Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2016; 2(1): 40-46.

Original Artcle

Longitudinal relationship between academic staffs' evaluation score by students and their characteristics: Does the choice of correlation structure matter?

Malihe Rezaei[1], Ali Akbar Haghdoost[2], Maryam Okhovati[3], Farzane Zolala[2], Mohammad Reza Baneshi[4*]

[1] Department of Biostatistics, School of Health, Kerman University of Medical Sciences, Kerman, Iran
[2] Department of Epidemiology, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[3] Department of Library and Information Science, Medical Informatics Research Center, Kerman University of Medical Sciences, Kerman, Iran
[4] Department of Biostatistics, Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

ARTICLE INFO

Received 17.10.2015
Revised 16.12.2015
Accepted 05.01.2016
Published 15.03.2016

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

Key words:
longitudinal data;
generalized estimating equations;
correlation structure;
evaluation;
simulation study

ABSTRACT

Each semester, students are asked to evaluate the academic staff through an online questionnaire. Generalized estimating equations model (GEE), taking into account the correlation between scores, is the established tool to analyze longitudinal data. The aim of this manuscript is to identify characteristics that influence staff score and to address the importance of selection of appropriate correlation structure. We analyzed scores of 336 staff in six consecutive semesters applying GEE with three correlation structures: exchangeable, autoregressive, and unstructured. We also compared the performance of these correlation structures via simulation study. Three normally distributed outcomes with exchangeable correlation structure were simulated. Four independent variables (two continuous and two binary) of which only one was related to the outcome were generated. In the empirical data set, time and academic degree were positively correlated with staffs' score. Our simulation study showed that the probability that autoregressive and unstructured correlation structures select wrong predictors as being significant is 1.3% and 3.7%. We concluded that evaluation of staff by students improved the quality of education. In addition, selection of inappropriate correlation structure affects the significance of variables.

Introduction

Academic staff plays the central role in education and training of specialists and qualified students (1). Therefore, the universities usually ask students to evaluate academic staff's teaching performance (2, 3). The evaluation process is carried out using different methods. Among these methods are: self-evaluation, evaluation by the head of the unit and also by other colleagues and student evaluation (1-5). Evaluation by students happens in universities across the world, as students directly monitor the quality of teaching by staff (3, 5, 6). It has been argued that as students directly receive the services, i.e., education, they can provide valuable information on the quality of teaching (1).

The evaluation process usually happens every semester. This provides longitudinal data sets where scores of staff over time are correlated. However, usually cross-sectional techniques are used to analyze such data (7). For example, in research conducted at Jondi-Shapour University of Medical Sciences of Ahvaz, longitudinal data were analyzed by chi-

* Corresponding Author: Mohammad Reza Baneshi, Postal Address: Research Centre for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Email: m_baneshi@kmu.ac.ir

Please cite this article in press as: Rezaei M, Haghdoost AA, Okhovati M, et al. Longitudinal relationship between academic staffs' evaluation score by students and their characteristics: Does the choice of correlation structure matter? *J Biostat Epidemiol. 2016; 2(1): 40-46*

square test and Pearson correlations (6). In another study in Arak, correlation and regression and non-parametric Spearman and Kendall and Kruskal–Wallis correlation tests were used (8).

The advantage of longitudinal analyzes over cross-sectional methods is that the correlations among observations are taken into account (9) and allows to explore the shape of the association between the outcome and time (10). The main difference between longitudinal methods and standard statistical methods, such as linear regression, is that the former incorporates a correlation structure into the analysis.

It has been argued that selection of the correct correlation structure leads to right inferences and estimating of accurate statistics (11). Main correlation structures usually applied are exchangeable, autoregressive, and unstructured (see methods section).

Another advantage of longitudinal models over cross-sectional methods is that the former allow for inter- and intra-person interpretations of time-dependent variables. For example, suppose the regression coefficient for the working experience variable is equal to three. One interpretation is that if a person has 10 years' working experience his/her evaluation score is three units higher than his colleague with 9 years' working experience. Another interpretation is that score of staff would increase 3 points each year.

