

Table 2 Standardized coefficients of the predictors derived from multiple logistic regression and the assigned score for each predictor

Predictor	Prevalence (%)	Recurrent ischemic stroke		Cardiovascular events		Score
		Standardized coefficient	P value	Standardized coefficient	P value	
Age, 65–75 years	41.5	0.277	0.0001	0.280	<0.0001	1
Age, >75 years	26.7	0.273	0.0001	0.278	<0.0001	2
Hypertension	75.5	-0.043	0.40	-0.036	0.47	1
Diabetes mellitus	26.4	0.149	0.0018	0.149	0.0011	1
Previous myocardial infarction(MI)	2.6	0.017	0.68	0.055	0.11	1
Other cardiovascular diseases except MI/AF	2.7	-0.022	0.67	-0.002	0.96	1
Peripheral artery disease*	0.6	-	-	-	-	1
Smoker	22.2	0.087	0.097	0.072	0.16	1
Previous stroke or TIA†	100	-	-	-	-	1
Stroke subtype except SAO	55.2	0.065	0.22	0.084	0.10	1
Waist circumference, ≥90 cm	27.5	0.145	0.0024	0.138	0.0026	1
Male gender	66.9	0.067	0.25	0.051	0.35	1

*Peripheral artery disease was excluded as no events occurred in patients with peripheral artery disease.

†The predictor of stroke/TIA was excluded as all patients had previous stroke.

AF, atrial fibrillation; TIA, transient ischemic attack; SAO, small artery occlusion.

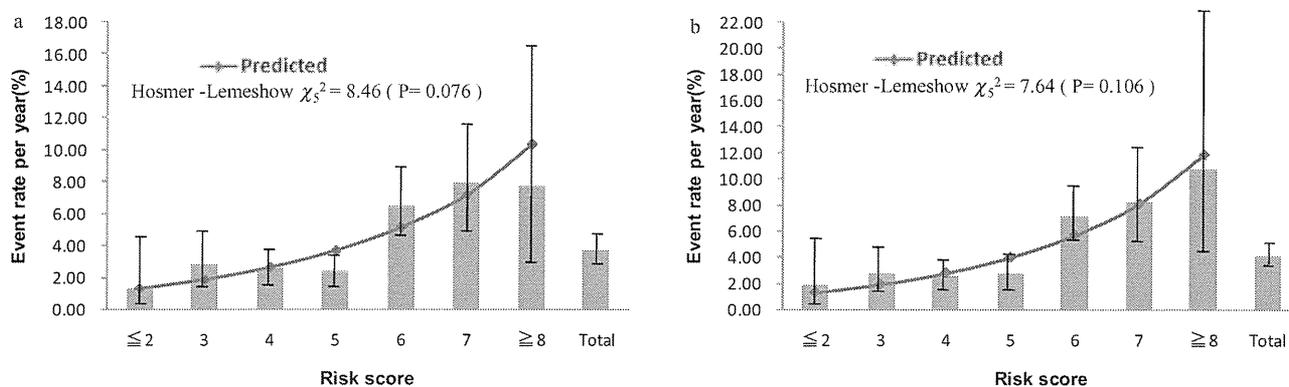


Fig. 1 Calibration between observed and predicted event rates per year. Bar chart represents the observed event rate. (a) Recurrent ischemic stroke rates by risk score. (b) Cardiovascular event rates by risk score.

Figure 1 shows the calibration curves and performance of the modified ESRS for both recurrent ischemic stroke and cardiovascular events. Predicted event rates were reasonably close to the corresponding observed event rates. Goodness-of-fit was confirmed using the Hosmer–Lemeshow test.

Figure 2 shows ROC curves comparing the predictive ability of the modified and unmodified ESRS. The modified version appeared to be superior to the original version ($P=0.094$ for recurrent ischemic stroke, $P=0.080$ for cardiovascular events). The c-statistic for recurrent ischemic stroke was 0.632 [95% confidence interval (CI): 0.579 to 0.684], and the c-statistic for cardiovascular events was 0.640 (95% CI: 0.590 to 0.689) in the modified ESRS.

Figure 3 shows the ability of dichotomized risk scores to discriminate between high and low recurrence rates. The modified ESRS (dichotomized between 5 and 6) seemed to be better able to discriminate between low and high rates of

recurrence for both ischemic stroke and cardiovascular events than the original ESRS.

Figure 4 shows the ability of the modified ESRS to predict recurrent ischemic stroke and cardiovascular events in men and women. In women, the cutoff was between 4 and 5 points, because it was 1 point less in women than in men. Good discriminatory ability was confirmed for men and women because the c-statistics were similarly high regardless of gender. Furthermore, the modified ESRS was highly predictive for both genders.

Discussion

Clinical prediction modeling is a useful method to assess risk of stroke. Although a well-known tool for that purpose, the ESRS may not have sufficient discriminatory ability. For instance, the c-statistic for the ESRS was 0.60 using REACH registry data and 0.61 using SCALA data

