

# Analysis of Quality of Life in Cancer Patients by Structural Equation Model

Hengqing Tong, Shudan Lu, Yang Ye, Yichao Pan

Department of Mathematics, Wuhan University of Technology, Wuhan, China.  
Email: tonghengqing@126.com

Received April 14<sup>th</sup>, 2010; revised May 8<sup>th</sup>, 2010; accepted May 27<sup>th</sup>, 2010.

## ABSTRACT

*Many people have been dead of cancer. The life quality of patients with cancer has aroused great concern from the public and specialists. In this paper, an index system of life quality is proposed to evaluate the quality of life, which includes 6 first-level indexes and 34 second-level indexes. Then, a structural equation model (SEM) based on these indexes and relationships among them is constructed for the analysis of quality of life in cancer patients. Furthermore, we offer a definite linear algorithm for the calculation of SEM. This method is more objective and scientific compared with traditional methods, such as descriptive analysis, some simple test methods and so on.*

**Keywords:** *Quality of Life, Structural Equation Model, Unit Vector Constraint, Definite Linear Algorithm*

## 1. Introduction

Nowadays, cancer has become the main threat to human health. For these diseases, it is hard to evaluate the treatment effect by cure rate, also survival rate is very limited. Compared with the traditional survival analysis method only considering survival rate or survival time, QOL evaluation focus more on subjective feeling, and can reflect the patient's health more fully. Considering that, the conception of QOL was accepted by the medical community in the 1970s, and there was ever a great upsurge in studies. QOL research of Clinical Oncology can go back to Karnofsky Performance Status (KPS) proposed by Karnofsky in the 1940s, which does not include the patient's subjective feelings, therefore, it is not a real QOL evaluation. The modern QOL research was initiated by Priestman and others with QOL measurement of breast cancer patients in 1976. QOL is an index system comprehensively evaluating individual physical, psychological, social support and demographic level.

QOL has been widely applied in various fields, and become an indispensable social important index and evaluation tool. The research of QOL has mainly three great effects in the treatment of cancer. 1) The evaluation of the effect of cancer treatment performs the selective therapy. 2) It has advantage of the selection and assessment of anti-cancer drug, antiemetic and anodyne and so on. U.S. FDA has decided that QOL must be the index in all anti-cancer drug evaluation, there should have the material both improving survival time and quality of life,

when the efficacy of new medicine is not better than the old, new role can be found by QOL evaluation, which is good for drugs listed, it also helps to find new unexpected side effects or toxicity. 3) It helps to comprehend the long-term existing status of therapy in patients with cancer and assists to adopt the best management countermeasure in patients with late cancer. Now the basic principle of cancer radical treatment is the comprehensive treatment, the ultimate goal of which is to extend lifetime and improve the quality of life. But for most of the definite diagnosis of cancer patients, they already lost radical therapy and need palliative treatment, then improving QOL becomes the primary consideration. Therefore, study of QOL is of great significance in theoretically and practically, which needs further research.

At present, the evaluation of QOL is a very weak link, it is not maturity to measure QOL. Main systematic evaluation methods of Quality of Life, which are in common use are descriptive analysis method and some simple test methods, including Correlation Analysis and T test and so on. The assessment is seriously hindered by the complexity of QOL. In this paper, we build a related structural equation model (SEM) to reveal the potential factor of QOL, and the relationships between them, in which the summarizing coefficients are calculated by samples, so it is more objective and convincing, and could offer more deep analysis for the index systems. It will provide the suitable intervention methods in order to improve and enhance the quality of life in patients with cancer.

### 2. SEM for the QOL of the Cancer

SEM is a fast-growing embranchment of Application Statistics, widely used in psychology, sociology and other fields. This model not only studies the interior relationship among various factors, but also the relative and causal relations among latent variables. The QOL of the patients is controlled by psychological factor and physiological factor, whose values cannot be measured directly, so it is not suitable to use the usual statistical methods to analyze the influencing factors of the QOL. Considering that, this paper is the application of SEM to analyze the influencing factors of the QOL in patients with cancer.

