

Validation of general linear modeling for identifying factors associated with Quality of Life: A comparison with structural equation modeling*

Naoko Kumagai^{1,2#}, Motonori Hatta³, Yashiyasu Okuhara², Hideki Origasa⁴

¹Integrated Center for Advanced Medical Technologies, Kochi Medical School, Kochi, Japan;

[#]Corresponding Author: kumagain@kochi-u.ac.jp

²Center of Medical Information Science, Kochi Medical School, Kochi, Japan; okuharay@kochi-u.ac.jp

³Data Management, Actelion Pharmaceuticals Japan Ltd., Tokyo, Japan; motonori.hatta@actelion.com

⁴Division of Biostatistics and Clinical Epidemiology, School of Medicine, University of Toyama, Toyama, Japan; horigasa@las.u-toyama.ac.jp

Received 3 September 2013; revised 8 October 2013; accepted 21 October 2013

Copyright © 2013 Naoko Kumagai *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Purpose: General linear modeling (GLM) is usually applied to investigate factors associated with the domains of Quality of Life (QOL). A summation score in a specific sub-domain is regressed by a statistical model including factors that are associated with the sub-domain. However, using the summation score ignores the influence of individual questions. Structural equation modeling (SEM) can account for the influence of each question's score by compositing a latent variable from each question of a sub-domain. The objective of this study is to determine whether a conventional approach such as GLM, with its use of the summation score, is valid from the standpoint of the SEM approach. **Method:** We used the Japanese version of the Mageri Foundation Respiratory Failure Questionnaire, a QOL measure, on 94 patients with heart failure. The daily activity sub-domain of the questionnaire was selected together with its four accompanying factors, namely, living together, occupation, gender, and the New York Heart Association's cardiac function scale (NYHA). The association level between individual factors and the daily activity sub-domain was estimated using SEM and GLM, respectively. The standard partial regression coefficients of GLM and standardized

path coefficients of SEM were compared. If these coefficients were similar (absolute value of the difference <0.05), we concluded that GLM was valid, as well as the SEM approach. **Results:** The estimates of living together were -0.06 and -0.07 for the GLM and SEM. Likewise, the estimates of occupation, gender, and NYHA were -0.18 and -0.20, -0.08 and -0.08, 0.51 and 0.54, respectively. The absolute values of the difference for each factor were 0.01, 0.02, 0.00, and 0.03, respectively. All differences were less than 0.05. This means that these two approaches lead to similar conclusions. **Conclusion:** GLM is a valid method for exploring association factors with a domain in QOL.

Keywords: General Liner Modeling; Latent Variable; Standardized Path Coefficient; Standard Partial Regression Coefficient; Structural Equation Modeling

1. INTRODUCTION

In medical treatment, QOL has been defined as a personal sense of well-being and a multidimensional factor that generally includes physical, psychological, social, and spiritual dimensions or domains [1]. The distinctive feature of the research objectives of QOL is that the focus is typically on broad questions [2]. These questions are made up of multiple scales, such as the binary scale, with "yes or no" questions, graded scales including options such as, "very bad," "bad," "average," "good," and "very good"; as well as continuous scales such as the

*Competing interests: The authors declare that they have no competing interests.

Authors' contributions: All authors have contributed substantially to the analysis of the data and preparation of the manuscript. All authors also read and approved of the final manuscript.

Visual Analogue Scale (VAS).

For a variety of QOL questionnaires, the general linear model, such as analysis of variance, is typically used to identify factors that are associated with a certain domain of QOL. Examples of these include research on the identification of a domain and related factors among HIV-positive individuals, as well as correlation studies on asymptomatic vertebral fractures and quality of life [3,4]. However, general liner modeling (GLM) uses the summative score obtained from scores on each question in a given sub-domain. This is because GLM cannot be used with multiple response variables. However, using the summation score ignores the influence of individual questions. In contrast, structural equation modeling (SEM) can deal with multiple responses and accounts for the influence of each question's score by compositing a latent variable from each question of a domain. The objective of this study is to determine the validity of a conventional approach involving the use of the summation score and GLM, as compared to the SEM approach.

