Structural Equation Modelling as an alternative to Multiple Linear Regression Models

Miriane Lucindo Zucoloto*
Universidade de São Paulo – USP, Ribeirão Preto, São Paulo, Brazil
mirianezucoloto@usp.br

João Maroco
ISPA-IU- Instituto Universitário de Ciências Psicológicas, Sociais e da Vida, Lisbon, Portugal
jpmaroco@ispa.pt

Edson Zangiacomi Martinez
Universidade de São Paulo – USP, Ribeirão Preto, São Paulo, Brazil
edson@usp.br

Juliana Alvares Duarte Bonini Campos
Universidade Estadual Paulista - UNESP, Araraquara, São Paulo, Brazil
jucampos@fcfar.unesp.br

Abstract

The structural equation modelling (SEM) is an extension of generalized linear models that considers a causal relationship model between variables with measurement error. This technique has been applied in studies with multiple dependents and independents variables, using latent and manifest variables, with simultaneous relationship of dependence and in highly complex models, being the analyzed model based on a theoretical framework established a priori. Thus, this study was conducted in order to present the technique of structural equation modelling as an alternative to the more common analysis, such as Multiple Linear Regression. We compared two analysis (Strctural Equation Modelling and Multiple Linear Regression) in order to evaluate the better model to assess the contribution of the pain severity, age and gender on the Oral Health Impact Profile-OHIP (dependent variable) in a sample of 1,007 Brazilian dental patients. The Oral Health Impact Profile (OHIP-14) and the Multidimensional Pain Inventory (MPI) were used. The sociodemographic and clinical variables collected were gender and age. We firstly performed a Multiple Linear Regression (MLR), considering the mean scores of the OHIP and Pain severity (MPI) as manifest variables. The existence of outliers was assessed by the square Mahalanobis distance (D^2) and the normality (univariate and multivariate) of the variables was evaluated by shape measures (Skewness and Kurtosis). The effects were considered statistically significant when $p < 0.05$. In a second moment the data were included in a Structural Equation Model (SEM), considering the factor "OHIP" as the central construct. The fit of the model was first analysed by the evaluation of the goodness of fit indices ($\chi^2/df \leq 3.0$, CFI and GFI$\geq$0.90 and RMSEA<0.10) and the contribution of the independent variables was based on statistical significance of causal paths (β), estimated by the z test, considering a significance level of 5%. As result of the MLR analysis, only the gender was statistically significant and the total explained variance of the model was 2.0%. For the SEM analysis, we observed adequate fit of the model to the data ($\chi^2/df=1.59$; CFI=0.99; GFI=0.98; RMSEA=0.04) and the independent variables contributed 20.0% to the variability of the central construct. Thus, the SEM can be a realistic alternative to reflect the complexity and multidimensionality present in certain theoretical discussions, providing more accurate and reliable results than usual techniques, such as multiple linear regression.

Keywords: statistical analysis; structural equation modelling; multiple linear regression; latent variable.

1. Introduction

The Structural Equation Modelling (SEM) was first documented in the first decades of the 20th century with widespread use in social sciences and humanities. Currently, SEM is considered an important tool for data analysis in various fields of knowledge (Maroco, 2014).

SEM is an extension of generalized linear models that considers a causal relationship model between variables and a measurement error. Thus, it can be described as a combination of classical techniques of Factor Analysis - which defines a measurement model operationalising latent variables or constructs - and Linear Regression - establishing, in the structural model, the relationship between the different study variables (Maroco, 2014, Kline, 1998, Hair et al., 2005).

In general, SEM is characterized by their ability to specify, estimate and test hypothetical relationships among a group of latent or manifest variables (Kline, 1998).

Unlike the more usual techniques of analysis, in which the data are responsible for the deductions of the theories, SEM is used to test a set of dependency relationships through a model which operationalizes this theory, i.e., the structure of the structural equation model is based on a theoretical framework established a priori (Maroco, 2014, Hair et al., 2005)

Thus, this study intends to present the Structural Equation Modelling as an alternative to traditional analysis methods, such as Multiple Linear Regression, considering real data from a study on the oral impact profile.

2. Methods

Below are listed the alternatives, possibilities and assumptions of SEM as compared with traditional analysis techniques (Hair et al., 2005, Kline, 1998, Maroco, 2014).