There are always two systems of equations in a SEM. One is a structure system of equations among structural variables, and the other one is a measurement system of equations between structural variables and observed variables. Since the QOL of the patients with cancer cannot only reflect the curative effect, but also prevention and rehabilitation effects, it's important and urgent to quantitative analysis and build effective QOL index immediately. In our model, there are 6 latent variables and 34 observed variables. The variables are listed in **Table 1** as follows:

There are 6 structural variables ( $\xi_1, \eta_1 \sim \eta_5$ ) and 15 path relationships, which are expressed in **Figure 1** below (The path coefficients from the exogenous latent variable  $\xi_1$  to the endogenous latent variables  $\eta_1 \sim \eta_5$  are  $\gamma_1 \sim \gamma_5$ , expressed with dashed arrowheads; The relationships among independent variables are  $\beta_{ij}$ , expressed with real-line arrowheads).

The structural Equations are relationships among the latent variables. The structural equations can be expressed as follows:

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta_{21} & 0 & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & 0 & 0 \\ \beta_{41} & \beta_{42} & \beta_{43} & 0 & 0 \\ \beta_{51} & \beta_{52} & \beta_{53} & \beta_{54} & 0 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{pmatrix} + \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{14} \\ \gamma_{15} \end{pmatrix} \xi_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix} \tag{1}$$

In general, vector and matrix are used to describe the structural Equations. Let  $\xi' = (\xi'_1, \dots, \xi'_k)$  and  $\eta' = (\eta'_1, \dots, \eta'_m)$ . Then  $m \times m$  square matrix  $B$  is the coefficient matrix of  $\eta$ , then  $m \times k$  matrix  $\Gamma$  is the coefficient matrix of  $\xi$ ,  $\varepsilon'_\eta = (\varepsilon'_1, \dots, \varepsilon'_m)$  is the residual

vector, then the structural Equations (1) may be extended as:

$$\eta = B\eta + \Gamma\xi + \varepsilon_\eta \tag{2}$$

The structural variables are implicit and cannot be observed directly. Each structural variable is corresponding with many observed variables.

Suppose there are  $k$  independent structural variables and  $m$  dependent structural variables. The observed variables corresponding to the independent structural variable  $\xi_t$  are denoted as  $x_{tj}$ ,  $t=1, \dots, k$ ,  $j=1, \dots, K(t)$ , where  $K(t)$  is the number of observed variables corresponding to the independent structural variable  $\xi_t$ . In **Figure 1**,  $k=1$  and  $K(1)=6$ . The observed variables corresponding to the dependent variable  $\eta_i$  are denoted as  $y_{ij}$ ,  $i=1, \dots, m$ ,  $j=1, \dots, L(i)$ , where  $L(i)$  is the number of observed variables corresponding to the dependent structural variable  $\eta_i$ . In **Figure 1**,  $m=5$  and  $L(i)=6, 6, 6, 4, 6$ .

The observation equations can be expressed as the relationship from the observed variables to the structural variables:

$$\xi_t = \sum_{j=1}^{K(t)} \psi_{tj} x_{tj} + \varepsilon_{xt}, \quad t=1, \dots, k \tag{3}$$

$$\eta_i = \sum_{j=1}^{L(i)} \omega_{ij} y_{ij} + \varepsilon_{yi}, \quad i=1, \dots, m \tag{4}$$

where  $\psi_{tj}$ ,  $\omega_{ij}$  are the summarizing coefficients, and  $\varepsilon$  with subscript is a random error.