2. METHODS

2.1. Materials and Subjects

2.1.1. Materials

The Japanese version of the Mageri Foundation Respiratory Failure (MRF-28) Questionnaire is a 28-item, disease-specific, health-related QOL questionnaire for patients with chronic respiratory failure due to pulmonary diseases. The questionnaire is self-administered and easy to complete, with all items requiring either a "yes" or "no" answer [5]. It consists of four domains, namely, daily activity, cognitive function, invalidity, "other," and two general questions about the patient's health status [5].

2.1.2. Subjects

The sample included in-patients and out-patients with symptomatic and previous, asymptomatic heart failure at the University of Toyama Hospital in Japan. Participants were recruited between December 2005 and November 2006. The study was approved by the Ethics Committee at the University of Toyama; all the participants provided written, informed consent to take part [5]. We used this database. A total of 94 subjects enrolled for this study.

2.2. Independent Variables and Response Variables

For this study, we used one of four domains of the MRF-28 questionnaire as a response variable, namely, the daily activity domain (See **Table 1**). In addition, we used four factors as independent variables, namely, living together (cohabitation status), occupation, gender, and the New York Heart Association's cardiac function scale (NYHA). The associations between the daily activity domain and the four factors were estimated using GLM and SEM. The daily activity domain consists of 11 questions that require a "yes" or "no" answer. "Yes" was assigned a score of 1, while "no" was assigned a score of 0. More "yes" answers indicated a greater burden from daily activity. A summation score was obtained from adding the scores on all 11 questions. With regard to living together, individuals staying with someone obtained a score of 1, while those living alone obtained 0. Currently employed individuals obtained a score of 1, while the unemployed obtained 0. Males were assigned a score of 1, while females were assigned a score of 0. Scores on the NYHA were divided into two groups; Class 2 was assigned a score of 0, while Class 3 and 4 were each assigned a score of 1. These were shown in **Table 2** as Patient Characteristics.

Table 1. The 11 items in daily activity domain—Mageri Foundation Respiratory Failure Questionnaire (MRF-28).

MRF1.	Washing myself (face) usually makes me feel breathless
MRF2.	Combing my hair or shaving myself usually makes me feel breathless
MRF3.	Getting dressed usually makes me feel breathless
MRF4.	Because of my disease, I am unable to shower as I would like to
MRF5.	Because of my disease, I cannot put on my socks, stockings or shoes as I would like to
MRF6.	Because of my disease, I am not able to cook as I would like to
MRF7.	Because of my disease, I cannot do housework or light repairs around the house
MRF8.	Because of my disease, when I need to, I cannot bend over as I would like to
MRF9.	Because of my disease, when I need to, I cannot pick up light things as I would like to
MRF10.	Because of my disease, I cannot play with children as I would like to
MRF11.	Because of my disease, I cannot talk as much as I would like to

2.3. Statistical Analysis

2.3.1. GLM

GLM was a special case of SEM and could be expressed as **Figure 1**. The summation score was regressed by a model that included four factors, namely, living together, occupation, gender, and NYHA. Standard partial regression coefficients were estimated.

2.3.2. SEM

We plotted a path from the latent variable to each question and made a latent variable of daily activity (See

Table 2. Patient characteristics.

