EDITORIAL

Correlation analysis in clinical and experimental studies

Análise de correlação em estudos clínicos e experimentais

Hélio Amante Miot[1]

It is common for researchers conducting clinical or biomedical studies to be interested in investigating whether the values of two or more quantitative variables change in conjunction in a given individual or object of study. In other words, whether when the value of one variable increases, the value of another tends to increase/ or, inversely, reduce, progressively. There are many different statistical tests that explore the intensity and direction of this mutual behavior of variables, known as correlation tests.[1,2]

The first step in analyzing correlations between two quantitative variables should be to look at a scatter plot, in order to discern whether there is a gradual variability between the sets of variables, whether this variation is monotonic (predominantly increasing or decreasing), if it follows a proportional tendency (linear), and whether the underlying distribution of the data is normal.[2-4] Different combinations of these premises indicate a need for different techniques for correlation analysis.

Figure 1 illustrates the distribution of values of four hypothetical variables (V1, V2, V3, and V4), which exhibit data that follow a normal distribution (Shapiro-Wilk, p > 0.32).

Variables V1 and V2 exhibit simultaneously increasing values, which are distributed around an underlying imaginary (ideal) straight line, which describes the trajectory of the data. It can be stated that there is a positive linear correlation between V1 and V2. For example, Rossi et al. identified a strong positive correlation ($\rho = 0.82$; $p < 0.01$) between cores on the Venous Symptoms Clinical Severity Scale and pain in chronic venous disease.[5]

In contrast, variables V1 and V3 exhibit antagonistic behavior: when the values of one increase, the values of the other reduce. It can be stated that there is a negative linear correlation between V1 and V3, just as Ohki and Bellen identified a moderate negative correlation ($\rho = -0.65$; $p < 0.01$) between average

regional temperature and the incidence of venous thrombosis.[6]

It can also be observed that the values for the correlation between V1 and V3 are closer to the imaginary straight line than the values for the correlation between V1 and V2. This invites the conclusion that the relationship between the values of the variables V1 and V3 is stronger than the relationship between V1 and V2, even though the directions are opposite.

Comparisons of the data for V4, whether with V1, V2, or V3, do not reveal gradually increasing or decreasing behavior. This leads to the conclusion that V4 does not exhibit a correlation with the other variables.

The most widely-used technique for evaluating the correlation between two quantitative variables is Pearson's product-moment correlation coefficient, or Pearson's *r*, which requires that both samples follow a normal distribution and that the relationship between the two variables is linear.[2,7] Failure to adhere to these prerequisites leads to erroneous conclusions, even when working with large sample sizes.

However, it is very common that samples of clinical and demographic data do not follow a normal distribution (for example, the distributions of income, quality of life indexes, disease severity indexes, years of study, and number of children). The most widely used options for investigating correlations between variables that do not exhibit normal distributions are the Spearman rank order correlation and the Kendall rank correlation coefficient (*Tau*-b), which substitute the original data for their ordered ranks.[2,7,8] These methods are also used in cases in which at least one of the variables has ordinal characteristics (for example, functional class, educational level, cancer staging, social class).

Another advantage of using the Spearman and Kendall nonparametric tests is that they are not restricted to linear correlations, as long as they exhibit monotonic behavior. In other words, they must exhibit

[1] Universidade Estadual Paulista – UNESP, Faculdade de Medicina de Botucatu, Departamento de Dermatologia e Radioterapia, Botucatu, SP, Brasil.

