Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2016; 2(1): 25-31.

Original Article

Comparing the classic factor analysis with Bayesian factor analysis on the data of Family Dermatology Life Quality Index

Maryam Zamani[1], Abbas Bahrampour[2*], Nouzar Nakhaee[3]

[1] Department of Biostatistics, School of Health, Kerman University of Medical Sciences, Kerman, Iran
[2] Department of Biostatistics and Epidemiology, School of Health Kerman, Research Center for Modeling in Health Kerman University of Medical Sciences, Iran
[3] Department of Neurology, Neuroscience Research Center, Kerman University of Medical Sciences, Kerman, Iran

ARTICLE INFO

Received 03.05.2015
Revised 09.10.2015
Accepted 13.11.2015
Published 15.03.2016

Available online at:
http://jbe.tums.ac.ir

Key words:
Bayesian factor analysis;
Monte Carlo-Markov chain algorithm;
Family Dermatology Life Quality Index

ABSTRACT

Background & Aim: Chronic diseases impact not only on patients but also on their family members' lives. This study aims to determine dimensions of Family Dermatology Life Quality Index (FDLQI) questionnaire by the use of classic and Bayesian factor analysis (BFA) factor.

Methods & Materials: In this study, FDLQI questionnaire distributed among 100 family members of dermatological patients. BFA is exploited to determining dimension and contribution of items of the questionnaire in different sample size. BFA is examined using the Monte Carlo-Markov chain algorithm. All the above analysis is done in sample size 100, 130, 150.

Results: In this study, 100 family members of dermatological patient attended to evaluate Persian version of FDLQI questionnaire. A mean age of participants was 37.1 years (± 12.3). A mean score of FDLQI was 15.4 (± 5.5) with maximum and minimum scores of 30 and 6, respectively. Exploratory FA revealed a one-factor solution that accounted for 45.87% of the total variance. The unidimensional model was concordance by confirmatory FA. For more exploration, BFA was performed. Two factors extracted when iteration is done.

Conclusion: It appears that when sample size diminished, Cronbach's alpha and Kaiser–Meyer–Olkin increased. Among 10 items of the questionnaire, item 9 mostly appears differently in results.

Introduction

Chronic diseases impact not only on patients but also on their family members' lives. Various aspects of life could be affected such as social relationships, financial status, recreational activities, work attendance, and flexibility (1). Even though chronic dermatological diseases have no direct risk for life, they can affect the quality of life of patients due to their associated symptoms (itching and pain), psychological problems (low self-esteem and depression), by affecting social and familial relationships and as a result of therapeutic burdens (excessive time consumption and financial burden) (2). In other words, chronic dermatological conditions, in addition to having their primary impact on the patients, can have what is known as the "secondary impact" on patient's partners and immediate family members (3). The secondary impact of skin diseases on patient's family members can manifest itself in a number of ways such as psychological distress, burden of care, additional housework, effect on social life, recreational activities and holidays, financial strain, adverse impact on physical wellbeing, job/education, sleep, eating/drinking, need for support, problems with peoples' attitudes, and negative impact on marital relationships (4). All

* Corresponding Author: Abbas Bahrampour, Postal Address: Department of Biostatistics, Kerman University of Medical Sciences, Kerman, Iran. Email: abahrampour@yahoo.com

effects on the sufferers have been extensively documented in the literature, but the effects of dermatological diseases on the quality of life of patients' family members have been less recognized and published (4). Family Dermatology Life Quality Index (FDLQI) is the unique questionnaire that has already been designed for evaluating the quality of life of family members of patients with dermatological diseases and it can be used in all types of dermatological diseases (4, 5). Recently, this instrument translated into 14 different languages, and questions concepts are similar in different languages. For example, Sampogna et al. (6) used for the first time the Italian version of the FDLQI to evaluate the burden of recessive dystrophic EB on family caregivers. It seems to be a useful tool. Safizadeh et al. (7) described the cultural adaptation of FDLQI questionnaire and to assess psychometric properties of the Persian version.

This study is gathering with the aim of determining the dimensions of FDLQI questioner by comparing classic and Bayesian factor analysis (BFA). Studies which discussed in this field are less recognized and published.