Factor	Number of patients (%)
Living together (yes/no)	84 (89)/10 (11)
Occupation (yes/no)	25 (27)/69 (73)
Gender (male/female)	64 (68)/30 (32)
NYHA (class 2/class 3 and 4)	30 (32)/64 (68)

Figure 2); the association with a latent variable was estimated on the basis of Kendall's correlation. The goodness-of-fit of the SEM was evaluated using the Standardized Root Mean Square Residual (SRMR) (good

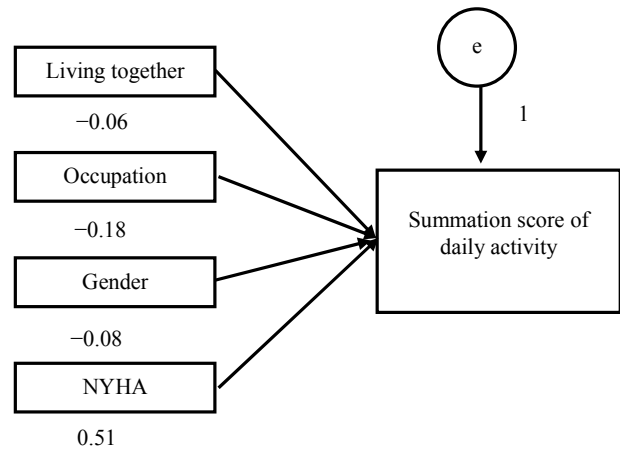


Figure 1. The association between summation score of daily activity and the four factors.