The study was carried out at Departamento de Dermatologia e Radioterapia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista (UNESP), Botucatu, SP, Brazil.
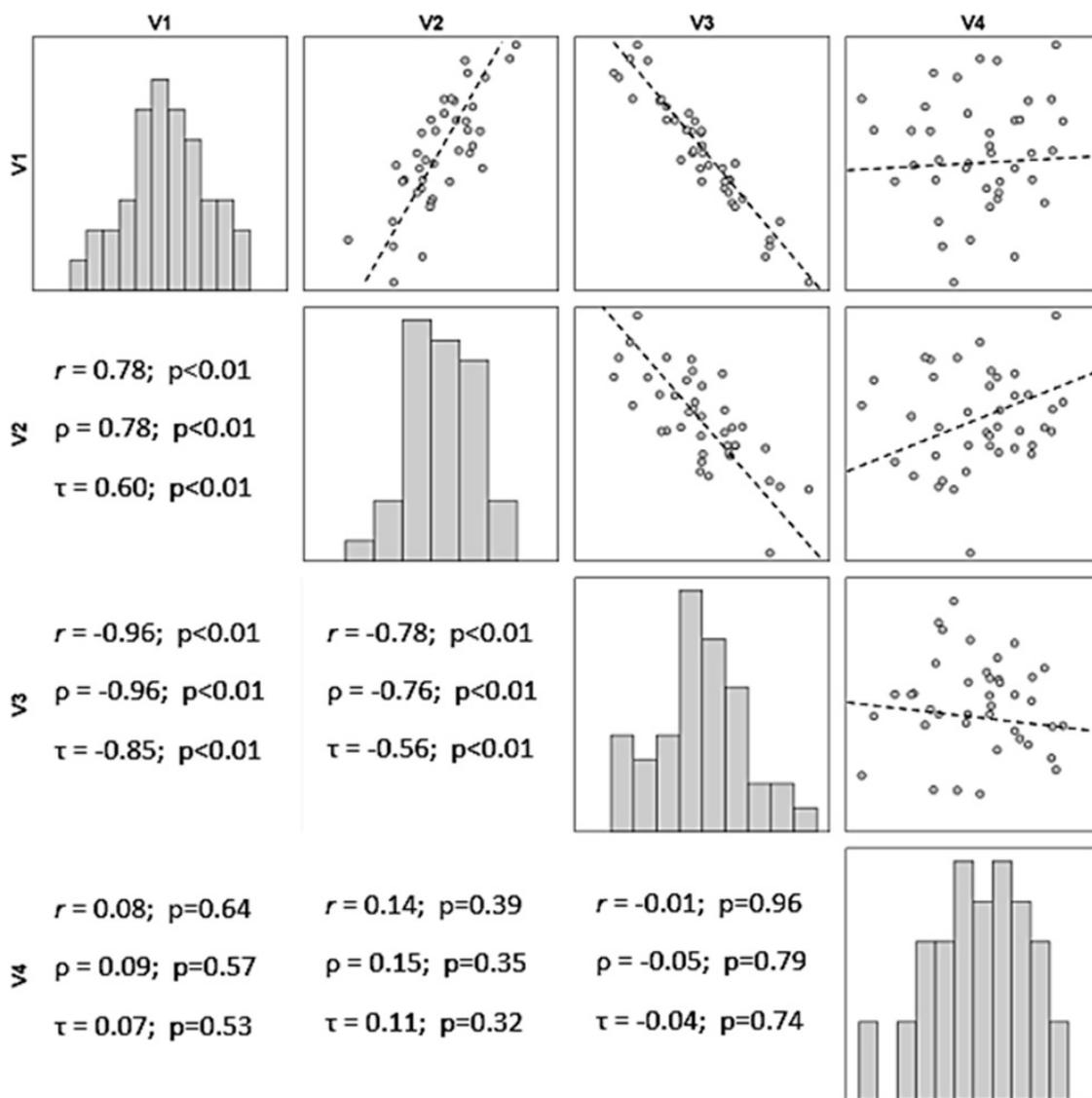
	V1	V2	V3	V4

Figure 1. Bar graphs, scatter plots, and correlation coefficients (*r*: Pearson, ρ: Spearman, and τ: Kendall *Tau*-b) for four hypothetical quantitative variables V1, V2, V3, and V4 (n = 40).

a gradual relationship in the same direction (rising or falling) for the whole domain of the data studied.

In Figure 2, it can be observed that there is no direct proportionality (linear) between the data for V1 and V5; rather there is an increase that is apparently exponential. Since the variation is monotonic (the data for V1 increase as a function of V5), the Spearman and Kendall coefficients can be used to estimate the correlation. In this example, to use the Pearson's coefficient, it will be necessary to log transform the data to achieve a certain linearity of correlation (Figure 2: V1 x V6). It should be noted that the ρ and τ coefficients give the same resultant values for the

correlations V1 vs. V5 and V1 vs. V6, since V6 is a transformation of V5 into monotonic data.