<u>Why use SEM as an alternative to traditional methods of statistical analysis?</u>

✓ Simultaneous dependent relationship between variables in the model:

Unlike the Multiple Linear Regression, the structural equation modelling, besides allowing multiple independent and dependent variables, allows to address simultaneous dependence relations (the variable may be dependent on a relationship and independent in another), and this set of relationships is the basis of SEM. These relationships are represented by parameters that indicate the magnitude of the effect that certain variables (independent) exert on another (dependent) according to a set of theoretical hypotheses that determine the structure of the model.

✓ Use of latent variables in the model:

SEM allows that abstract concepts are included in the analysis through latent variables (not directly measurable), being able to estimate the errors of measurement and taking into account the reliability and the validity of the results in the estimation of the model parameters.

✓ Analysis of complex theoretical models:

The complexity of theoretical models in some areas has grown with the progress of scientific research. Some theoretical models have multiple manifest and latent variables, differences between groups, hierarchical effects, multiple interactions, being inappropriate the use of the classical models of analysis in this cases. SEM also allows evaluating the global fit and the significance of the parameters in a theoretical generalization, unifying several multivariate statistical methods in a single methodological framework.

Steps of the Structural Equation Model

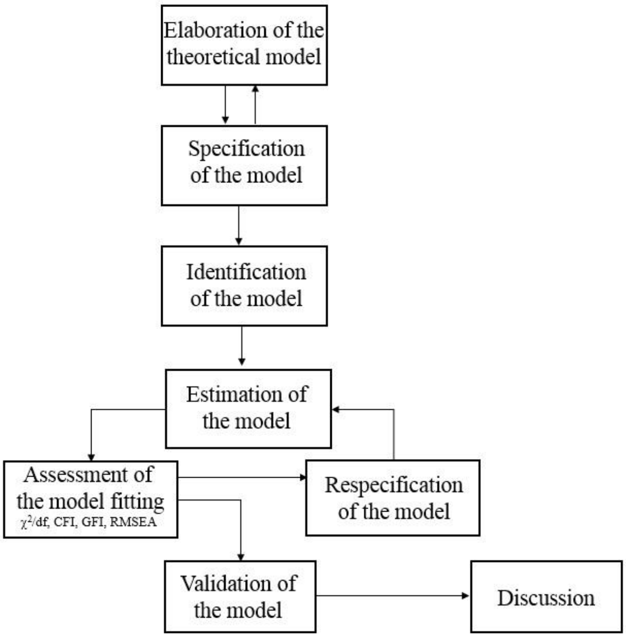

Figure 1. Steps of the Structural Equation Model.

Assumptions of the Structural Equation Model

- ✓ Independence of observations
- ✓ Multivariate normality
- ✓ Linearity
- ✓ Nonnull sample covariances
- ✓ Multiple indicators
- ✓ Absence of multicollinearity in the structural model
- ✓ strong measure
- ✓ Absence of outliers:

Application of techniques of Multiple Linear Regression and Structural Equation Modelling to assess the contribution of pain severity, age and gender to the Oral Health Impact Profile: a comparative study

Objective of the analysis: Evaluate the contribution of pain severity, age and gender on the oral health impact profile (OHIP) in a sample of 435 Brazilian dental patients.

Study variables:
Independent variables: Gender, age and pain severity.
Dependent variable: Oral Health Impact Profile (OHIP).

Instruments:

- ✓ Oral Health Impact Profile (OHIP-14) – Original factors: Functional limitation, Physical pain, Physical disability, Psychological discomfort, Psychological disability, Social disability and Handicap (Slade & Spencer, 1994, Slade, 1997, Oliveira & Nadanovsky, 2005). To define the structural model of this study, the factors functional limitation, psychological discomfort and social disability of the OHIP were chosen due to a greater contribution to the impact on oral health.

- ✓ Multidimensional Pain Inventory (MPI) – Part 1 (Silva & Ribeiro-Filho, 2006): Factor: Pain severity – items: 1, 2 and 3.

Sample characteristics: Patients enrolled at Faculty of Dentistry of Araraquara – São Paulo, Brazil. 74.7% were female and the mean age was 39.9 (SD=0.65) years.

Analysis 1: Multiple Linear Regression

Firstly, the significance of the effect of pain severity, age and gender on the Oral Health Impact Profile was assessed by a Multiple Linear Regression estimation of the parameters through the maximum likelihood method. The OHIP scores were obtained by average of the responses given to the 3 factors chosen of the OHIP (functional limitation, psychological discomfort and social disability), while the scores of pain severity was obtained by average of the responses given to the items 1, 2 and 3 of the Part 1 of the Multidimensional Pain Inventory. The existence of outliers was assessed by the square Mahalanobis distance (D^2) and the normality (univariate and multivariate) of the variables was evaluated by shape measures (Skewness and Kurtosis). The effects were considered statistically significant when $p < 0.05$. The analysis was performed in the software SPSS® Amos (v.21, SPSS Inc., Chicago, IL).