Methods

FDLQI is a specific tool for the evaluation of the impact of dermatological diseases on the quality of life of patients' family members. In this study, our data are collected through questionnaire which was distributed among 100 family members of patients with dermatological diseases after obtaining their consent for participating in the study. The inclusion criteria for subjects were as follows: age over 18 years, able to read and understand Persian, and living in the same household. FDLQI questionnaire contains 10 questions evaluating the outcomes of dermatological diseases on different aspects of life of patient's family members over the previous month. The items assess physical and psychological well-being, interpersonal relationships, social life, leisure activities, burden of care, job/study activities and housework, and expenditure. Each question has four response options: "not at all/not relevant," "a little," "quite a lot," and "very much" scored as 0, 1, 2 and 3, respectively. Final score (range = 0-30) is obtained by adding the scores of 10 questions. The higher score shows more effect on the quality of life. Those having a severe non-dermatological disease were excluded (5). The subjects were recruited from outpatient clinics of Kerman University of Medical Sciences and private clinics.

In our analysis, we first examine the suitability of data for FA using Kaiser–Meyer–Olkin (KMO) and Bartlett's tests. To assess the internal consistency reliability of the scale, Cronbach's alpha was computed. Numbers of factor models are determined through scree plot. No rotation is used in this analysis. We use Monte Carlo-Markov Chain (MCMC) algorithm to achieve the Bayesian calculation required in the FA. Comparing BFA and classic FA provide chance to see the way that items present. Our analysis is done on sample size of 100, 130,150. A sample size of 130 and 150 have generated with simulation. According to the questionnaire's responses simulation was done for sample size more than 100. All the above analysis is done on these sample sizes. R software (R Foundation for Statistical Computing, Vienna, Austria) is used in this study. Loading factors, which they are higher than the minimum acceptable level of 0.4, were significant. According to summarizing and being significant, the number of factor was determined.

FA

FA is a multivariate statistical method. This method is a very helpful when there exist a large number of observable factors. These observable factors which normally have large dimensions can be expressed by a set of unobservable factors that relatively have much smaller dimensions (8). This is one of the major causes that have made FA very popular recently (9).

The factor model can be expressed as follows:

$$Y = \Gamma_q f + \mu + e \tag{1}$$

Where, Y is a (p × 1) random vector and $\Gamma_q \in R^{p \times q}$ is co-called loading factor, f is a (q × 1) random vector of latent (common) factors, $\mu \in R^{p \times 1}$ denotes the mean of Y, $e \in R^{p \times 1}$ is the error term (10).

Assumptions used for construction of the above model are as follows:

$$E(f) = 0 \quad Var(f) = I \quad E(e) = 0 \quad Cov(e_k, e_l) = 0 \quad k \neq l \qquad (2)$$

$$Var(e) = \Psi = diag(\varphi_1, \dots \varphi_p)$$

$$Cov(e, f) = 0$$

$$Var(y) = \Sigma = \Gamma_q \Gamma'_q + \Psi$$

$$Cov(y_i, y_i) = \sum_{k=1}^{q} \gamma_{ik}\gamma_{jk} \qquad i \neq j$$

In FA, to nearest points placed around factors axis and natural grouping variable specified, rotation is made. Orthogonal and oblique methods are two approaches used for rotation purpose. It is worth which mentioning that Bayesian method does not require any rotation (9).

The following assumptions are considered for the likelihood function:

i. $\varepsilon_i \sim N(0, \Psi)$ $\Psi \equiv diag(\varphi_1, \dots \varphi_p)$, $\varphi_j > 0$, $j = 1, \dots, p$ (3)
ii. $(f_i|q) \sim N(0, R)$ $q \leq p$ $R \approx I_q$
iii. ε_i and f_i are independent.