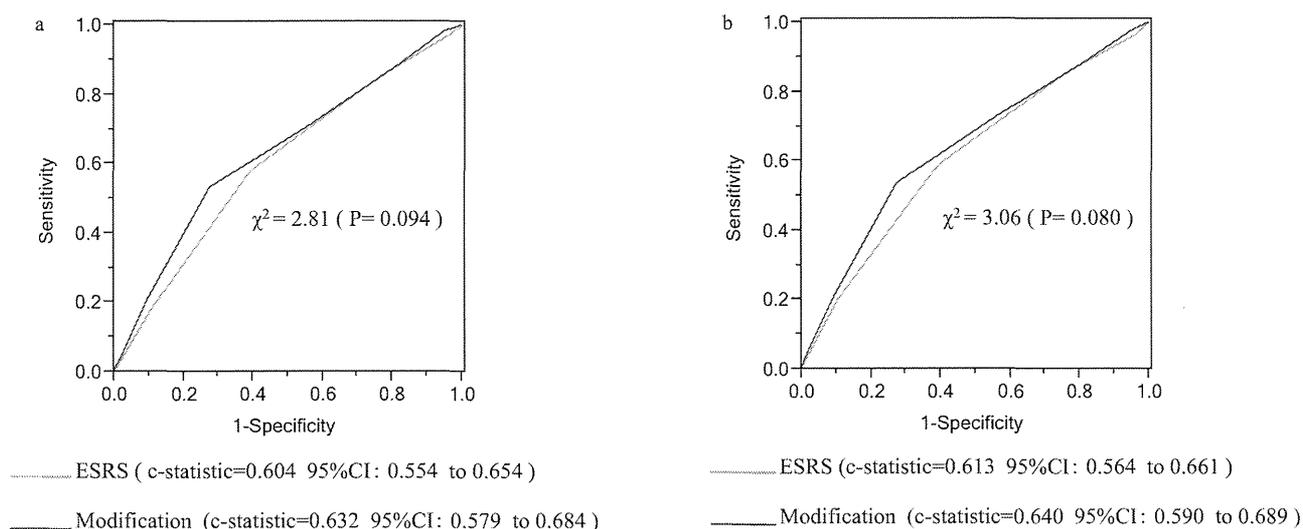


Fig. 2 Receiver operating curves comparing the predictive ability of the Essen stroke risk score (ESRS) and its modification. (a) Recurrent ischemic stroke. (b) Cardiovascular events.

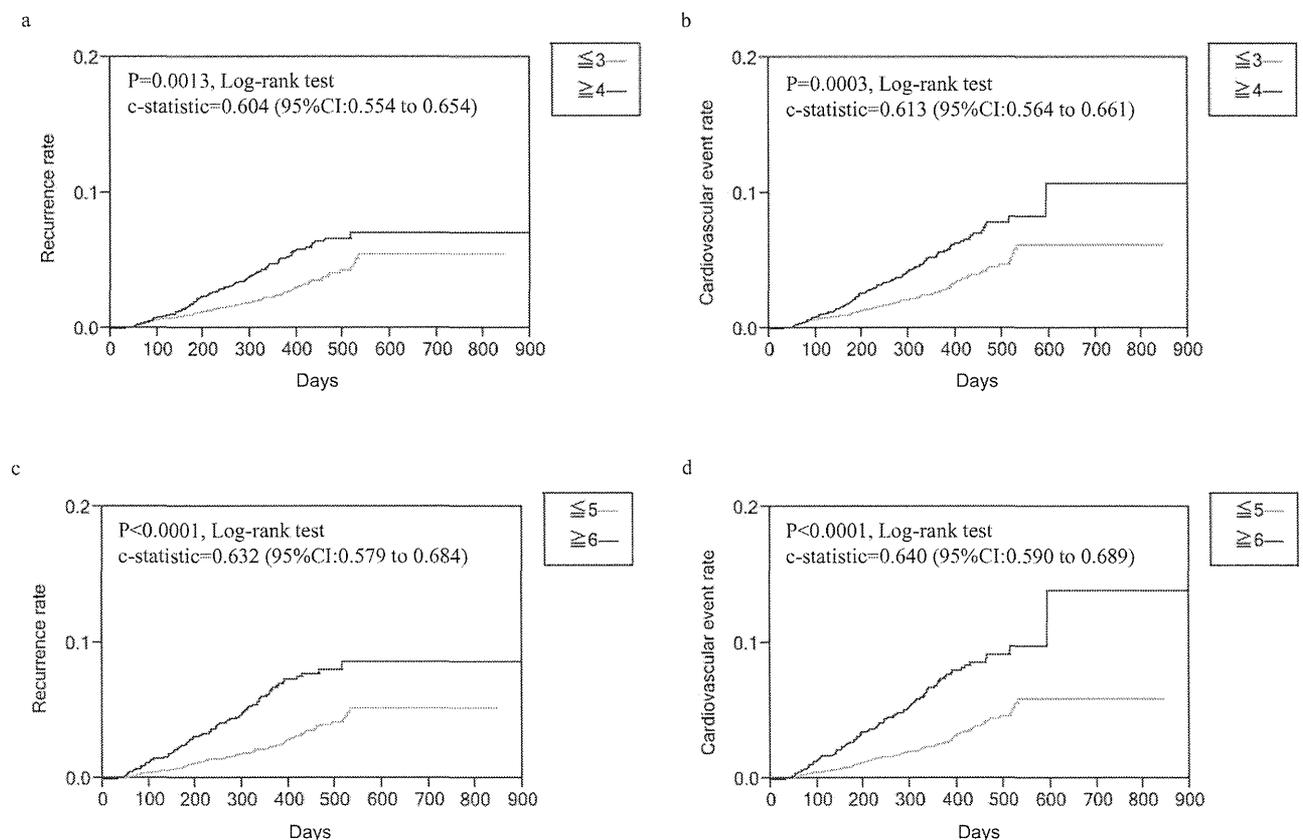


Fig. 3 Kaplan-Meier curves comparing the predictive ability of the Essen stroke risk score (ESRS) and its modification. (a) Predictability of recurrent ischemic stroke by the ESRS. (b) Predictability of cardiovascular events by the ESRS. (c) Predictability of recurrent ischemic stroke by the modification. (d) Predictability of cardiovascular events the modification.