The aim of this manuscript is to investigate whether staff characteristics affect their teaching scores. In addition, we addressed the importance of selection of appropriate correlation structure on estimates derived using empirical and simulated data.

Methods

Our sample comprised 336 academic staff affiliated to the Kerman University of Medical Sciences (KUMS). Evaluation scores were collected from Education and Development Center. To ensure confidentiality, names were removed from the dataset.

Evaluation scores were available for six consecutive semesters and ranged 1-5. We transformed the data to range of 0-100 to make the interpretation easier. Independent variables were: gender (male and female), teaching experience (in years), scientific degree (instructor, assistant professor, associate professor, and professor), and school in which staff works (medical, dentistry, pharmacy, health, management and information science, paramedical, and nursing).

As data in some semesters for some members were not available, to maximize the power, missing data were imputed using expectation maximum algorithm.

Generalized estimating equations (GEE) model was fitted to assess the impact of independent variables on the outcome (i.e., staffs' teaching scores). GEE model fits a simple linear regression model to the data by assuming that all observations are independent. The second step is to calculate the residuals followed by calculation of correlation correction factor so as to correct standard errors (SEs). In other words, the second regression model is fitted to the data by integrating a correlation matrix into the analysis. These steps are repeated until convergence (12).

The general form of GEE is written as equation (1):

$$y_{it} = \beta_0 + \sum_{J=1}^{J} \beta_{1j}\, x_{itj} + \beta_2 t + \cdots + \mathrm{CORR}_{it} + \varepsilon_{it} \quad (1)$$

Where, y_{it} is response variable for i^{th} individual at time t, β_0 is constant (intercept), β_{1j} is regression coefficient of independent variable j, x_{itj} is the independent variable j for i^{th} individual at time t, J is the total number of independent variables, β_2 is regression coefficient for variable time, t is the time, CORR is correlation structure, and ε_{it} is error for i^{th} individual at time t.

An important issue in GEE analysis is selection of appropriate correlation structure (10). The most commonly used correlation structures are exchangeable, autoregressive, and unstructured. Exchangeable correlation structure assumes that the correlation of observations is constant and does not change over time. Autoregressive correlation structure assumes that the correlation of observations is an exponent subordinate of the time interval between repeated observations which means that correlation between first time and the second time is more than the fourth time and first time. The unstructured correlation structure assumes that there is no clear trend for correlation of observations (13).

We fitted the GEE model with the above three correlation structures. We used the following three criteria for selection of appropriate correlation matrix: (a) considering the correlation matrix of observations, (b) simplicity of the selected correlation structure, and (c) quasi likelihood under the independence model criterion (QIC) (10). QIC is the modified version of Akaike information criterion, which takes into account the correlation between observations. Model with the lowest QIC provides the best fit to the data (14, 15).

To further explore the behavior of correlation structures, we simulated four normally distributed outcomes (Y_i) with constant variance and with exchangeable correlation structure with of about 0.40. Two Bernoulli independent variables (x_1 and x_2) and two continuous independent variables with a normal distribution (x_3 and x_4) were generated. The data were generated so that x_3 was associated with the outcome, while x_1, x_2, and x_4 did not. This process was repeated 100 times while sample size at each simulated data set was 200.

We fitted exchangeable, autoregressive, and unstructured correlation models and counted the number of times x_3 lost their significance and also number of times x_1, x_2, and x_4 wrongly selected as being significant.

Results

Academic staff affiliated to KUMS was 61% male and 39% female. The mean (standard deviation) of teaching experience was 15 (8.45) years. In total, 46% and 21% were assistant and associate professor, respectively. About 13% and 20% were instructor and professor. More than half of them staff were affiliated to medical school. Frequency of missing rate in evaluation scores in six semesters were 14.6%, 14%, 31%, 34.5%, 29%, and 12.5%, respectively.

Mean (SE) of evaluation score for females and males were 75.14 (0.85) and 74.83 (0.854), respectively (Table 1). The highest mean was observed for those in professor category (77.12 with SE of 1.21). In terms of school, the highest and lowest scores were seen in members of dental and paramedical schools, respectively [78.57 (1.25) vs. 72.68 (3.12)].