One can conclude from i and ii that, $(y_i|\mu, \Gamma, f_i, q) \sim N(\mu + \Gamma f_i, \Psi)$. For simplicity, μ is assumed to be zero. Thus, we can write observations likelihood as:

$$p(Y|F, \Gamma, \psi, q) \propto |\psi|^{-\frac{n}{2}} e^{-\frac{1}{2} tr\psi^{-1}(Y-F\Gamma')'(\alpha Y - F\Gamma')} \qquad \Psi > 0 \quad (3)$$

$$Y' \equiv (y_1, \dots, y_n) \qquad F' \equiv (f_1, \dots, f_n)$$

And the joint function of y_i, f_i is as follows:

$$p(fi, yi|\mu, \Gamma, R, \psi, q) \propto e^{-\frac{1}{2}(f_i - \hat{f}_i)'(R^{-1} + \Gamma'\psi\Gamma)^{-1}(f_i - \hat{f}_i)} e^{-\frac{1}{2}(y_i - \mu)'(\psi + \Gamma R \Gamma')^{-1}(y_t - \mu)} \qquad (4)$$

The natural conjugate family of generalized is used for posterior distribution of the parameters. It is assumed that Γ dependents on Ψ. However, parameters F and Ψ are supposed to be independent. The same holds for F and Γ. Hence, the prior joint distribution for parameters F, Γ, Ψ can be written as:

$$P(F, \Gamma, \Psi|q) \equiv P(\Gamma|q, F, \Psi) \times P(F|q, \Psi) \times P(\Psi|q) \qquad (5)$$

$$p(\Gamma|\psi, q) \propto |\psi|^{-\frac{q}{2}} e^{-\frac{1}{2} tr\psi^{-1}(\Gamma - \Gamma_0)H(\Gamma - \Gamma_0)'} \qquad (6)$$

$$p(\psi) \propto |\psi|^{-\frac{b}{2}} e^{-\frac{1}{2} tr\psi^{-1}B} \qquad v > 2p \qquad (7)$$

$$P(F|q) \propto e^{-\frac{1}{2} trF'F} \qquad (8)$$

Where $\Psi > 0$, $H > 0$, $B > 0$ and B is a diagonal matrix.

One should note that (Γ| Ψ) has joint normal distribution; however, Ψ^{-1} has Wishart distribution. Furthermore, (H, Γ_0), (B, υ) are hyper parameters that needs to be evaluated. f_i are independent and has a normal distribution (9). In Bayesian statistics, hyper parameters are priori distribution parameters.

Based on the principle of Bayesian, posterior distribution is proportional to the product of the likelihood function in prior density function. Posterior distributions in FA in Bayesian method with uncertain parameters of interest are as follows:

$$p(F, \Gamma, \psi|Y, q) \propto e^{-\frac{1}{2} trF'F} |\psi|^{\frac{(n+q+v)}{2}} e^{-\frac{1}{2} tr\psi^{-1}U} \qquad (9)$$

$$U \equiv (Y - F\Gamma')'(Y - F\Gamma') + (\Gamma - \Gamma_0)H(\Gamma - \Gamma_0)' + B \qquad (10)$$

MCMC algorithm

It is shown that BFA can be done through MCMC approach (11). This approach helps Bayesian analysis to overcome some computational limits and enables statisticians to extract reasonable priors (10). By utilizing MCMC algorithm, random sample of posterior distribution can be generated (9). The importance of this method is in complex model where there is a rare possibility to directly sample from the posterior distribution. This approach suggests one of the easiest ways to have stable results (8).

Results

Mean age of participants was 37.1 (± 12.3) years and 58 were females. Descriptive analysis of main data set is shown in table 1. The mean FDLQI the mean FDLQI score was 15.4 (standard deviation = 5.5, median = 15, range = 6-30). The highest- and lowest-scoring FDLQI items were financial burden and effect on housework, respectively (Table 2). According to the FA only one general factor can explain whole items (Figure 1).

Table 1. Demographic characteristics of dermatological patient's family members (n = 100)

Characteristic	Frequency
Age-group	
≤ 40 years	63
> 40 years	37
Gender	
Male	42
Female	58
Education	
Illiterate	20
Primary school	24
Secondary school	20
High school	25
Diploma	
Collage	11
Relationship to patient	
Parent	42
Sibling	27
Spouse	11
Son/daughter	10
Others	10
Disease duration	
1 month	57
1-3 months	14
≥ 3 months	29

It turns out that KMO and Bartlett criterion of main data set is 0.85. Furthermore, scree plot suggests that information can be summarized in one factor (Figure 1). Exploratory FA revealed the presence of one-factor structure underlying the items of the FDLQI, which explained 45.87% of the total variance (Table 3). The Cronbach's alpha of the Persian version of FDLQI was 0.87. For more exploration, BFA was done to see how many factors will be extract when iteration is done. The result of classic FA and BFA are in table 3.