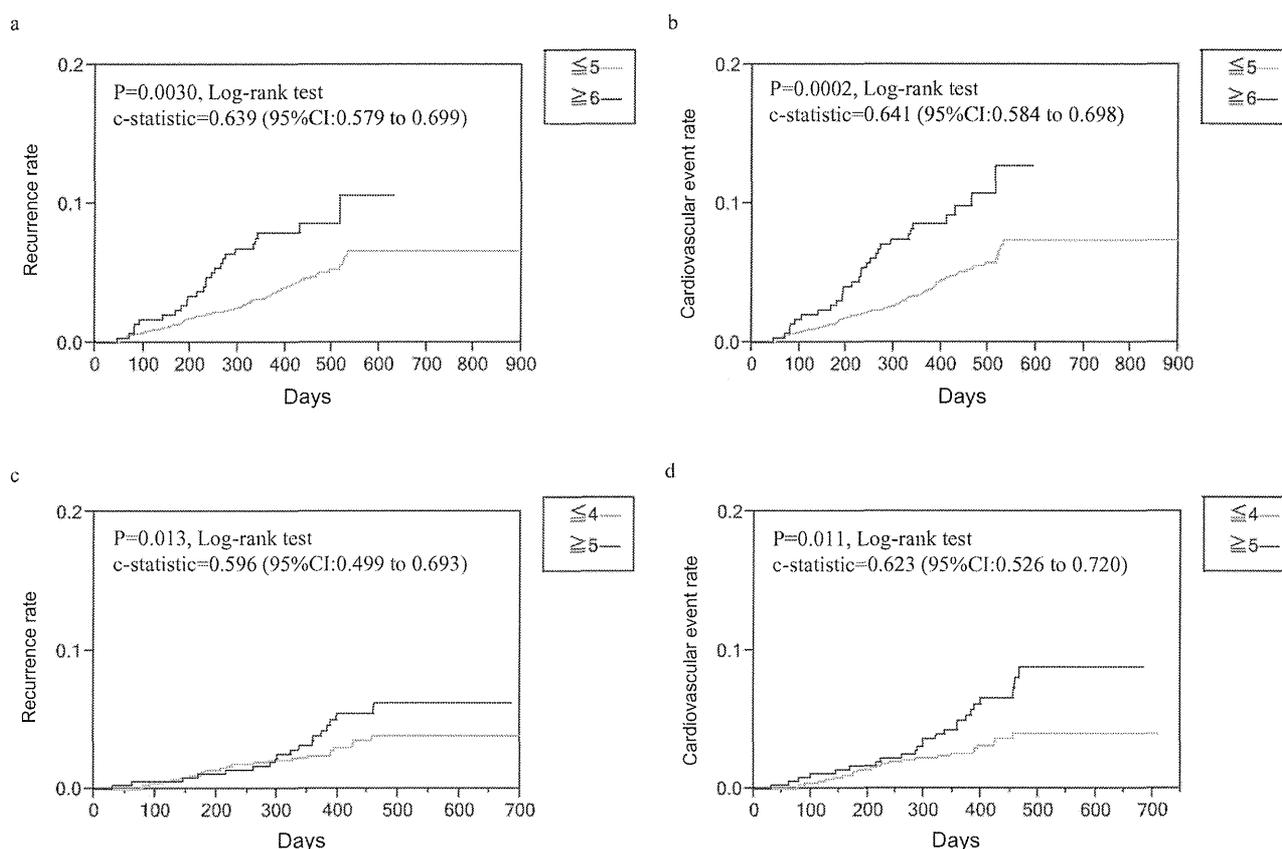


Fig. 4 The ability of the modified Essen stroke risk score to predict recurrent ischemic stroke and cardiovascular events by gender. (a) Men for recurrent ischemic stroke. (b) Men for cardiovascular events. (c) Women for recurrent ischemic stroke. (d) Women for cardiovascular events.

(10,11). We modified the ESRS by adding waist circumference (a measure of abdominal adiposity), stroke subtype [defined by the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification] (18), and gender. The modification improved discriminatory ability (c-statistic of 0.63 or 0.64 from ROC curves) and predictive ability (Kaplan–Meier curves).

Should waist circumference be included in the risk score? National and international guidelines have differed about the value of using clinical measures of adiposity to predict cardiovascular disease risk in primary prevention. The World Health Organization (19) and the US National Heart, Lung, and Blood Institute (20) recommended assessment of waist circumference although it was omitted from several risk scores such as the Framingham risk score (3). In a recent report, hazard ratios for developing ischemic stroke and coronary heart disease increased 1.25 and 1.32, respectively, for each 12.6cm increase in waist circumference after adjustment for age, gender, and smoking status (21). Another large case-control study also found that the waist-to-hip ratio could be important in prediction of ischemic stroke (13). Univariate analysis of the EVEREST data also suggested waist circumference as a possible risk.

The waist circumference cutoff points for risk assessment might be another issue. We selected 90 cm as the cutoff in both

men and women. The cutoff is 85 cm for men and 90 cm for women according to national guidelines defining metabolic syndrome in Japan (22) and 90 cm for men and 80 cm for women according to the International Diabetes Federation (23).

Small artery disease (lacunar infarct), atherothrombotic stroke, and stroke of undetermined etiology occurred in 45%, 10%, and 35%, respectively, of the EVEREST population. This is because the TOAST classification system (18) was used to define stroke subtypes. Accordingly, atherothrombotic stroke was defined only if stenosis was apparent in the brain. Further from univariate analysis, patients with small artery disease had a significantly lower rate of recurrent ischemic stroke. Stroke subtypes were considered to be good predictors.

Although we added a point for PAD to the ESRS, the importance of PAD could not be confirmed in the analysis because the prevalence of PAD was unexpectedly low (i.e. 0.6%) and no events occurred in patients with PAD. Peripheral artery disease has only recently begun to be recognized in Japan, so the diagnosis of PAD may still be ambiguous.

This study has several limitations. First, the number of events was not sufficiently large because only 121 events of recurrent ischemic stroke occurred in 3292 patients. Second, although a Cox proportional hazards model should have been used for assessing the discriminatory and predictive ability,

multiple logistic regression was applied in this study because all participants were followed up for one-year that was quite homogeneous. Third, we did not confirm the predictive ability using a validation cohort again. However, the EVEREST data were externally validated using a cross-validation method. Random samples each consisting of 80% of all data were selected and the cross-validation method was repeated 100 times. Distribution of 100 different c-statistics ranged from 0.58 to 0.70 with a mean of 0.63, which was almost the same as the c-statistic derived from the whole data set. Although we are confident that our result has been validated externally, we expect that an external validation using other data will be conducted in the future.

Recently, a similar and important INSIGHT registry (24) was published where the rate recurrent ischemic stroke was 6.2% per year (reported 6.7% with a mean follow-up of 13 months), as compared to 3.7% per year in our EVEREST registry (14). Our accrual criterion was less than six-months after the index event; however, there were no data available about the gap between the index event and accrual. On the other hand, a mean duration after the index event was 0.9 month in the INSIGHT. Therefore, it is considered that the INSIGHT accrued patients at more acute phase than the EVEREST.

In conclusion, the modified risk score including points for waist circumference, stroke subtype by etiology, and gender in addition to the variables used in the original ESRS improved the ability to predict recurrent cardiovascular events within one-year.

Patients with a high modified ESRS score should be candidates for intensified secondary prevention strategies.

Acknowledgements

We would like to thank all the physicians who participated in the EVEREST registry. We also thank all the patients involved in the registry.