In biomedical sciences, the Spearman coefficient (ρ or rho) is the most widely used for evaluating the correlation between two quantitative variables, probably because it is similar to the Pearson method, once the data have been substituted for their ordered ranks. However, care should be taken when generalizing conclusions of interpretations of the correlation between the values of the ranks of data from the variables and the original data.

In contrast, the Kendall Tau-b coefficient (τ or t_b), has mathematical properties that make it more robust to extreme data (outliers), give it a greater capacity for

ρ=0.94; τ=0.84 (p<0.01)
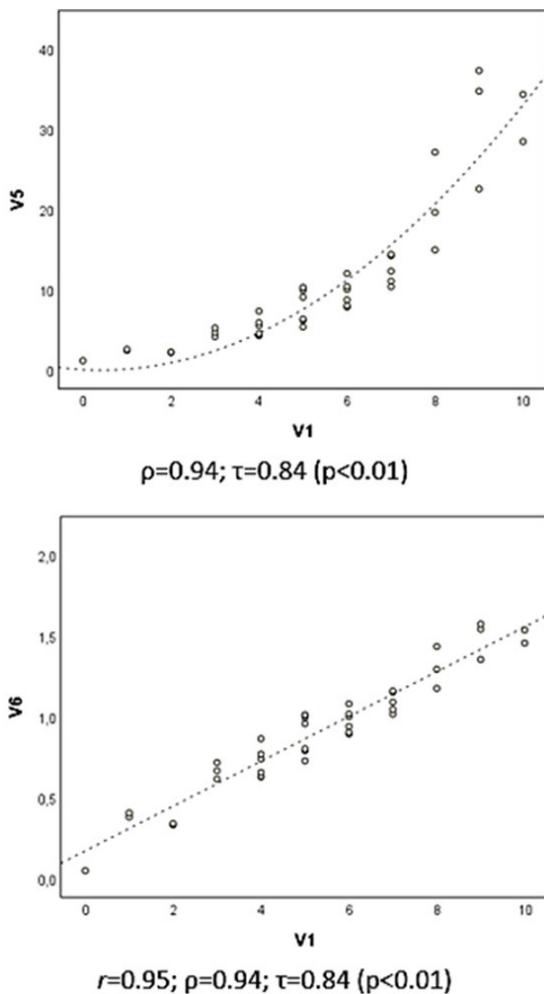

r=0.95; ρ=0.94; τ=0.84 (p<0.01)

Figure 2. Scatter plots and correlation coefficients for the hypothetical variables V1, V5 and V6 (n = 40), where V6 is the result of log transformation of V5 (V6 = $\log_{10}$ V5).

populational inference and a smaller estimation error. While significance (p-value) and direction (+ or -) are similar to those of the Spearman method, the coefficient returns less extreme values and interpretation is different, since it signifies the percentage of observed pairs that take the same direction in the sample (agreement) minus the pairs that do not agree. For example, a τ coefficient of 0.60 signifies that 80% of pairs agree, while 20% disagree (τ = 0.80 - 0.20 = 0.60).[9]

Transformation of data (for example, logarithmic, square root, 1/x) in order to obtain a normal distribution to enable Pearson's coefficient to be tested is a valid option for samples with asymmetrical data distributions (Figure 2: V1 x V6). However, it should be borne in mind that, in common with techniques that employ ordered ranks, transformation of data alters the scale

between measures and impacts on direct interpretation of the measures of effect.[7]

The magnitude of the effect of the correlation between two or more variables is represented by correlation coefficients, which take values from -1 to +1, passing through zero (absence of correlation). Positive coefficients ($r > 0$) indicate a direct relationship (Figure 1: V1 x V2) between variables; while negative coefficients ($r < 0$) indicate an inverse correlation (Figure 1: V1 x V3 and V2 x V3).