Analysis 2: Structural Equation Modelling

Validation Study:
Prior and required step to start analyze the structural model. The OHIP and MPI validations was conducted in previous studies (Zucoloto et al., 2014, Zucoloto, 2014).

Model Fit:
Estimation Method: Maximum Likelihood
Evaluation in two steps: i) Evaluation of the quality of the model fit: Indexes: ratio of chi-square by the degrees of freedom ($\chi 2/df \leq 2.0$), comparative fit index ($CFI \geq 0.9$), goodness of fit index ($GFI \geq 0.9$) and root mean square error of approximation ($RMSEA \leq 0.10$); ii) Path (β) significance: z test, ($p<.05$) (Maroco, 2014).
Software: SPSS® Amos (v.21, SPSS Inc., Chicago, IL).

3. Results

Analysis 1

No outlier was detected and no variable had severe deviations from the normal distribution (Sk=0.33-1.10; Ku=0.48-1.20). The model fitted for evaluation the contribution of pain severity, age and gender on OHIP presents explained variance of 2%. Only gender was statistically significant ($\beta=0.13$; $p=0.006$). Figure 1 presents the model with standardized estimates of regression coefficients and of the variability of the Oral health Impact Profile (OHIP).

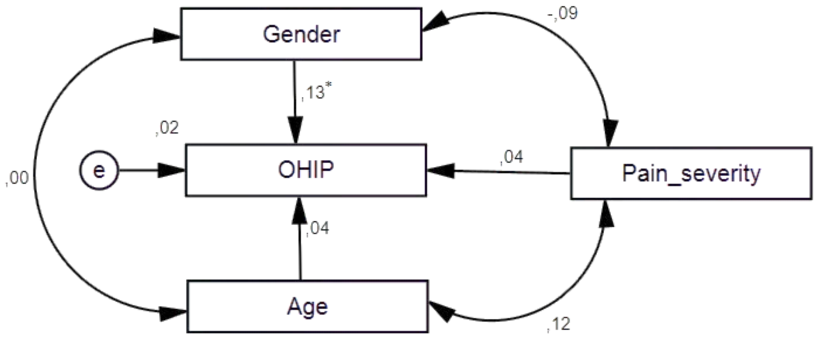

Figure 1. Multiple Linear Regression Model of the Oral Health Impact Profile (OHIP) in relation to the pain severity, gender and age. *Indicates statistical significance ($p<0.05$). Explained variance: 2.0%.

Analysis 2

Figure 2 shows the structural model with the estimated measurement weights between factors and the standardized contribution of each predictor variable (β) on Oral Health Impact Profile (OHIP).