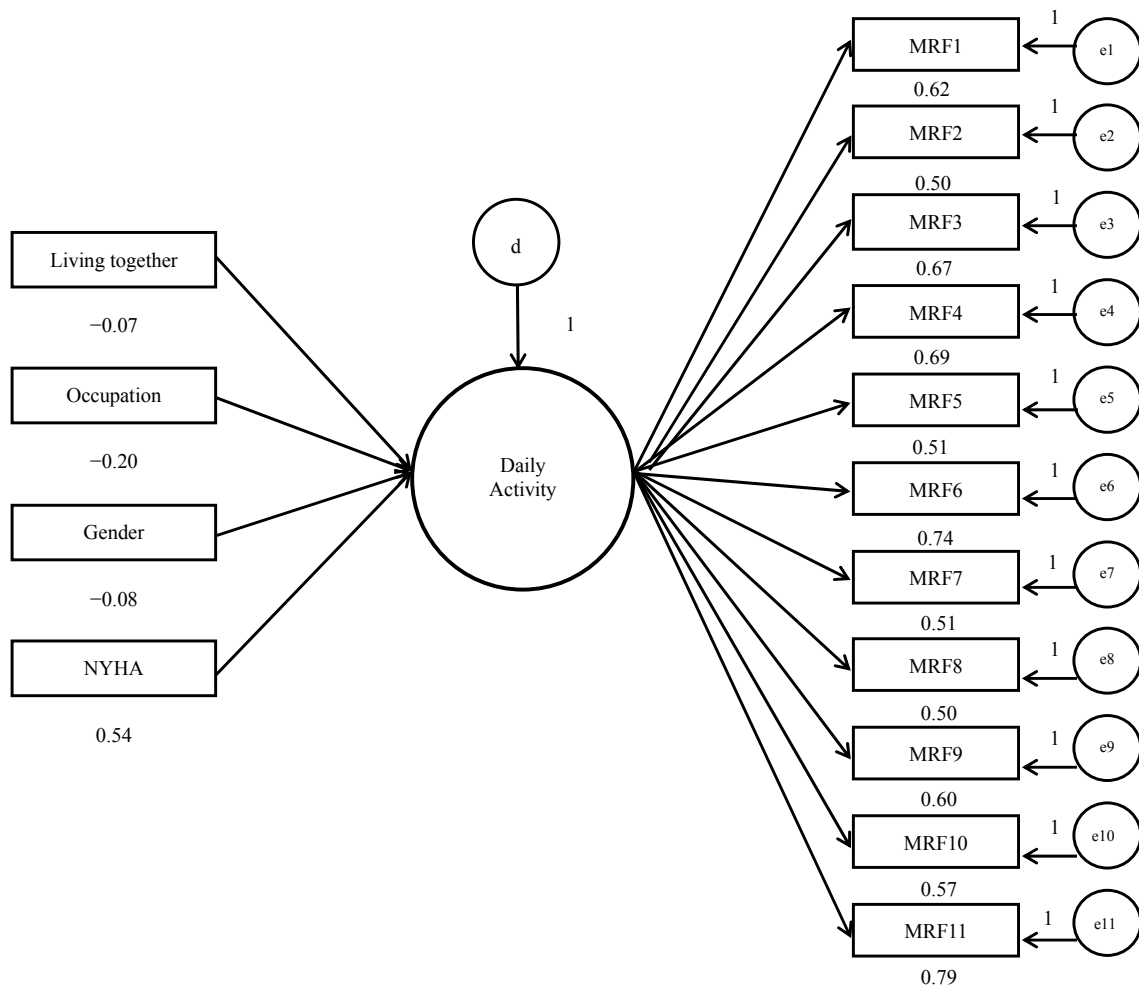


Figure 2. The association between daily activity and the four factors.

models <0.08), Goodness of Fit Index (GFI) (good models >0.95), and Normed Fit Index (NFI) (good models > 0.90) [6].

Then, we examined the extent of the difference between the standardized path coefficient from a factor to the latent variable and the standard partial regression coefficient of GLM. If the absolute value of the difference is small (<0.05 of scale difference for scale of 0 - 1), that is, less than 0.05, then the assumption is that GLM is suitable, as well as SEM. We utilized the Statistical Analysis system (SAS Institute, Cary, NC, USA).

3. RESULTS

The structural equation model presented in **Figure 2** (SRMSR = 0.078; GFI = 0.96; NFI = 0.93) depicted acceptable fits. The standard partial regression coefficient and standardized path coefficient of GLM, and of SEM for each factor—living together, occupation, gender, and NYHA class—were, respectively, as follows: -0.06 to -0.07; -0.18 to -0.20, -0.08 to -0.08, 0.51 to 0.54, as shown in **Table 3**. The absolute values of the differences were 0.01, 0.02, 0.00, and 0.03, respectively. All were less than 0.05. Both approaches showed similar estimates; in addition, the positive and negative signs were the same. As scores on the NYHA increased, alluding to severity of cardiac dysfunction, so did the burden of daily activities. Further, unemployed individuals also experienced more of this burden than those in occupation. This is most likely due to the fact that people who are not in employment often have disabilities that, to some extent, interfere with daily activities. People without cohabitants felt more burdened than those who were cohabiting, probably due to lack of assistance. In terms of gender, women tended to feel more burdened than men.

4. DISCUSSION

In this study, we used real quantitative data in order to assess whether GLM is appropriate for the identification of associated factors within QOL domains, as compared to SEM. The association between factors in the daily

Table 3. Comparison between the general liner model (GLM) and structural equation model (SEM).

Method	GLM	SEM
	Standard partial regression coefficient	Standardized path coefficient
Living together	-0.06	-0.07
Occupation	-0.18	-0.20
Gender	-0.08	-0.08
NYHA	0.51	0.54

activity domain was similar for both modeling approaches.

The general linear model was expressed as: $Y_{\text{summation}} = \beta_1 X_{1\text{living together}} + \beta_2 X_{2\text{occupation}} + \beta_3 X_{3\text{gender}} + \beta_4 X_{4\text{NYHA}} + e$ (Equation (1)). The structural equation model was given as follows:

$$Y_{\text{MRF1}} = \gamma_1 F + e_1$$

$$Y_{\text{MRF2}} = \gamma_2 F + e_2$$

$$Y_{\text{MRF3}} = \gamma_3 F + e_3$$

$$Y_{\text{MRF4}} = \gamma_4 F + e_4$$

$$Y_{\text{MRF5}} = \gamma_5 F + e_5$$

$$Y_{\text{MRF6}} = \gamma_6 F + e_6$$

$$Y_{\text{MRF7}} = \gamma_7 F + e_7$$

$$Y_{\text{MRF8}} = \gamma_8 F + e_8$$

$$Y_{\text{MRF9}} = \gamma_9 F + e_9$$

$$Y_{\text{MRF10}} = \gamma_{10} F + e_{10}$$

$$Y_{\text{MRF11}} = \gamma_{11} F + e_{11}$$

$$F = \alpha_1 X_{1\text{living together}} + \alpha_2 X_{2\text{occupation}} + \alpha_3 X_{3\text{gender}} + \alpha_4 X_{4\text{NYHA}} + d \text{ (Equation (2))}$$