Each correlation test has its own coefficient, demanding its own interpretation. In general, for the coefficients Pearson's r and Spearman's ρ, values from 0 to 0.3 (or 0 to -0.3) are biologically negligible; those from 0.31 to 0.5 (or -0.31 to -0.5) are weak; from 0.51 to 0.7 (or -0.51 and -0.7) are moderate; from 0.71 to 0.9 (or -0.71 to 0.9) are strong correlations; and correlations > 0.9 (or < -0.9) are considered very strong.[8]

One peculiarity of Pearson's r coefficient is that the square of its value provides an estimate of the percentage of variability in the values of one variable that is explained by the variability in the other. For example, a coefficient of $r = 0.7$ indicates that 49% of the variability of one variable can be explained by, or is followed by, the variation in the values of the other, in the sample tested.

In clinical and biomedical studies, the majority of coefficients with biological significance fall in the range of 0.5 to 0.8 (or -0.5 to -0.8). This is the result of errors of measurement, laboratory techniques, or variation of instruments, which affect the precision of measurements, and also, and primarily, because biological phenomena are affected by multifactorial influences and complex interactions, in which the variation of a single variable cannot totally explain the behavior of another.[2]

Tests of the significance of a correlation between quantitative variables are based on the null hypothesis that there is no correlation between the variables ($r = 0$), which makes the p-value subject to influence both from the dimension of the effect and from the sample size. This means that caution is necessary when interpreting coefficients that result in a weak correlation (r < 0.3), but have highly significant p-values, caused by overly-large sample sizes. Calculation of sample sizes for analysis of correlations has been explored in an earlier edition of this periodical.[10]

Correlation coefficients have inferential properties and, in scientific texts, should preferably be expressed with their 95% confidence intervals and significance (p-value), for example: ρ = 0.76 (95%CI 0.61-0.91), p < 0.01.[11,12] In the case of multiple comparisons, coefficients can be shown on their own, in the form of a

matrix, and with their significance indicated to facilitate interpretation of the data, as Brianezi et al. presented the 28 correlations between seven histological parameters in a single table.[13] Special cases involving hundreds or thousands of correlations may demand graphical representation techniques, such as the color heatmaps often used in genome studies, just as Hsu et al. represented 4,930 correlations between (85x58) genomic and metabolomic variables.[14]

In the case of ordinal data with few categories (for example, satisfaction scores, quality of life items, socioeconomic status), investigations based on the polychoric test of correlation may be more robust (smaller type I errors) than using the Spearman and Kendall tests.[15] Although rarely used, there are also methods for assessing the correlations between variables of a categorical nature (for example, Cramér's V coefficient) and between dichotomous and quantitative variables (for example, the point-biserial correlation coefficient), but these approaches are beyond the scope of this text.[7]

In special situations in which linear correlations between different variables must be analyzed in conjunction (for example, questionnaire items) in order to understand the overall variation in conjunction of the variables, analysis of the correlation between "n" variables can be assessed using the consistency type Intraclass Correlation Coefficient (ICC). There are different ways to analyze ICCs, which result in indicators of different magnitudes.[16] An ICC (random or two-factor mean, mean measures) returns the same value as Cronbach's α coefficient, used to measure the internal consistency of scales.[17]

The identification of a significant correlation between two or more variables should be interpreted with caution, since statistical analysis does not provide evidence of direct dependence or even of causality between the variables, just that they tend to vary in conjunction.[1,18,19] However, despite the risk of fallacious conclusions of causality or on the basis of results of correlations between variables, correlation tests are important exploratory techniques for investigation of associations between the behavior of groups of variables, facilitating construction of hypothetical models that should then be confirmed by means of dedicated experiments. Indeed, this occurs with ecological clinical studies that often employ correlation techniques for data analysis and provide a basis for subsequent investigations of the phenomena indicated by correlations between indicators and population groups.[20-22] The same applied to genome-wide and proteomic studies, considered exploratory, which study the patterns of correlations of findings with clinical variables in order to indicate models for later confirmation.[19,23]

Indeed, performing multiple correlation tests on a sequence of variables increases the chances of identifying, by chance, correlations described as "spurious", which should be evaluated in terms of their biological plausibility and confirmed later using appropriate investigation techniques. Use of techniques for correction of p-values adjusted for multiple correlations is always recommended in these conditions.[7,19,24-27]

Another limitation of an inferential nature in correlation analyses is rooted in their incapacity for extrapolation of conclusions to different data intervals or different populations from those studied.