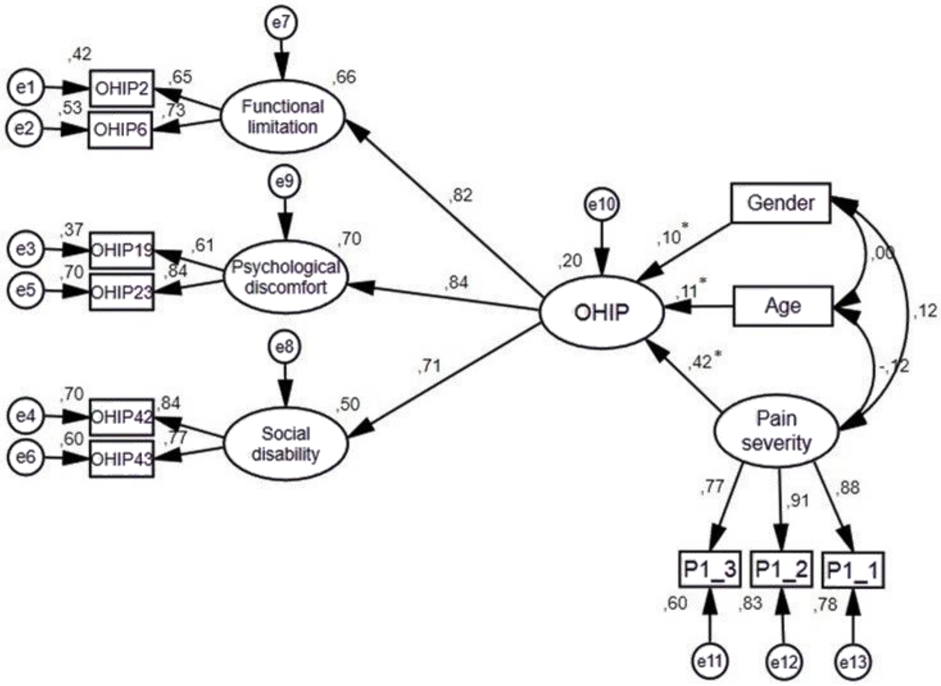

Figure 2. Structural model for assessment the contribution of the pain severity, age and gender on oral health impact profile (OHIP) (χ^2/df=1.59; CFI=0.99; GFI=0.98; RMSEA=0.04). *Indicates statistical significance ($p<0.05$). Explained variance: 20.0%.

It can be noted an adequate fit of the structural model to the data. Gender, age and pain severity had a positive and significant contribution on Oral Health Impact Profile. The higher the contribution of age and pain severity, the greater is the impact on oral health, being mainly detected in female gender.

4. Discussion

Despite the fact that Structural Equation Modelling techniques have emerged decades ago, its use in some areas, as in health science, is still incipient. Thus, the main objective of this work was to present the SEM as a realistic alternative to analyze complex models using latent and manifest variables, increasingly common in medical researches.

As mentioned, as the advantages of structural equation modeling over multiple linear regression can be summarized in the incorporation of latent and manifest variables simultaneously in the model and allows simultaneous dependency relationships between variables (Maroco, 2014). This can be observed in the analysis presented, in which the presence of latent variables in structural equation model greatly enhances the measure of the construct, improving the explained variance. According Hair et al. (2005) the use of latent variables have practical and theoretical justifications in improving statistical estimation, which better represents the theoretical concepts, seeking to extend the constructs which still have determinants little known and studied. Thus, the use of the SEM is advantageous because it provides estimates of strength for all hypothetical relationship in a theoretical scheme, bringing a more accurate measure and assuming that we cannot measure a construct perfectly and that the measurement error always exists.

As presented, the analysis of structural equations requires the performance of several steps and fulfillment of some assumptions, that becomes a complex method and difficult to implement and interpretation. However, with the advent of some software such as AMOS, LISREL and MPLUS this type of analysis has become increasingly accessible to professionals not only from exact sciences but also from humanities, social and health sciences.

Finally, in the predictive model proposed as example (Figure 1), we can observe an adequate fit of the structural model to the data and a significant relationship of the pain, age and gender on impacts of oral health. This relationship accounts for 20.0% of the variability of the central construct, which can be considered high, in view of the complexity and multidimensionality of the variables studied.

5. Conclusion

The Structural Equation Modelling can be a realistic alternative to reflect the complexity and multidimensionality present in certain theoretical discussions, providing more accurate and reliable results than usual techniques, such as multiple linear regression.

References

Hair, J.F., Black, W.C., Babin, B., Anderson, R.E. & Tatham, R.L. (2005). Multivariate data analysis, 6th ed.: Prentice Hall.

Kline, R.B. (1998). Principles and practice of structural equation modeling. New York: The Guilford Press.

Maroco, J. (2014). Análise de Equações Estruturais, 2nd ed. Lisboa: ReportNumber.

Oliveira, B. & Nadanovsky, P. (2005). Psychometric properties of the Brazilian version of the Oral Health Impact Profile - short form. Community Dent Oral Epidemiol, 33, 307-314.

Silva, J.A. & Ribeiro-Filho, N.P. (2006). Avaliação e Mensuração de Dor - Pesquisa, Teoria e Prática, 1st ed. Ribeirão Preto: Funpec.

Slade, G.D. (1997). Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol, 25, 284-290.

Slade, G.D. & Spencer, A.J. (1994). Development and evaluation of oral health impact profile. Community Dent Health, 11, 3-11.

Zucoloto, M.L. (2014). Contribuição dos impactos na saúde bucal e de variáveis sociodemográficas e clínicas na qualidade de vida relacionada à saúde. Brasil: Universidade Estadual Paulista - UNESP.

Zucoloto, M.L., Maroco, J. & Campos, J.A.D.B. (2014). Psychometric properties of the oral health impact profile and new methodological aproach. J Dent Res, 93, 645-650.