References

- Hata J, Tanizaki Y, Kiyohara Y *et al.* Ten-year recurrence after first-ever stroke in a Japanese community: the Hisayama study. *J Neurol Neurosurg Psychiatry* 2005; **76**:369–72.
- Uchiyama S, Goto S, Matsumoto M *et al.* Cardiovascular event rates in patients with cerebrovascular disease and atherosclerosis at other vascular locations: results from 1-year outcomes in the Japanese REACH registry. *J Neurol Sci* 2009; **287**:45–51.
- D'Agostino RB, Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *JAMA* 2001; **286**:180–7.
- Gage BF, Waterman AD, Shannon W, Boehler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National registry of atrial fibrillation. *JAMA* 2001; **285**:2864–70.
- van Walraven C, Hart RG, Wells GA *et al.* A clinical prediction rule to identify patients with atrial fibrillation and a low risk for stroke while taking aspirin. *Arch Intern Med* 2003; **163**:936–43.
- Rothwell PM, Gilles MF, Flossmann E *et al.* A simple score (ABCD) to identify individuals at high early risk of stroke after transient ischaemic attack. *Lancet* 2005; **366**:29–36.
- Johnston SC, Rothell PM, Nguyen-Huynh MN *et al.* Validation and refinement of scores to predict very early stroke risk after transient ischaemic attack. *Lancet* 2007; **369**:283–92.
- Easton JD, Saver JL, Albers GW *et al.* Definition and evaluation of transient ischemic attack. *Stroke* 2009; **40**:2276–93.
- CAPRIE Steering Committee. A randomized, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* 1996; **348**:1329–39.
- Weimar C, Diener H-C, Alberts MJ *et al.* The Essen stroke risk score predicts recurrent cardiovascular events: a validation within the REduction of Atherothrombosis for Continued Health (REACH) registry. *Stroke* 2009; **40**:350–4.
- Weimar C, Goetler M, Rother J *et al.* Predictive value of the Essen stroke risk score and ankle brachial index in acute ischaemic stroke patients from 85 German stroke units. *J Neurol Neurosurg Psychiatry* 2008; **79**:1339–43.
- Kernan WN, Viscoli CM, Brass LM *et al.* The stroke prognosis instrument II (SPI-II): a clinical prediction instrument for patients with transient ischemia and nondisabling ischemic stroke. *Stroke* 2000; **31**:456–62.
- O'Donnell MJ, Xavier D, Liu L *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010; **376**:112–23.
- Suzuki N, Houkin K, Terayama Y *et al.* One-year atherosclerotic vascular events rates in outpatients with recent non-cardioembolic ischemic stroke: the EVEREST (Effective Vascular Event REduction after Stroke) registry. *J Stroke Cerebrovasc Dis* 2012; **21**:245–53.
- Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; **115**:928–35.
- Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statist Med* 1996; **15**:361–87.
- Hosmer DW, Lemeshow SA. Goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics Series A* 1980; **10**:1043–69.
- Adams HP Jr, Bendixen BH, Kappelle LJ *et al.* Classification of subtype of acute ischemic stroke: definition for use in a multicenter clinical trial. *Stroke* 1993; **24**:35–41.
- WHO Consultation of Obesity. Presenting and Managing the Global Epidemic. Geneva, World Health Organization, 2000.
- National Institutes of Health. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—the evidence report. *Obes Res* 1998; **6**(Suppl.):51–209S.
- The Emerging Risk Factors Collaboration. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *Lancet* 2011; **377**:1085–95.
- Matsuzawa Y, Ikeda Y, Katayama S *et al.* Working committee of determining diagnostic criteria for metabolic syndrome. Definition and diagnostic criteria of metabolic syndrome. *Jap J Intern Med* 2005; **94**:794–809. (in Japanese).
- International Diabetes Federation. The IDF Consensus Worldwide Definition of the Metabolic Syndrome. Brussels, IDF, 2006.
- Weimar C, Siebler M, Brandt T *et al.* Vascular risk prediction in ischemic stroke patients undergoing in-patient rehabilitation – insights from the investigation of patients with ischemic stroke in neurologic rehabilitation (INSIGHT) registry. *Int J Stroke* 2012; doi: 10.1111/j.177-4949.2011.00752.x.

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

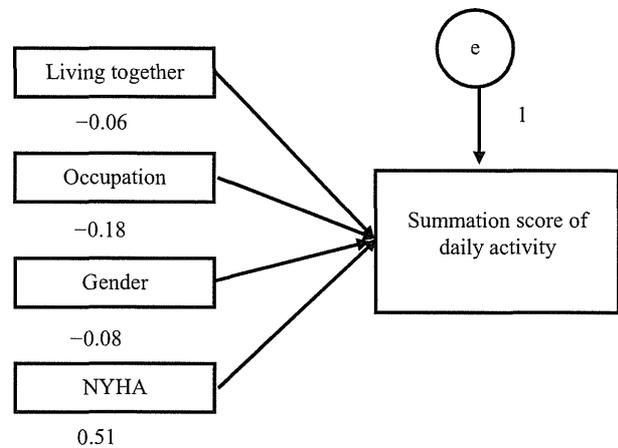


Figure 1. The association between summation score of daily activity and the four factors.