Correlation analyses were not developed a priori with the purpose of predicting values or for inference of the participation of multiple variables in the explanation of a phenomenon and there are regression or multivariate analysis techniques that can be used for this purpose.[7] Although there are partial correlation techniques that adjust the correlation values for the behavior of confounding variables (identical to the standardized β coefficient in multivariate linear regression), and polynomial transformation techniques for correction of non-monotonic correlations, an experienced statistics professional should be consulted for planning and execution of analyses of greater complexity.

Correlation analyses can also be employed to compare parallelism of measures between two different scales for measurement of the same phenomenon, such as psychometric quality-of-life scales,[28] or clinimetric scales, such as pressure ulcer risk scales. However, researchers very often use them erroneously to test agreement of data or consecutive measures of the same phenomenon (for example, test-retest,[29] calibration of measurement instruments, interrater comparisons), even though there are more appropriate methods for these purposes.[30]

Finally, strategies for evaluation of correlations between variables should be encouraged in clinical and biomedical research, since they maximize understanding of the phenomena studied. However, due to the peculiarities inherent to the different methods, they must be described in detail in the methodology and in the presentation of results.

■ REFERENCES

1. Liu J, Tang W, Chen G, Lu Y, Feng C, Tu XM. Correlation and agreement: overview and clarification of competing concepts and measures. Shanghai Arch Psychiatry. 2016;28(2):115-20. PMid:27605869.

2. Zou KH, Tuncali K, Silverman SG. Correlation and simple linear regression. Radiol. 2003;227(3):617-22. http://dx.doi.org/10.1148/radiol.2273011499. PMid:12773666.

3. Scott M, Flaherty D, Currall J. Statistics: are we related? J Small Anim Pract. 2013;54(3):124-8. http://dx.doi.org/10.1111/jsap.12029. PMid:23458641.

4. Miot HA. Assessing normality of data in clinical and experimental trials. J Vasc Bras. 2017;16:88-91. http://dx.doi.org/10.1590/1677-5449.041117. PMid:29930631.

5. Rossi FH, Volpato MG, Metzger PB, et al. Relationships between severity of signs and symptoms and quality of life in patients with chronic venous disease. J Vasc Bras. 2015;14(1):22-8. http://dx.doi.org/10.1590/1677-5449.20140039.

6. Ohki AV, van Bellen B. Regional incidence of venous thromboembolism in Brazil. J Vasc Bras. 2017;16:227-31. http://dx.doi.org/10.1590/1677-5449.000517. PMid:29930651.

7. Norman GR, Streiner DL. Biostatistics. The bare essentials. 4th ed. Shelton: People's Medical Publishing House; 2014.

8. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71. PMid:23638278.

9. Arndt S, Turvey C, Andreasen NC. Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. J Psychiatr Res. 1999;33(2):97-104. http://dx.doi.org/10.1016/S0022-3956(98)90046-2. PMid:10221741.

10. Miot HA. Sample size in clinical and experimental trials. J Vasc Bras. 2011;10:275-8. http://dx.doi.org/10.1590/S1677-54492011000400001.

11. Kim N, Fischer AH, Dyring-Andersen B, Rosner B, Okoye GA. Research Techniques Made Simple: Choosing Appropriate Statistical Methods for Clinical Research. J Invest Dermatol. 2017;137(10):e173-8. http://dx.doi.org/10.1016/j.jid.2017.08.007. PMid:28941476.

12. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. Am J Physiol Endocrinol Metab. 2004;287(2):E189-91. http://dx.doi.org/10.1152/ajpendo.00213.2004. PMid:15271643.

13. Brianezi G, Minicucci EM, Marques ME, Miot HA. Evaluation epidermal p53 immunostaining by digital image analysis. Skin Res Technol. 2013;19(1):e108-12. http://dx.doi.org/10.1111/j.1600-0846.2012.00616.x. PMid:22672779.

14. Hsu HH, Araki M, Mochizuki M, et al. A Systematic Approach to Time-series Metabolite Profiling and RNA-seq Analysis of Chinese Hamster Ovary Cell Culture. Sci Rep. 2017;7(1):43518. http://dx.doi.org/10.1038/srep43518. PMid:28252038.