Table 1. Mean of evaluation score of members of Kerman University of Medical Sciences based on gender, academic degree, and school

Variable	Mean	SE
Gender		
Man	74.83	0.854
Woman	75.13	0.85
Scientific degree		
Professor	77.12	1.21
Assistant professor	73.05	0.94
Associate professor	74.48	1.24
Coach	75.29	1.49
Working location college		
Health	73.62	1.82
Management and Information	75.04	2.14
Paramedics	72.68	3.12
Nursing	75.39	1.64
Pharmacy	76.74	1.54
Dentistry	78.57	1.25
Medical	72.85	0.72

SE: Standard error

The correlation coefficient between evaluation scores varied from 0.20 to 0.40 which suggests that an exchangeable correlation structure might adequately fit the data. Based on the QIC values, the unstructured model showed the poorest performance. However, no remarkable difference between exchangeable and autoregressive models was seen (Table 2). Therefore, model with exchangeable structure correlation was chosen as the appropriate model.

Based on the best model, each semester, evaluation scores were increased by about one point (Table 2). The autoregressive model revealed similar estimate, but the estimated coefficient in the unstructured model was nearly half, at 0.54. In all models, the mean score of professors was significantly higher than associate professors. Furthermore, scores of members of pharmacy and dentistry schools were significantly higher than members of medical school (about five points). In the unstructured correlation structure, the difference between mean scores of health school and management and information science school with the medical school were also significant. Impact of gender and teaching experience was not significant in neither of models (Table 2).

In our simulation study, the performance of all three correlation structures was similar in terms of power. That is, all three models were able to capture the significance of correct independent variables. However, the autoregressive and unstructured models had higher type one error. In 1.3% and 3.7% of scenarios, a wrong variable was selected as being significant.

Discussion

Many studies have performed to address the evaluation of academic staff by students [3]. However, the majority of them applied cross-sectional analysis to data with longitudinal nature. The use of cross-sectional techniques leads to bias in estimation of SEs, which causes wrong inferences. This makes a comparison of our results with other studies difficult. In this study, considering the correlation of evaluation scores during the six semesters, we applied GEE with different correlation structures. We have observed that choice of correlation structure affects the significance of variables.

Based on our results, a linear trend was seen in evaluation scores over time. In our study, the evaluation scores have statistically significant trend and have been growing over time. This result may indicate improving and promoting quality of education process in university and evaluation process of faculty by students. This was result consistent with results observed analyzing evaluation score of members of Arak University of Medical Sciences (8). Similar results reported from similar studies implemented in Texas University, Queensland in Australia, and Canada. In a similar study performed in Jondi-Shapour University, ignoring the correlation between evaluation scores, no trend was observed (16).

We also observe that the degree of staff could affect their evaluation scores. The mean score for professors as about 4 units more than assistant professors and was of statistical significance, score of instructors was higher than assistant professors who were not of statistical significance**.** Findings of Shahid Bahonar University suggested that evaluation score for professor and associate professor was significantly higher than instructor and assistant professor (17). On the other hand, in Ahvaz University, score of instructors was significantly higher than others (18).

Table 2. Assessment of variables that influence staff score using different correlation structures

