SE, Standard error; T, Testosterone

[a] any outcome can be used as a fixed factor like Female pattern hair loss, menstrual irregularity, infertility, etc. We used hirsutism as an example in this study.

Table (4): The use of ANCOVA, univariate linear regression model analysis.

Dependent variables	Significance of Levene's Test adjustment		Significance of hirsutism as a fixed factor adjustment		Observed power Adjustment		Estimated nonadjusted marginal means[a]		Estimated adjusted marginal means[a]	
	Before	After	Before	After	Before	After	Hirsutism	No Hirsutism	Hirsutism	No Hirsutism
T	0.108	0.110	0.032	0.030	0.572	0.582	37.786 ± 0.719	34.993 ± 1.088	37.810 ± 0.722	34.938 ± 1.100
E2	<0.001	<0.001	0.914	0.981	0.051	0.050	77.930 ± 1.761	78.275 ± 2.666	78.058 ± 1.770	77.981 ± 2.694
T/E2	<0.001	<0.001	<0.001	<0.001	0.981	0.964	6.945 ± 0.193	5.533 ± 0.292	6.922 ± 0.194	5.586 ± 0.295
E2/T	0.919	0.926	0.227	0.191	0.226	0.257	0.311 ± 0.014	0.342 ± 0.022	0.310 ± 0.014	0.345 ± 0.022
Log T/E2	0.065	0.065	0.001	0.001	0.929	0.916	0.696 ± 0.013	0.617 ± 0.019	0.696 ± 0.013	0.617 ± 0.019
Log E/2T	0.051	0.051	0.004	0.005	0.819	0.804	-0.700- ± 0.014	-0.627- ± 0.021	-0.699- ± 0.014	-0.627- ± 0.021

Abbreviations: E2, Estradiol; T, Testosterone
[a] The means were expressed as (Mean ± Standard Error)

Table (5): Linear regression analysis using parametric correlation with Pearson's coefficient after adjustment for the hirsutism and Composite Body Shape Factor

Variables	Pearson's coefficient Adjusted for Hirsutism	1-tailed significant level	Pearson's coefficient Adjusted for hirsutism and the Composite Body Shape Factor	1-tailed significant level
T	**-0.062-**	**0.016**	**0.001**	**0.490**
E2	**0.003**	**0.457**	**-0.022-**	**0.222**
T/E2	**-0.116-**	**< 0.001**	**0.055**	**0.029**
E2/T	**0.035**	**0.114**	**0.014**	**0.318**
Log T/E2	**-0.099-**	**< 0.001**	**0.023**	**0.214**
Log E/2T	**0.083**	**0.002**	**-0.016-**	**0.286**

Abbreviations: E2, Estradiol; T, Testosterone