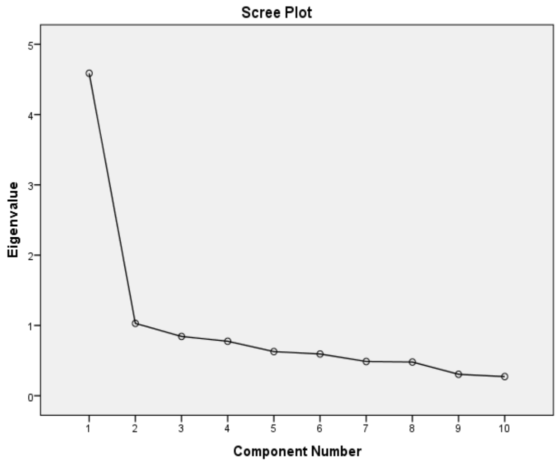

Figure 1. Scree plot of 10 items (only one eigenvalue above 1)

Cronbach's alpha for sample size 130 is 0.79. According to the above analysis KMO and Bartlett criterion for this is 0.81. Using the scree plot (Figure 2), FA resulted in a two-factor solution that accounted for 46.94% of the variance. Furthermore, classic FA and BFA are in the table 4. As iteration is done BFA revealed two factors.

Table 2. Mean scores of FDLQI scores and percentage of item responses

Item	Mean (SD)	Percentage of item responses			
		Not at all/not relevant	A little	Quite a lot	Very much
Emotional impact	1.79 (0.73)	4	27	55	14
Physical well-being	1.69 (0.73)	4	35	49	12
Relationships	1.42 (0.78)	10	46	36	8
People's reaction	1.62 (0.84)	7	40	37	16
Social life	1.66 (0.87)	10	30	44	16
Leisure activities	1.14 (0.85)	23	47	23	7
Burden of care	1.65 (0.81)	7	35	44	14
Housework	1.11 (0.87)	28	38	29	5
Job/study	1.49 (0.88)	16	29	45	10
Financial burden	1.84 (0.75)	2	31	48	19

FDLQI: Family Dermatology Life Quality Index, SD: Standard deviation

Table 3. Loading factors of classical and BFA for sample size 100

Item number	Classical loading factor		Bayesian loading factor	
	F1	F2	F1	F2
Emotional impact	0.702	-0.098	-0.707	-0.157
Physical well-being	0.703	0.165	-0.694	-0.216
Relationships	0.699	0.475	-0.629	-0.947
People’s reaction	0.648	-0.33	-0.658	0.013
Social life	0.711	0.102	-0.721	-0.259
Leisure activities	0.744	0.257	-0.732	-0.308
Burden of care	0.649	-0.452	-0.672	0.034
Housework	0.728	-0.182	-0.802	0.099
Job/study	0.522	0.451	-0.469	-0.264
Financial burden	0.640	-0.318	-0.643	-0.017
KMO	0.85			

KMO: Kaiser–Meyer–Olkin, BFA: Bayesian factor analysis

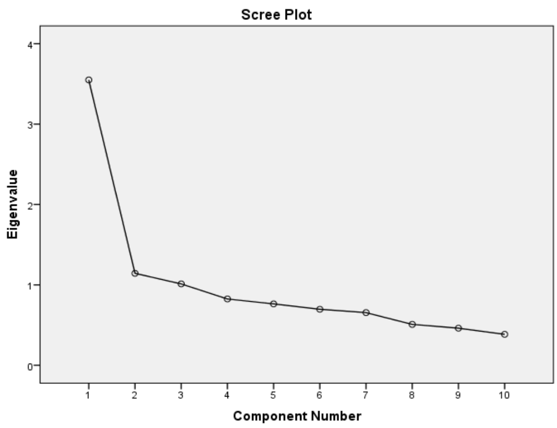

Figure 2. Scree plot of 10 items (only two eigenvalue above 1)