Type of model	**GEE models**								
Correlation structure	**Exchangeable**			**Autoregressive**			**Unstructured**		
Variables	**B**	**SE**	**P**	**B**	**SE**	**P**	**B**	**SE**	**P**
Time	0.98	0.16	< 0.001	1.06	0.17	< 0.001	0.54	0.16	< 0.001
Degree									
Instructor	2.26	1.83	0.210	2.37	1.81	0.190	1.92	1.70	0.250
Associate professor	1.49	1.29	0.240	1.33	1.30	0.300	1.71	1.30	0.180
Professor	4.13	1.34	0.001	4.00	1.35	0.003	3.29	1.30	0.010
Assistant professor	*	*	*	*	*	*	*	*	*
School									
Dentistry	5.55	1.36	< 0.001	4.97	1.35	< 0.001	5.91	1.32	< 0.001
Pharmacy	3.91	1.64	0.017	4.11	1.66	0.010	5.26	1.90	< 0.001
Nursing	2.48	1.90	0.190	2.40	1.90	0.200	2.39	1.98	0.220
Paramedical	-.14	3.25	0.960	0.10	3.26	0.970	1.93	2.96	0.510
Management and information	2.21	2.23	0.320	2.29	2.06	0.260	4.51	1.77	0.620
Health	0.76	1.96	0.690	0.77	1.92	0.680	4.03	1.50	< 0.001
Medical	*	*	*	*	*	*	*	*	*
Gender									
Woman	0.38	0.98	0.690	0.30	0.98	0.750	0.59	1.00	0.540
Men	*	*	*	*	*	*	*	*	*
Work experience(year)	-0.07	0.07	0.280	-0.04	0.07	0.480	0.00	0.06	0.970
QIC		389035.9			389677.0			400783.2	

*Reference group. GEE: Generalized estimating equations, SE: Standard error, QIC: Quasi-likelihood under the independence model criterion

Similar results reported in Arak University where a score of instructors was higher than assistant professors (8).

We observed that mean evaluation score for women was slightly higher than male which was not significant. This was similar to results of Arak University and Ahvaz Jondi-Shapour University (6, 8). However, at the Texas University, men score were more than women (19).

In this study, the evaluation score for the pharmacy and dentistry schools was significantly greater than medical college. Evaluation score of members of paramedical school was smaller than the medical school which was not statistically significant. An opposite finding reported from Arak University, in which score of staff of medical school was less than paramedical school (8). In a previous study conducted in Kerman University during 2001 to 2006, no difference between schools was seen (20).

In our study, work experience did not reach significance level. Similar results were reported in Jondi-Shapour University (16, 20). However, results of Shahid Bahonar University showed a positive correlation between work experience and evaluation score (17).

None of the similar studies implemented across Iranian universities justified their findings. Therefore, comparison of the differences between our results with others is difficult.

The main aspect of our work was to address the impact of the selection of correlation structure on GEE results. It has been claimed that even when the correlation structure is incorrectly selected, GEE provides unbiased estimates of the coefficients of the independent variables (9, 21). However, our simulation study revealed that the choice of correlation structure change affects the subset of independent variables that have a significant impact on the response variable (11).

In Wisconsin Epidemiologic Study of Diabetic Retinopathy, influence of 13 risk factors on a binary response variable (diabetic retinopathy in each of the two eyes was assessed. Since observations of two eyes tend to be correlated, the GEEs with three correlation structures (independence, autoregressive, and exchangeable) are used**.** Simulation studies suggested that the independent, exchangeable, and autoregressive models were 48.8%, 44.4%, and 45%, respectively, likely to select the right predictors (22).

In another simulation study, considering the binary and Gaussian response, selection probability of correct subset from independent variables was different applying different correlation structures. For Gaussian response, independence, exchangeable, and autoregressive structures were 76.3%, 74.8%, and 73.5%, respectively, likely to detect correct independent variables (23).

One of the main limitations of our work is that we only conducted one small simulation study under fixed conditions. Other issues that might influence the performance of models are sample size, number, and degree of correlation between variables and distribution of outcome. Therefore, detailed simulation studies should be designed to address the impact of all these factors together.

Conclusion

In the GEE, selecting the correct correlation structure is important, and it affects the significance of the independent variables. In addition, education deputy should consider the factors that influence the evaluation score for efficiently planning.

Acknowledgments

We would like to thank Education and Development Office of KUMS who facilitated our access to the data.

References

1. Resifar A, Khaghanizadeh M, Ebadi A, Masoumeh M. An overview of the current evaluation methods in universities. With emphasis on the evaluation of teachers by students. Iranian Bimonthly of Education Strategies in Medical Sciences 2008; 1(1): 10-8. [In Persian].
2. Rajabi S, Papzan A. Combination of

qualitative and quantitative research methods used to design a tool for evaluation of faculty members' performance case study: agricultural college, Razi University of Kermanshah province. Higher Education 2010; 3(1): 151-69. [In Persian].