15. Jin S, Yang-Wallentin F. Asymptotic Robustness Study of the Polychoric Correlation Estimation. Psychometrika. 2017;82(1):67-85. http://dx.doi.org/10.1007/s11336-016-9512-2. PMid:27660261.

16. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420-8. http://dx.doi.org/10.1037/0033-2909.86.2.420. PMid:18839484.

17. Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's Alpha. Psychometrika. 2009;74(1):107-20. http://dx.doi.org/10.1007/s11336-008-9101-0. PMid:20037639.

18. Jupiter DC. How are we related? Causality, correlation, and association. J Foot Ankle Surg. 2012;51(5):702-3. http://dx.doi.org/10.1053/j.jfas.2012.05.025. PMid:22766190.

19. Ness RO, Sachs K, Vitek O. From correlation to causality: statistical approaches to learning regulatory relationships in large-scale biomolecular investigations. J Proteome Res. 2016;15(3):683-90. http://dx.doi.org/10.1021/acs.jproteome.5b00911. PMid:26731284.

20. Ojha RP, Offutt-Powell TN, Evans EL, Singh KP. Correlation coefficients in ecologic studies of environment and cancer. Arch Environ Occup Health. 2011;66(4):241-4. http://dx.doi.org/10.1080/19338244.2010.539641. PMid:22014198.

21. Schmitt JV, Chinem VP, Marques MEA, Miot HA. Increase in the incidence of basal cell carcinoma in a university hospital between 1999 and 2009. An Bras Dermatol. 2011;86(2):375-7. http://dx.doi.org/10.1590/S0365-05962011000200029. PMid:21603829.

22. Schmitt JV, Miot HA. Distribution of Brazilian dermatologists according to geographic location, population and HDI of municipalities: an ecological study. An Bras Dermatol. 2014;89(6):1013-5. http://dx.doi.org/10.1590/abd1806-4841.20143276. PMid:25387516.

23. Kumari S, Nie J, Chen HS, et al. Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. PLoS One. 2012;7(11):e50411. http://dx.doi.org/10.1371/journal.pone.0050411. PMid:23226279.

24. Boardman TJ, Moffitt DR. Graphical Monte Carlo type I error rates for multiple comparison procedures. Biometrics. 1971;27(3):738-44. http://dx.doi.org/10.2307/2528613. PMid:5116579.

25. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811-8. http://dx.doi.org/10.1002/sim.4780090710. PMid:2218183.

26. Miot HA, Brianezi G. Morphometric analysis of dermal collagen by color clusters segmentation. An Bras Dermatol. 2010;85(3):361-4. http://dx.doi.org/10.1590/S0365-05962010000300010. PMid:20676470.

27. Ludbrook J. Multiple comparison procedures updated. Clin Exp Pharmacol Physiol. 1998;25(12):1032-7. http://dx.doi.org/10.1111/j.1440-1681.1998.tb02179.x. PMid:9888002.

28. Maranzatto CF, Miot HA, Miot LD, Meneguin S. Psychometrican analysis and dimensional structure of the Brazilian version of melasma quality of life scale (MELASQoL-BP). An Bras Dermatol. 2016;91(4):422-8. http://dx.doi.org/10.1590/abd1806-4841.20165014. PMid:27579735.

29. Paula HR, Haddad A, Weiss MA, Dini GM, Ferreira LM. Translation, cultural adaptation, and validation of the American Skindex-29 quality of life index. An Bras Dermatol. 2014;89(4):600-7. http://dx.doi.org/10.1590/abd1806-4841.20142453. PMid:25054747.

30. Miot HA. Agreement analysis in clinical and experimental trials. J Vasc Bras. 2016;15:89-92. http://dx.doi.org/10.1590/1677-5449.004216. PMid:29930571.

Correspondence
Hélio Amante Miot
Universidade Estadual Paulista – UNESP
Av. Prof. Mário Rubens Guimarães Montenegro, s/n - Distrito de Rubião Junior
CEP 18618-687 - Botucatu (SP), Brasil
Tel.: +55 (14) 3882-4922
E-mail: heliomiot@gmail.com

Author information
HAM - Tenured professor, Departamento de Dermatologia e Radioterapia, Faculdade de Medicina de Botucatu, Universidade Estadual Paulista (UNESP).