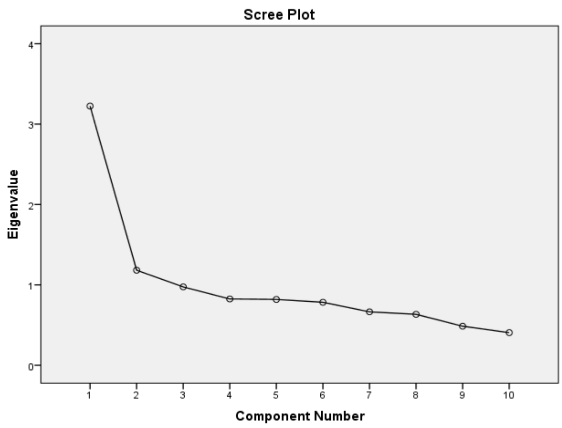

Figure 3. Scree plot of 10 items (only two eigenvalue above 1)

Lastly, KMO and Bartlett test for sample size 150 turns to be 0.78 while Cronbach’s alpha for this data set is 0.76. By using exploratory FA, two factor with explaining 44.1% total variance derived from this data set (Figure 3). Way of positioning items in latent factors under two methods of analysis come in table 5.

Exploratory FA revealed one factor in sample size of 100 and in other sample sizes two factors. However, BFA in sample size of 150 showed one factor and in other sizes two factors.

For the main data the exploratory FA, “leisure activities” has high loading factor while in BFA, “relationship” and” housework” had high loading score.

Table 4. Loading factor of classic and BFA for sample size 130

Item number	Classical loading factor		Bayesian loading factor	
	F1	F2	F1	F2
Emotional impact	0.574	0.040	−0.416	0.252
Physical well-being	0.614	0.263	−0.386	0.368
Relationships	0.581	-0.360	−0.285	0.560
People’s reaction	0.591	-0.370	-0.490	0.213
Social life	0.671	-0.461	-0.0459	0.426
Leisure activities	0.630	-0.101	-0.360	0.472
Burden of care	0.624	0.159	-0.737	-0.085
Housework	0.667	0.164	-0.528	0.278
Job/study	0.401	0.523	-0.231	0.250
Financial burden	0.560	0.354	-0.544	0.039
KMO	0.081			

KMO: Kaiser–Meyer–Olkin, BFA: Bayesian factor analysis

Table 5. Loading factor of classic and BFA for sample size 150

Item number	Classical loading factor		Bayesian loading factor	
	F1	F2	F1	F2
Emotional impact	0.540	0.134	-0.391	-0.186
Physical well-being	0.624	0.178	-0.416	-0.241
Relationships	0.523	0.459	-0.353	-0.339
People's reaction	0.578	-0.419	-0.422	-0.036
Social life	0.656	0.037	-0.518	-0.189
Leisure activities	0.602	0.300	-0.402	-0.364
Burden of care	0.624	-0.383	-0.727	0.294
Housework	0.606	-0.150	-0.453	-0.101
Job/study	0.375	0.498	-0.258	-0.165
Financial burden	0.486	-0.489	-0.368	0.075
KMO	0.78			

KMO: Kaiser–Meyer–Olkin, BFA: Bayesian factor analysis

In exploratory FA recognized "social life" and "job/study" have high loading factor in sample size of 130. However, BFA marked "burden of care" and "relationships" as high loading factors. At last in sample size 150, "job/study" and "social life" in classic FA and "burden of care" in BFA are seen to have high loadings.

Discussion

Dermatological diseases have side effect on the quality life of family members of patients. Therefore, beside caring dermatological patients, it is a very important to pay appropriate attention to their family members. Persian version of FDLQI showed that the family members of patients with dermatological diseases showed a lower quality of life comparing to the original sample of Barsa et al.'s study (5) while in Persian version most affected domains were, respectively, "financial burden" and "emotional impact," and "housework" had been affected less than other domains, in the study on original questionnaire, "emotional impact" and "job/study" were, respectively, the most and least affected domains (7). Other studies have also shown that type of the involved disease and cultural conditions effects on family members of patients (12).