3. Siamian H, Bala Ghafari A, Aligolbandi K, Rezanezhad SF, Sharifi Nick M, Shahrabi A, et al. Characteristics of a good university lecturer according to students. J Mazandaran Univ Med Sci 2013; 22(96): 106-13. [In Persian].
4. Aliasgharpour M, Monjamed Z, Bahrani N. Factors affecting students' evaluation of teachers: comparing viewpoints of teachers and students. Iran J Med Educ 2010; 10(2): 186-95. [In Persian].
5. Mianroodi AA, Asgari Baravat Z, Khanjani N. Explaining the pros and cons of different sources of faculty evaluation from the viewpoints of Medical University Academics. Strides Dev Med Educ 2012; 9(1): 65-77. [In Persian].
6. Shakurnia A, Motlagh ME, Malayeri AR, Jouhanmardi A, Komaili Sani H. Students' opinion on factors affecting faculty evaluation in Jondishapour Medical University. Iran J Med Educ 2005; 5(2): 101-10. [In Persian].
7. Carle AC. Evaluating College Students' Evaluations of a Professor's Teaching Effectiveness across Time and Instruction Mode (Online vs. Face-to-Face) Using a Multilevel Growth Modeling Approach. Computers & Education 2009; 53(2): 429-35.
8. Rafiei M, Mosayebi G. Results of six years professors' evaluation in Arak University of Medical Sciences. J Arak Univ Med Sci 2010; 12(4): 52-62. [In Persian].
9. Nakai M, Ke W. Statistical models for longitudinal data analysis. Applied Mathematical Sciences, 2009; 3(40): 1979-89.
10. Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide. 2nd ed. Cambridge, UK: Cambridge University Press; 2013.
11. Cui J, Qian G. Selection of working correlation structure and best model in gee analyses of longitudinal data. Communications in Statistics - Simulation and Computation 2007; 36(5): 987-96.
12. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. Statist Med 1998; 17: 1261-91.
13. Ghisletta P, Spini D. An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. Journal of Educational and Behavioral Statistics 2004; 29(4): 421-37.
14. Hin LY, Wang YG. Working-correlation-structure identification in generalized estimating equations. Stat Med 2009; 28(4): 642-58.
15. Jang MJ. Working correlation selection in generalized estimating equations [Thesis]. Iowa, IA: University of Iowa; 2011.
16. Shakournia A, Elhampour H, Mozafari A, Dashtbozorgi B. Ten year trends in faculty members' evaluation results in Jondishapour University of Medical Sciences. Iran J Med Educ 2008; 7(2): 309-16. [In Persian].
17. Lesani M, Mohamadi V. Assess trend of teaching quality for faculty from the mathematics and computer school of Shahid Bahonar University of Kerman. Proceedings of the 5th Conference on Quality Assessment in University Systems; 2011 Apr 27; Tehran, Iran. [In Persian].
18. Ghafourian Boroujerdnia M, Shakurnia AH, Elhampour H. The opinions of academic members of Ahvaz University of Medical Sciences about the effective factors on their

evaluation score variations. Strides Dev Med Educ 2006; 3(1): 19-25. [In Persian].

19. McPherson MA, Jewell RT, Kim M. What determines student evaluation scores? A random effects analysis of undergraduate economics classes. Eastern Economic Journal 2009; 35(1): 37-51.
20. Fattahi Z, Mousapour N, Haghdoost A. The trend of alterations in the quality of educational performance in faculty members of Kerman University of Medical Sciences. Strides Dev Med Educ 2005; 2(2): 63-71. [In Persian].
21. Ziegler A. Generalized estimating equations. New York, NY: Springer Science & Business Media; 2011.
22. Pan W. Akaikes information criterion in generalized estimating equations [Online]. [cited 2000]; Available from: URL: http://sph.umn.edu/faculty1/wp-content/uploads/2012/11/rr2000-013.pdf
23. Wang YG, Hin LY. Modeling strategies in longitudinal data analysis: covariate, variance function and correlation structure selection. Computational Statistics & Data Analysis 2010; 54(12): 3359-70.