The relationships between the structural variables and the observed variables can also be expressed as follows:

$$\begin{pmatrix} x_{t1} \\ \vdots \\ x_{tK(t)} \end{pmatrix} = \begin{pmatrix} v_{t1} \\ \vdots \\ v_{tK(t)} \end{pmatrix} \xi_t + \begin{pmatrix} \varepsilon_{xt1} \\ \vdots \\ \varepsilon_{xtK(t)} \end{pmatrix}, \quad t=1, \dots, k \tag{5}$$

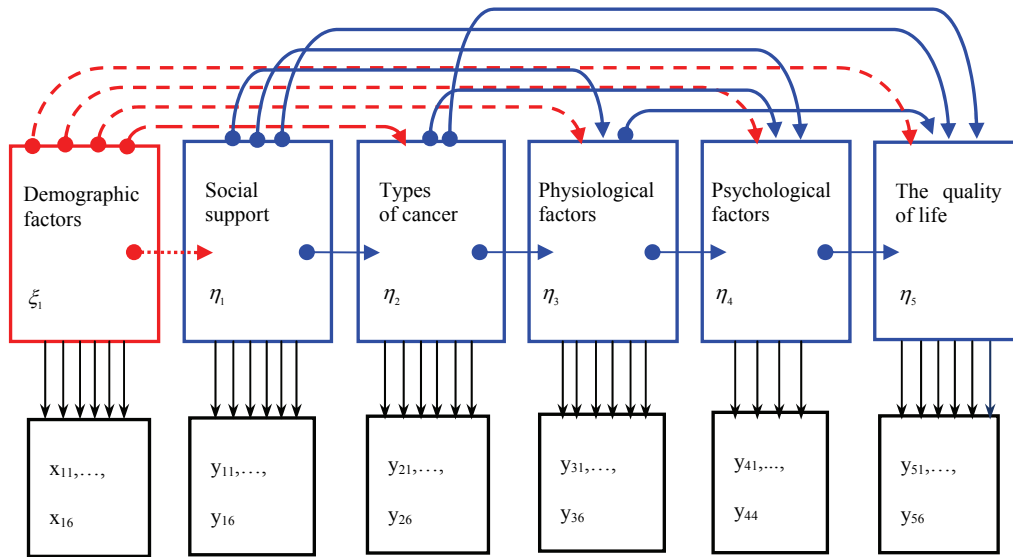
$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{iL(i)} \end{pmatrix} = \begin{pmatrix} \lambda_{i1} \\ \vdots \\ \lambda_{iL(i)} \end{pmatrix} \eta_i + \begin{pmatrix} \varepsilon_{yi1} \\ \vdots \\ \varepsilon_{yiL(i)} \end{pmatrix}, \quad i=1, \dots, m \tag{6}$$

where  $v_{tj}$ ,  $\lambda_{ij}$  are the carrying coefficients, and  $\varepsilon$  with subscript is still a random error.

Denoting observation vectors as  $x'_t = (x'_{t1}, \dots, x'_{tK(t)})$ ,  $y'_i = (y'_{i1}, \dots, y'_{iL(i)})$ , and denoting coefficients as

**Table 1. Index of variables**

Structural variables	Observed variables		
Demographic factors $\xi_1$	Age $x_{11}$	Sex $x_{12}$	Character $x_{13}$
	Level of education $x_{14}$	Occupation $x_{15}$	Marital status $x_{16}$
Social support $\eta_1$	Interpersonal relationship $y_{11}$	Family ties $y_{12}$	cognitive structure $y_{13}$
	Nursing $y_{14}$	Economic capability $y_{15}$	Medical facilities and service $y_{16}$
Types of cancer $\eta_2$	Cancer sites $y_{21}$	Cancer incidence rate $y_{22}$	Cure rate $y_{23}$
	Cancer benign or malignant $y_{24}$	Cancer period $y_{25}$	Type of medical care system $y_{26}$
Physiological factors $\eta_3$	The degree of physical pain $y_{31}$	Fatigue $y_{32}$	Anorexia $y_{33}$
	Health condition before the cancer $y_{34}$	Natural immunity $y_{35}$	Nutritional state $y_{36}$
Psychological factors $\eta_4$	Personal knowledge of cancer $y_{41}$	Self assessment $y_{42}$	Emotional stability $y_{43}$
	Psychological therapy $y_{44}$		
The quality of life $\eta_5$	Emotions and psychology $y_{51}$	Functional status $y_{52}$	Satisfaction rate with the treatment $y_{53}$
	Hope for the future $y_{54}$	Survival time $y_{55}$	Life satisfaction $y_{56}$