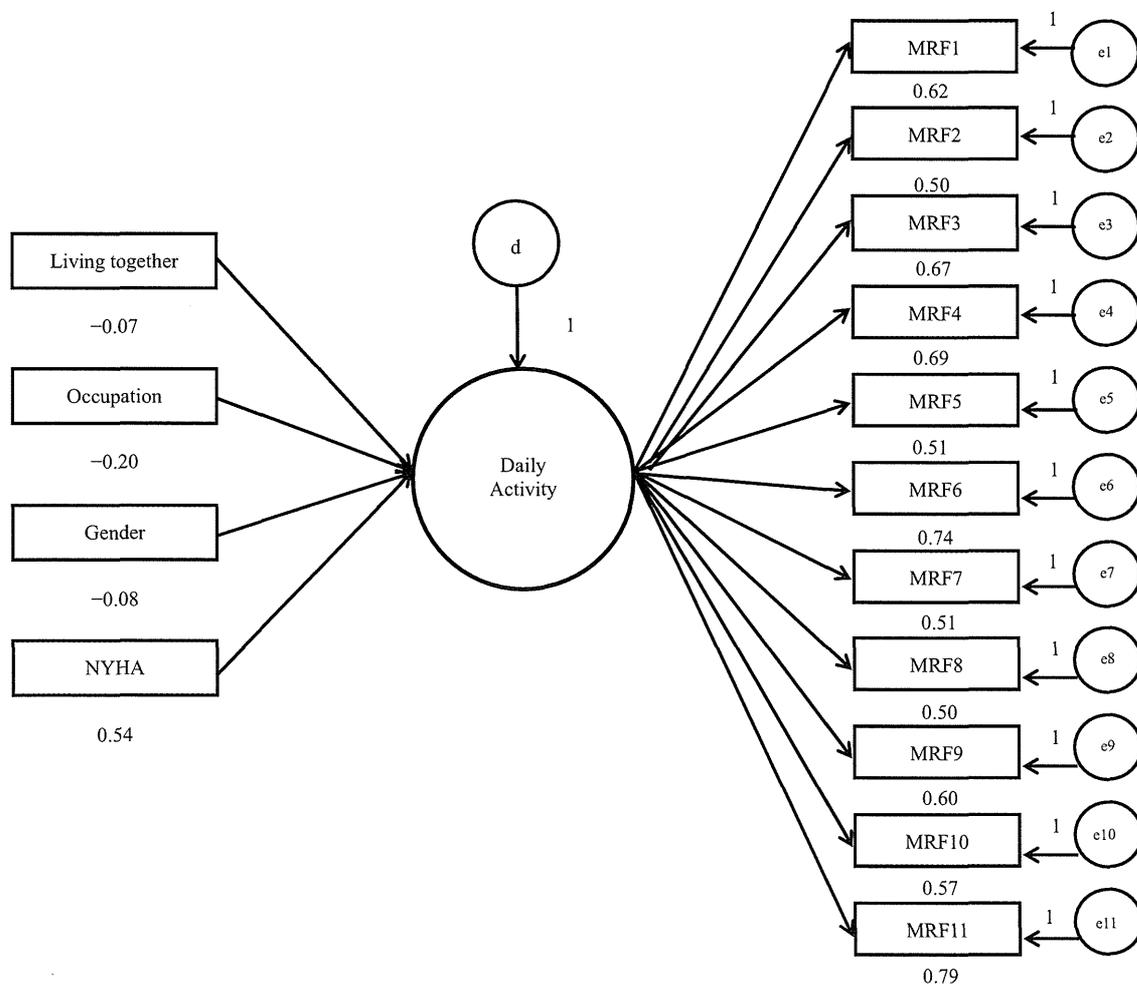


Figure 2. The association between daily activity and the four factors.

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:

$$\begin{aligned} Y_{\text{MRF1}} &= \gamma_1 F + e_1 \\ Y_{\text{MRF2}} &= \gamma_2 F + e_2 \\ Y_{\text{MRF3}} &= \gamma_3 F + e_3 \\ Y_{\text{MRF4}} &= \gamma_4 F + e_4 \\ Y_{\text{MRF5}} &= \gamma_5 F + e_5 \\ Y_{\text{MRF6}} &= \gamma_6 F + e_6 \\ Y_{\text{MRF7}} &= \gamma_7 F + e_7 \\ Y_{\text{MRF8}} &= \gamma_8 F + e_8 \\ Y_{\text{MRF9}} &= \gamma_9 F + e_9 \\ Y_{\text{MRF10}} &= \gamma_{10} F + e_{10} \\ Y_{\text{MRF11}} &= \gamma_{11} F + e_{11} \end{aligned}$$

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

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 antiretroviral 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>

漢方と教育研修

日本社会事業大学大学院 福祉マネジメント研究科 教授 鶴岡浩樹

POINT

- ① 漢方薬の処方、近年、日常的な診療行為となっている
- ② 漢方薬のRCTを集めたエビデンスレポートが、学会ホームページより無料で閲覧できる
- ③ 漢方教育は著しく発展しているが、プライマリ・ケア向けの仕組みが十分ではない
- ④ プライマリ・ケア医にとって統合医療とは生態学的接近(五十嵐の10の軸)にほかならない

特集

はじめに

2011年に実施された漢方薬処方に関する全国調査によれば、医師の約9割(89.0%)が漢方薬を処方していることが分かった¹⁾。漢方薬を処方することは今や日常的な診療行為である。地域医療の現場で使用する家庭医も多いのではないだろうか。漢方は伝統中国医学に端を発し、日本の風土や気候に適した様式に改良され、日本人の生活に寄り添いながら発展した、日本古来の伝統医学であり、わが国においては、医療制度の上でも保険適用が認可されている。漢方を補完代替医療(complementary and alternative medicine: CAM)と扱うことに抵抗する専門家が多いこともうなずける。一方で、世界の趨勢をみると漢方はCAMの一つと位置付けられている。現代西洋医学とは明らかに概念も理論体系も異なる。

Evidence-based medicine(EBM)や医学教育の流れを語る上では、CAMの動向を無視して、漢方を論じることができない。本稿では漢方をCAMとして考え、CAMのエビデンスと医学教育という視点からまとめた。

CAMとは

CAMとは、現代西洋医学以外の医療の総称で、漢方、鍼灸などの東洋医学から民間療法、健康食品、祈祷まで幅広い領域を示す。世界共通の定義は難しく、類義語も複数あるため混乱を招いている。代替医療(alternative medicine)という用語は、代替技術(alternative technology)の一つとして1990年代前半まで米国で流行したが、反体制的な意味を含み、この用語を単独で使用することは少なくなった。相補医療/補完医療(complementary

medicine)はヨーロッパで最も親しまれている用語である。文献検索で最もヒットするタームだ。そのほかにも非西洋医学(non-western medicine)、非通常医療(unconventional medicine)、非正統医療(unorthodox medicine)、伝統医学(traditional medicine)、民間療法(folk medicine)、ホリスティック医学(holistic medicine)などがある。学術的には相補代替医療／補完代替医療(complementary and alternative medicine)が一般的で、CAMと記し「カム」もしくは「キャム」と呼ぶ²⁾。

世界の動向

CAMを評価しようとする試みは1970年代の世界保健機関(World Health Organization: WHO)にさかのぼる。発展途上国への医療支援をする中で、健康観や文化観が異なる住民に、現代医学を受入れてもらうことの難しさに直面した。世界人口の8割が伝統医学治療者の恩恵を受けていたという当時の状況もあった。コスト・時間・マンパワーなど、支援に至るまでの労力は想像を超え、現地の伝統医学治療者の協力なしでは成就し得ないことに思い知らされる³⁾。WHOは伝統医学の研究に乗り出し、数々のガイドラインを刊行した⁴⁾。これらの仕事は今振り返っても価値のある内容だ。