The different results were due to the summation or latent variable of F, which consists of a correlation between each question and its error. Therefore, we assume that if a question of a given domain strongly correlated with the domain, and there was a homogenous association between the factors and the domain, then GLM and SEM would estimate similar results. Cronbach's alpha for questions of daily activity was 0.9. This is considered considerably high. The questions of the sub-domain were closely related as a group. Under the well-constructed QOL sub-domain, the association between the factors was similarly estimated using the GLM and SEM approaches.

As a limitation, at least 100 cases, although 200 are preferable, are required for SEM. Our study had 94 cases, which is considerably smaller than the required number of cases. However, goodness-of-fit was appropriate, which means that the small sample size may not have a major influence [7]. As other limitations, it may need simulation study which examined among various values of Cronbach's alpha to confirm our conclusions. Therefore, further more studies would be needed.

5. CONCLUSIONS

The QOL sub-domain is generally constructed with a high Cronbach's alpha (>0.9) [8].

Although the high Cronbach's alpha may not be directly related to validity of GLM, we assume that well-constructed sub-domains result in GLM and SEM that yield suitable results.

6. ACKNOWLEDGEMENTS

The authors would like to thank all the patients who participated so

willingly in the study.

REFERENCES

- [1] Ferrell, B.R. and Hassey, D.K. (1997) Quality of life among long-term cancer survivors. *Oncology*, **11**, 565-571.
- [2] Fayers, P.M. and Machin, D. (2005) Quality of life: The assessment, analysis and interpretation of patient-reported outcomes, Japanese version. Nakayama Shoten Co., Ltd., Tokyo.
- [3] Shan, D., Ge, Z., Ming, S., Wang, L., Sante, M., *et al.* (2011) Quality of life and related factors among HIV-positive spouses from serodiscordant couples under anti-retroviral therapy in Henan Province, China. *PLoS One*, **6**, e21839. <http://dx.doi.org/10.1371/journal.pone.0021839>
- [4] Lopes, J.B., Fung, L.K., Cha, C.C., Gabriel, G.M., Takayama, L., Figueiredo, C.P. and Pereira, R.M. (2012) The impact of asymptomatic vertebral fractures on quality of life in older community-dwelling women: The São Paulo Ageing & Health Study. *Clinics (Sao Paulo)*, **67**, 1401-1406. [http://dx.doi.org/10.6061/clinics/2012\(12\)09](http://dx.doi.org/10.6061/clinics/2012(12)09)
- [5] Hatta, M., Joho, S., Inoue, H. and Origasa, H. (2009) A health-related quality of life questionnaire in symptomatic patients with heart failure: Validity and reliability of a Japanese version of the MRF28. *Journal of Cardiology*, **53**, 117-126. <http://dx.doi.org/10.1016/j.jjcc.2008.09.011>
- [6] Hu, L. and Bentler, P.M. (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, **6**, 1-55. <http://dx.doi.org/10.1080/10705519909540118>
- [7] Hoelter, D.R. (1983) The analysis of covariance structures: Goodness-of-fit indices, sociological. *Methods and Research*, **11**, 325-344. <http://dx.doi.org/10.1177/0049124183011003003>
- [8] Altman, D.G. and Bland J.M. (1997) Cronbach's alpha. *British Medical Journal*, **314**, 572. <http://dx.doi.org/10.1136/bmj.314.7080.572>