**Figure 1. Model of QOL index**

$\psi'_i = (\psi'_{i1}, \dots, \psi'_{iK(i)})$ ,  $\omega'_i = (\omega'_{i1}, \dots, \omega'_{iL(i)})$ , then the Equations (2)-(4) can be written into:

$$SEM^+ \begin{cases} \eta = B\eta + \Gamma\xi + \varepsilon_\eta \\ \xi_t = \psi'_t x_t + \varepsilon_{x_t}, t = 1, \dots, k \\ \eta_i = \omega'_i y_i + \varepsilon_{y_i}, i = 1, \dots, m \end{cases} \quad (7)$$

We call  $SEM^+$  the structural equation model with positive observation.

Letting  $\nu'_i = (\nu'_{i1}, \dots, \nu'_{iK(i)})$ ,  $\lambda'_i = (\lambda'_{i1}, \dots, \lambda'_{iL(i)})$ , then the observation Equations (5) and (6) can be expressed as:

$$x_t = \nu'_t \xi_t + \varepsilon_{x_t}, t = 1, \dots, k \quad (8)$$

$$y_i = \lambda'_i \eta_i + \varepsilon_{y_i}, i = 1, \dots, m \quad (9)$$

Combine the Equations (2), (8) and (9) as:

$$SEM^- \begin{cases} \eta = B\eta + \Gamma\xi + \varepsilon_\eta \\ x_t = \nu_t \xi_t + \varepsilon_{xt}, t = 1, \dots, k \\ y_i = \lambda_i \eta_i + \varepsilon_{yi}, i = 1, \dots, m \end{cases} \quad (10)$$

and call  $SEM^-$  the structural equation model with converse observation.

### 3. The Least Square Solution by the Modular Constraint of Structural Vector

By analyzing the observation equations of SEM carefully, we can discover the least square relationship between each structural variable and its observed variables, and obtain the least square solution of structural variable by the modular constraint of structural vector. Now, we give the MCLS algorithm of structural vector in SEM which is studied by penman as follows:

**Algorithm 1.** The modular constraint least square solution (MCLS) of structural vector in SEM.

**Step 1.** In  $SEM^-$ , suppose that  $\xi_t, \eta_i$  all are unit vectors, calculate the least square estimates of the coefficients between the structural variable and its observed variables:

$$\hat{\nu}_{tj}^2 = x_{tj}x'_{tj}, \quad j = 1, \dots, K(t), \quad t = 1, \dots, k \quad (11)$$

$$\hat{\lambda}_{ij}^2 = y_{ij}y'_{ij}, \quad j = 1, \dots, L(i), \quad i = 1, \dots, m \quad (12)$$

**Step 2.** In  $SEM^-$ , calculate the least square estimates of structural variable by making use of  $\hat{\nu}_{tj}, \hat{\lambda}_{ij}$ :

$$\hat{\xi}_{ts} = \frac{\hat{\nu}_t X'_{ts}}{\hat{\nu}_t \hat{\nu}'_t}, \quad \hat{\eta}_{is} = \frac{\hat{\lambda}_i Y'_{is}}{\hat{\lambda}_i \hat{\lambda}'_i}, \quad s = 1, \dots, N, \quad (13)$$

$$t = 1, \dots, k, \quad i = 1, \dots, m$$

Here  $X_{ts}, Y_{is}$  are the transverse vector of the observation data matrix,  $X'_{ts} = (x_{t1s}, \dots, x_{tK(t)s})$ ,  $Y'_{is} = (y_{i1s}, \dots, y_{iL(i)s})$ .

**Step 3.** In  $SEM^+$  (or (3) and (4)), make use of  $\hat{\xi}_t, \hat{\eta}_i$  obtained in Step 2 to calculate regression coefficients  $\psi_{tj}, \omega_{ij}$  according to a common linear regression method.