CAMが世界的な脚光を浴びるようになったのは1990年代である。1992年に米国国立保健衛生研究所(National Institutes of Health: NIH)が代替医療局(Office of Alternative Medicine: OAM)を設立したこと、1993年にEisenbergのCAMの全米利用率調査の結果がNew England Journal of Medicine誌に報告されることが、1990年代のCAMブームのきっかけとなった。この調査では米国成人の34%(6,000万人)がCAMを利用しているというものであった。この結果を少ないとする声もあるが、家庭医よりCAM治療者への受診者が多いことや、CAMの利用を主治医に伝えない患者が7割以上という結果に、世界の医療従事者が驚かされた⁵⁾。これを機に先進各国で利用率調査がなされ、世界規模でCAMへの関心が高まっていることが分かった。それまでCAMを排斥しようと

していた医学界は、CAMを評価しようと180度方向を変える。1998年にはOAMがNational Center for Complementary and Alternative Medicine(NCCAM)に昇格し、豊富な資金をベースにその活動が加速した。同年、米国医師会雑誌(Journal of American Medical Association: JAMA)がCAM特集号を刊行し世界的な反響を呼んだ。この特集号には、米国の利用者が増え続けているというEisenbergの追跡調査に加え、CAMのランダム化比較試験(randomized controlled trial: RCT)が多数報告され、CAM領域におけるEBMの重要性が認識された⁶⁾。1999年には英国医学会雑誌(British Medical Journal: BMJ)が「相補医療のABC」と題した連載を行い⁷⁾、CAMの現状とエビデンスを世界の医師に伝える重要な役割を果たした。こうして一般社会においても医療界においてもCAMは無視できない存在となった。欧米では東洋医学をCAMの代表的療法として捉えており、漢方や鍼治療への認識は高い。

21世紀になると、現代医学とCAMを統合しようとする試みが始まった。これを統合医療(integrative medicine)と呼び、EBMをモノサシに統合していく方向性が見えてきた。

漢方のエビデンス

EBMのうまれた1990年代初頭、CAMにEBMを適用することはナンセンスと思われていた。エビデンスのないものがCAMという考え方が支配していた⁸⁾。1990年代の爆発的なCAMブームの中で、CAMも医薬品と同様にRCTで評価しようとする動きが活発化する。この時代の流れは、漢方医学にも大きな影響を与えた。

漢方医学を牽引する日本東洋医学会では、2001年にEBM特別委員会が設置された。当時としては画期的な活動で、秋葉哲生委員長を中心に、1986年から2002年の漢方関係の論文が整理された。この作業により漢方におけるエビデンスの基盤が形成された。2005年からはEBMの第一人者、津谷喜一郎が委員長となり、第2期の作業が始まった。網羅性と透明性を高めるため、システマティック・レビューに準じた作業方針がとられ

表1 構造化抄録(SA)の項目

1	目的
2	研究デザイン
3	セッティング
4	参加者
5	介入
6	主なアウトカム評価項目
7	主な結果
8	結論
9	漢方的考察
10	論文中の安全性の評価
11	Abstractorのコメント
12	Abstractorの名前と作成日

た。収集する臨床試験はRCTに限定し、世界的な標準に基づく構造化抄録(structured abstract : SA)集を作成することとした。著者も委員(Abstractor)として2005年から2013年までSAを作成する作業に加わった。世界的な標準とはAltmanの8項目で、1)目的、2)研究デザイン、3)セッティング、4)参加者、5)介入、6)主なアウトカム評価項目、7)主な結果、8)結論、から成る⁹⁾。JAMAやBMJのほか、EBMの2次情報誌として知られるEvidence-Based Medicine、ACP Journal Clubで使用されている項目と同じだ。EBM実践者にとっては吟味しやすい項目といえる。これに同学会独自の4項目、9)漢方的考察、10)論文中の安全性の評価、11)Abstractorのコメント、12)Abstractor名と作成日、を加えた。すなわち「漢方薬のエビデンスレポート」と称されるこのSAは、表1のように計12項目から構成される。RCTの収集はThe Cochrane Library (CENTRAL)、医学中央雑誌(医中誌web)、日本漢方生薬製剤協会提供のデータベースより検索された。

2010年までに345件のRCTと1件のメタアナリシスが収集され、これらのSAが作成された。SAは同学会ホームページ(www.jsom.or.jp/medical/ebm)より、会員でなくても無料で閲覧できる。2011年には14件、2012年には20件のRCTが追加された。よって、現在、約380件の漢方薬のRCTがインターネット上でSAという様式で瞬時に手に入れることができる。注目すべきは、花粉症と小青竜湯、癒着性イレウスと大建中湯、アルツハイマー型認知症と抑肝散、COPDと麦門冬湯、GERDと六君子湯など、日常病に関するRCT

表2 地域医療に関連したRCTのトピック

疾患名	漢方薬
花粉症	vs 小青竜湯
癒着性イレウス	vs 大建中湯
大腸癌術後	vs 大建中湯
気管支喘息	vs 柴朴湯
COPD	vs 麦門冬湯
COPD	vs 補中益気湯
かぜ症候群後遷延性咳嗽	vs 麦門冬湯
インフルエンザ	vs 麻黄湯
アルツハイマー型認知症	vs 抑肝散
脳血管認知症	vs 釣藤散
GERD	vs 六君子湯
過敏性腸管症候群	vs 桂枝加芍薬散
更年期ホットフラッシュ	vs 桂枝茯苓散
高血圧	vs 大柴胡湯、釣藤散
高血圧	vs 黄連解毒湯
高脂血症	vs 大柴胡湯
肥満	vs 防風通聖散
脳卒中後遺症	vs 当帰芍薬散
带状疱疹後神経痛	vs 補中益気湯
リンパ浮腫	vs 牛車腎気丸
前立腺肥大症	vs 牛車腎気丸

が蓄積され始めたことである。そのほとんどが日本発のRCTであることも特記しておきたい。数はまだ少ないが、地域医療の現場で利用できそうなRCTのトピックを表2に列挙した。ここに挙げたトピックは、全てがポジティブな結果のものではない。ぜひ、同学会のホームページからSAにアクセスしていただき、意思決定の判断材料に役立ててほしい。