As results showed when the sample size decreased KMO and Cronbach's alpha increased. Classic FA extract one factor for main data set and it is in line with the FA results of the original version which was in favor of one dimensionality of the tool (5) while BFA do the same for sample size 150. Sampogna et al. (6) showed one dimensionality of the questionnaire in the Italian version, too. However, in two last sample size exploratory FA summarized information on two factors which are obtained from analysis. In other study (13), 132 family member's patient which FA of 10 items of the final questionnaire revealed two factors.

BFA has a major advantage over the traditional maximum likelihood FA of incorporating prior information into the model. With complicated models, it is rare that samples from posterior distribution can be obtained directly. The methods of "MCMC" are proved to be the easiest way to get reliable results (8). Ansari and Jedidi (14) reported that Bayesian approach also has promise for estimating more complex data structures.

Press and Shygmasv showed BFA techniques which loading scores are estimable along other parameters (15).

In 1981, Lee studied FA in view point of Bayesian in which posterior information of parameters included in analysis. Finally, the conclusion is that Bayesian estimates of FA are considerably better. Hence, if appropriate prior be available, the Bayesian approach can be an interesting method for FA.

Studies in medical fields by topic of comparing classic FA with BFA are very rare. Moreover, no studies have been done on FDLQI questionnaire with BFA method.

The results Safizadeh et al. (7) showed that the Persian version of FDLQI has acceptable factorial validity and internal consistency reliability and could be used in related studies. Following the same reason, we attempt to do further analysis.

This research is a part of Maryam Zamani thesis of Master of Science in biostatistics.

Acknowledgments

None.

References

1. Baanders AN, Heijmans MJ. The impact of chronic diseases: the partner's perspective. Fam Community Health 2007; 30(4): 305-17.
2. Safizadeh H, Shamsi-Meymandy S, Naeimi A. Quality of life in Iranian patients with acne. Dermatology Research and Practice 2012; 2012: 4.
3. Finlay AY. Quality of life indices. Indian J Dermatol Venereol Leprol 2004; 70(3): 143-8.
4. Basra MK, Finlay AY. The family impact of skin diseases: the greater patient concept. Br J Dermatol 2007; 156(5): 929-37.
5. Basra MK, Edmunds O, Salek MS, Finlay AY. Measurement of family impact of skin disease: Further validation of the Family Dermatology Life Quality Index (FDLQI). J Eur Acad Dermatol Venereol 2008; 22(7): 813-21.
6. Sampogna F, Tabolli S, di Pietro C, Castiglia D, Zambruno G, Abeni D. The evaluation of family impact of recessive dystrophic epidermolysis bullosa using the Italian version of the Family Dermatology Life Quality Index. J Eur Acad Dermatol Venereol 2013; 27(9): 1151-5.
7. Safizadeh H, Nakhaee N, Shamsi-Meymandi S, Pourdamghan N, Basra MK. Preliminary reliability and validity of Persian version of the Family Dermatology Life Quality Index (FDLQI). Qual Life Res 2014; 23(3): 869-75.
8. Faghih M. Bayesian factor analysis: applied on quality of life data [Thesis]. Shiraz, Iran: Shiraz University of Medical Sciences; 2012. [In Persian].
9. Rowe DB. Correlated bayesian factor analysis [Thesis]. San Diego, CA: University of California; 1998.
10. Cao Y, Evans M. A Bayesian approach to factor analysis via comparing prior and posterior concentration [Thesis]. Toronto, ON: University of Toronto; 2010.
11. Ghosh J, Dunson DB. Default prior distributions and efficient posterior computation in Bayesian factor analysis [Thesis]. Durham, NC: Duke University; 2009.
12. Knafl KA, Gilliss CL. Families and chronic illness: a synthesis of current research. Journal of Family Nursing 2002; 8(3): 178-98.
13. Basra MK, Sue-Ho R, Finlay AY. The Family Dermatology Life Quality Index: measuring the secondary impact of skin disease. Br J Dermatol 2007; 156(3): 528-38.
14. Ansari A, Jedidi K. Bayesian factor analysis for multilevel binary observations. Psychometrika 2000; 65(4): 475-96.
15. Merkle E. Bayesian estimation of factor analysis models with incomplete data [Thesis]. Columbus, OH: Ohio State University; 2005.