**Step 4.** In  $SEM^+$  (or (2)), make use of  $\hat{\xi}_t, \hat{\eta}_i$  obtained in Step 2 to calculate the estimates of coefficient matrices  $B, \Gamma$ .

Notice that (2) is a common linear regression Equation system, we can use Two Step Least Square to calculate it.

### 4. Definite Linear Algorithm of Prescription Constraint

Since the solutions of  $SEM^+$  or  $SEM^-$  are not unique, and they may differ by a multiple (the proof is omitted). Therefore, in the structural Equation (1) or (2), if each structural variable is multiplied by the same multiple, its coefficient solution is the same. Taking note of this, the solution of structural equations is irrelevant to the modular length of the structural variable. However, it is not reasonable to assume that the modular length of each structural variable is 1. On the other hand, if each modular length of the structural variable is not the same in the possibly existing optimal solution set, then MCLS is not good.

One reasonable way is to let each structural variable have an undetermined parameter of the modular length and combine the structural Equation (1) or (2) to find the solution. The error square sum of this solution includes  $m+k$  modular length parameters. Changing these modular length parameters to minimize the error square sum, we can obtain a reasonable modular length of the structural variable.

Another possible way is to find a more reasonable constraint to replace the modular constraint. After getting MCLS, we can change the modular length of the structural variable in observation equations to make the path coefficient between structural variables and observation variables satisfy the prescription condition. In Equations (3) and (4), the prescription conditions are:

$$\sum_{j=1}^{K(t)} \psi_{tj} = 1, \quad \psi_{tj} \geq 0, \quad t = 1, \dots, k \quad (14)$$

$$\sum_{j=1}^{L(i)} \omega_{ij} = 1, \quad \omega_{ij} \geq 0, \quad i = 1, \dots, m \quad (15)$$

If the corresponding path coefficients of MCLS are non-negative at the beginning, then it is simple. We just need to divide the two sides of the Equations (3) and (4) by a constant. This constant should be the sum of the corresponding path coefficients in MCLS. For example, in the Equation (3), if  $\sum_{j=1}^{K(t)} \psi_{tj} = c_t$ , then the two sides of the Equation (4) are divided by the constant  $c_t$ , so the modular length of structural variables becomes  $1/c_t$ , and  $\sum_{j=1}^{K(t)} \psi_{tj} = 1$ .

If the corresponding path coefficients of MCLS are negative at the beginning, we cannot copy the method of prescription regression proposed by Fang (1982, 1985), because regression dependent variables are not completely known. Now we know the direction of regression dependent variables, but the modular length is undetermined. According to the theorem in Fang (1982), if the initial regression coefficients have negative ones, whose

prescription regression coefficient should be 0. So we can first make ordinary regression about MCLS, where the modular length of dependent variables is 1. If there are some non-positive terms in the initial regression coefficients, we can get rid of these variables, and thus the corresponding regression coefficient is 0. Then the two sides of the Equations (3) and (4) can be divided by a constant that should be the sum of the corresponding path coefficients in MCLS, as discussed in the previous paragraph.

Of course we can improve the constraint of the prescription condition. If some regression coefficient is 0, its corresponding variable may be removed from the model, which is not a desired situation. To avoid this, we may change the prescription condition and let  $\psi_{tj} \geq \delta$ ,  $\omega_{ij} \geq \delta$ , where  $\delta > 0$ . If some initial regression coefficients are less than  $\delta$ , they all are changed as  $\delta$ , and the corresponding independent variables with coefficient  $\delta$  should be moved to the left side of the equation in regression process.

Summarizing the above discussion we can continue to improve the algorithm of MCLS.

**Algorithm 2.** Improvement on Step 3 of Algorithm 1.

**Step 5.** After getting the estimate of structural variables  $\hat{\xi}_t, \hat{\eta}_i$  in Step 2, calculate the summarizing coefficients  $\psi_{tj}, \omega_{ij}$  by prescription regression, and recalculate the estimates of  $\xi_t, \eta_i$ .