余談であるが、Abstractorのコメントは地域医療教育の知恵が埋め込まれている。漢方医学の領域では、複数の流派が存在する上に、EBMの批判的吟味という文化が浸透していなかった。エビデンスレポートのコメント次第ではRCTを実施しようという文化が育まれないのではないかという議論もあった。そこで地域医療の基本となるコミュニケーション・スキル教育で用いられるPNPの法則を転用した。すなわち、コメントをする際に、positive-negative-positiveの順序で記載するようにAbstractorのトレーニングを行った¹⁰⁾。

漢方薬のエビデンスレポートは英訳もされている。現在ではThe Cochrane Libraryから検索でき、そのまま日本東洋医学会のホームページに飛べるようなシステムになっている。PubMedにしてもThe Cochrane Libraryにしても、漢方薬は“Kampo”というタームで検索を行う。“Kanpo”で

特集

はないことに注意してほしい。

漢方の卒前教育・卒後教育

全国医学部におけるCAM教育の実態はどうなっているのだろうか。著者が行った全医学部の教務担当事務員に行った電話調査では、1999年に全80校中16校(20%)がCAM教育を導入していた。教育内容は漢方・鍼をはじめとする東洋医学が95%を占めた。担当教員の所属は麻酔科が最も多かった。同時期の米国の調査では、教育内容はバラエティに富み、担当教員の所属は家庭医療学などプライマリ・ケア部門が多かった。2004年に追跡調査を行ったところ、CAM教育を行う医学部は全80校中69校(86%)と急増した。担当教員の所属は69校中37校(53%)で東洋医学の専門家であった。1999年と2004年の変化を表3に記したので参照されたい¹⁰⁾。欧米と比較すると、日本のCAM教育の導入の早さは驚異的である。その背景には、2001年に公表されたモデル・コア・カリキュラムに「和漢薬を概説できる」という一文が加わったことが大きく影響している。2011年にはこの内容が「和漢薬(漢方薬)の特徴および使用の現状を概説できる」と改訂された。このように、漢方医学は必修化され、現在、日本の全医学部で教育されている。しかし、その内容についてはバラツキがあると推測する。2004年の調査ではCAM教育に費やす時間は、90分から1,000分以上と著しいバラツキを認めた。1,000分以上行っている医学部もわずか5校であった。表3に示したように、ほとんどが講義形式で、実習やbedside learning(BSL)を行っている大学は極めて少なかった。推測であるが、現在も授業の形式に大きな変化はないように思える。その根拠として、CAM教育担当者に行ったインタビュー調査について紹介したい。この調査は2004年の全国調査で積極的にCAM教育を行っていた医学部の教員を17名選出し、2005年に半構造化面接による質的研究を実施した。そこで浮かび上がった教育上の問題点は、指導者不足、研修を積むための施設不足、指導医や専門医を養成する施設の不足、学派と教育の問題、適切な教科書がない、大学に講座や臨床部門

表3 CAMの医学部教育:1999年と2004年の比較

	1999年	2004年
大学数	80	80
回答数	80(100%)	80(100%)
CAM教育実施校	16(20%)	69(86%)
教育内容		
漢方	15(93%)	69(100%)
鍼	8(50%)	21(30%)
CAM	0(0%)	14(20%)
EBM	0(0%)	7(10%)
担当教員		
専門家	3(18%)	27(53%)
薬理学	2(12%)	17(24%)
麻酔科	6(37%)	5(7%)
人類学	0(0%)	5(7%)
教育形式		
講義	16(100%)	69(100%)
実習	4(25%)	4(5%)
BSL	2(12%)	3(4%)
チュートリアル	0(0%)	4(5%)

として定着が困難。製薬会社主導による教育に懐疑的、学内での抵抗、などであった。各校の教育内容にバラツキがあるのも、マンパワー不足と環境整備が不十分であることに集約される¹²⁾。学派の問題も悩ましく、細かく分類すれば20ほどの学派が存在すると言われている。日本東洋医学会ではこのような学派を超え、漢方教育の標準化を目指し、2つの教科書を作成した。同学会学術教育委員会が編集した『学生のための漢方テキスト(南光堂、2007)』と『専門医のための漢方医学テキスト:漢方専門医研修カリキュラム準拠(南光堂、2010)』である。漢方医のエキスパートになるには、東洋医学講座があるような大学病院と関連病院で研鑽を積み、同学会の専門医を取得するという道がある。しかし、家庭医やプライマリ・ケア医にとって、このコースを選択するのは勇気がいることであろう。なぜなら、地域医療を実践するためにさまざまな知識と技術を学ばなければならない。内科以外にも、小児科、メンタルヘルス、小外科、初期救急、在宅医療、ヘルスプロモーション活動、EBM、NBM、等々。優先順位を考えると漢方医学の順位はどうしても落ちてしまう。地域医療を実践しながら、週1回の研修日を利用して、漢方医に学ぶなど研鑽のやり方が現実的かもしれない。

表4 五十嵐の10の軸

総合医療の最も重要な基盤は		この基盤と場を背景にして、 総合医療は次のことを実現する	
①近接性	無差別性 患者や問題を選ばない 精神的 良好な医師患者関係 時間的 時間外の初期救急も含めて 経済的 費用効果思考で行動	⑥質の保証	QOL(いきがい、自己実現)の維持向上を 尺度とした医療、保健、福祉の質を保証 する思考と行動ができる
②日常性	日常問題、日常病 単純な頻度ではなく、重症度、 影響度の大きい順に	⑦個別性	個別の事情に応じた思考と行動 多くの選択肢を示しつつ、 患者の自己決定の支援ができる
この基盤のもと以下の場で、 そのニーズを反映して仕事をする		⑧生態学的接近	多面的、学際的、有機的、総合的な 思考と行動ができる
③全人	生物医学的 視点と平行して 心理的 社会的 倫理的 視点からも思考と行動ができる	これらを実現するためには、 以下の役割と責任が必要である	
④家庭	家庭を一単位とした思考と行動ができる	⑨役割	患者の道案内役、弁護士役 患者や医療関係者の調整役、聴き役、 説明役、連絡役を担う思考と行動
⑤地域	地域を一単位とした思考と行動ができる 保健、医療、福祉を統合した地域医療を 実践する	⑩責任	継続性(当面の問題、生涯にわたる継続性) 責任制(主治医としての) 民主制(患者との対等な関係)