- 1) Make use of  $\hat{\xi}_t, \hat{\eta}_i$  directly in Step 2 and calculate  $\hat{\psi}_{tj}, \hat{\omega}_{ij}$  in  $SEM^+$  by common regression.
- 2) For any  $t$ , if there are  $\hat{\psi}_{tj} \geq \delta$ , ( $\delta \geq 0$ ) for all  $j$ , and  $\sum_{j=1}^{K^{(t)}} \psi_{tj} = c_t$ , then divide both sides of Equation (3) by  $c_t$ . Similarly, for any  $i$ , if there are  $\omega_{ij} \geq \delta$ , ( $\delta \geq 0$ ) for all  $j$ , and  $\sum_{j=1}^{L^{(i)}} \omega_{ij} = c_i$ , then divide both sides of Equation (4) by  $c_i$ .

After checking all  $t, i$ , go to Step 4 in Algorithm 1.

- 3) For any  $t, i$ , if there is some  $j$  so that  $\hat{\psi}_{tj} < \delta$ , or  $\omega_{ij} < \delta$ , ( $\delta \geq 0$ ), then let the corresponding term be fixed, *i.e.*,  $\hat{\psi}_{tj} = \delta$  or  $\omega_{ij} = \delta$ . After checking all  $j$ , go to Step 1 and Step 2 in this algorithm.

Note that if some regression coefficient is fixed in common regression, the corresponding independent variables with its coefficient  $\delta$  should be moved to the left side of the equation and combined with the dependent

variable to regression. After regression the corresponding independent variable with its coefficient  $\delta$  should be moved to the right side of the equation.

## 5. Conclusions and Discussions

In this paper, a structural equation model is proposed to analyze the quality of life in cancer patients. Besides, a definite linear algorithm for SEM based on the modular constraint of structural variable, the least square theorem, and the prescription regression method is introduced, this model is more objective and scientific compared with traditional methods, such as descriptive analysis method and some simple test methods, including Correlation Analysis and T test and so on, because the summarizing coefficients of this evaluation system are calculated by samples rather than designed arbitrarily. Therefore, we can have a better understanding of the relationships among the indexes, which will do a great favor to decision-making analysis of Quality of Life in Cancer Patients.

## REFERENCES

- [1] P. Ravasco, I. Monteiro-Grillo and M. E. Camilo, "Does Nutrition Influence Quality of Life in Cancer Patients Undergoing Radiotherapy?" *Radiother Oncology*, Vol. 67, No. 2, May 2003, pp. 213-220.
- [2] C. C. Gotay, E. I. Korn, M. S. McCabe, *et al.*, "Building Quality of Life Assessment into Cancer Treatment Studies," *Oncology*, Vol. 6, 1992, pp. 25-28.
- [3] D. F. Cella and E. A. Cherin, "Quality of Life During and after Cancer Treatment," *Comprehensive Therapy*, Vol. 14, 2001, p. 69.
- [4] F. Gu, "Quality of Life Measurement in 233 Cancer Patients," *Chinese Cancer*, Vol. 8, No. 8, 1999, pp. 356-358.
- [5] J. P. Wang and J. N. Cui, "Quality of Life and Factors that Influencing it among Cancer Patients in China," *Chinese Journal of Clinical Psychology*, Vol. 8, No. 1, 2000, pp. 23-26.
- [6] H. Q. Tong, "Data Analysis and Statistical Computation (DASC)," Electronic Publication, Science Press of China, Beijing, 2005.
- [7] C. M. Wang and H. Q. Tong, "Best Iterative Initial Values for PLS in a CSI Model," *Mathematical and Computer Modeling*, Vol. 46, No. 3-4, August 2007, pp. 439-444.
- [8] Y. H. Jiang, Z. Li and Y. F. Cheng, "The Relation of Psychological Factor and Cancer," *Chinese Journal of Clinical Rehabilitation*, Vol. 6, No. 1, 2002, pp. 92-93.