生態学的接近

2001年にBMJから刊行された“orthodox meets alternative”と題した統合医療特集は話題を呼んだ。21世紀の幕開けに2羽のフラミンゴが互いの首をからませた特集号の表紙は印象的であった¹³⁾。

統合医療にはいくつかの定義があるが¹⁴⁾、誌面の関係上、本稿では詳細を紹介できない。大事なことはプライマリ・ケアの概念と共通項が多いことである。例えば、統合医療は「五十嵐の10の軸」ですべて説明が可能なのである¹⁵⁾(表4)。興味のある読者は別稿を参照されたい¹⁶⁾。さて、「五十嵐の10の軸」には「生態学接近」という項目がある。「多面的、学際的、有機的、総合的な思考と行動ができる」と説明されている。現場ではさまざまな局面に遭遇し、サイエンスだけではどうにもならない状況に追い込まれる。患者の物語、家族の思い、地域の健康観などさまざまな文脈を考慮しているうちに、CAMとも出くわす。CAMに興味を持ち使いたいと思う患者の物語を、丸ごと受けとめて最良の決断をしなければならない。五十嵐のいう「質の保証」を担保し、「個別性」を実現するために、状況を把握しなければならない。そのためには多面的、学際的に勉強し、総合的な意思決定をしなければならない。すなわちこの一連の行為が「生態学的接近」であり、統合医療のプ

ロセスそのもの、と著者は考えている。

欧米では多くのプライマリ・ケア医がCAMや統合医療に興味を持っている。彼らはCAMを日常的によく遭遇するトピックと考えているからだ。おそらく彼らは現場で否応なしに「生態学的接近」を迫られるからであろう。家庭医やプライマリ・ケア医によるCAMの医学論文が多いのはそのためであろう。米国の医学部のCAM教育の担当者に家庭医が多いことにも納得がいく。米国では、アリゾナ大学のProgram in Integrative Medicine(PIM)が中心となって、Integrative Family Medicine(IFM)という研修プログラムが行われている。IFMはその名のとおり統合医療を実践する家庭医を意味するが、これは統合医療を担う西洋医学側の医師として家庭医やプライマリ・ケア医が適していることを反映している。IFMプログラムはNIHがサポートしており、アリゾナ大学をはじめ米国の複数の大学で実践されている。すでに数名の日本人医師がPIMのIFMを修了している。

おわりに

日本の漢方医学関係者には「東西統合」という狭義の統合医療を主張する専門家がいる。彼らの主張は漢方をはじめとする東洋医学と現代西洋医学の統合を目指すことである。怪しげな療法も

含まれるCAMとは一緒にしてもらいたくないという立場だ。一方、欧米型のように、漢方医学もCAMの一つと位置付け、CAM全般を抱合した広義の統合医療を主張する専門家がいる。

プライマリ・ケア医の立場からは、生態学的接近を続けながら、この論争を見守るしかない。地域医療の実践をジャズの演奏と考え、基礎となるコード進行に、漢方のエビデンスを含めてみてはどうか。地域医療の実践を、ジャズのアドリブと位置付け、アドリブの中で漢方も使ってみよう。これを実践するには、プライマリ・ケア医向けの漢方教育あるいはCAM教育の仕組みが必要である。

参考文献

- 1) 漢方薬処方実態調査2011. 日本漢方生薬製剤協会ホームページ <http://www.nikkankyo.org/aboutus/investigation/investigation01.html>
- 2) 鶴岡浩樹, 鶴岡優子: 相補代替医療(CAM)とプライマリ・ケア ①: 世界の動向と日本の現状. 日本医事新報 2002; 4102: 25-29.
- 3) WHO/R. バンナーマン, J. パートン, 陳文傑 編, 津谷喜一郎 訳: 世界伝統医学大全. 平凡社, 1991. [Bannerman R, et al: Traditional medicine and healthcare coverage. Geneva, WHO, 1984.]
- 4) 津谷喜一郎, 鶴岡浩樹: WHOガイドラインとコクラン共同計画の活動: 代替医療の合理的使用のために. In: 今西二郎 編: 別冊医学のあゆみ 代替医療のいま. 医歯薬出版. 東京, 2000. p23-31.
- 5) Eisenberg DM, Kessler RC, Foster C: Unconventional medicine in the United States. Prevalence, cost, and pattern of use. N Engl J Med 1993; 328: 246-252.
- 6) 津谷喜一郎: 「大変なことがわかった」の意味. JAMA日本語版. 1999; 7: 90.
- 7) Zollman C, Vickers A: ABC of complementary medicine. BMJ Books, London, 2000.
- 8) 津谷喜一郎, 鶴岡浩樹 訳: エビデンスに基づく医療と相補医療. 医道の日本. 1999; 58(7): 88-91. [Vickers A: Evidence-based medicine and complementary medicine. Evidence-Based Medicine 1998; 3: 168-169.]
- 9) Altman DG, Gardner MJ: More informative abstracts. Ann Inter Med 1987; 107: 790-791.
- 10) 鶴岡浩樹, 岡部哲郎, 津谷喜一郎: 漢方薬RCTの構造化抄録におけるコメント記載の改善: 第2回エビデンスレポート・タスクフォース・ワークショップ報告. 日本東洋学会雑誌 2009; 60(2): 177-184.
- 11) 鶴岡優子, 鶴岡浩樹: 全国医学部における相補代替医療教育の現状: 追跡調査1999-2004. 医学教育 2005; 36(5): 323-328.
- 12) 鶴岡浩樹: 卒前教育における代替医療の現状と今後の課題(課題番号16790303). 平成16年度~18年度科学研究費補助金(若手研究B)研究成果報告書. 自治医科大学, 2007.
- 13) Rees L, Weil A: Integrated medicine. Br Med J 2001; 322: 119-120.
- 14) 織田 聡: 日本型統合医療の理念. 日本統合医療学会誌 2012; 5(1): 12-18.
- 15) 五十嵐正統, 山田隆司: 五十嵐の10の軸, その真骨頂を垣間見る. 月刊地域医学 2007; 21(1): 4-12.
- 16) 鶴岡浩樹: プライマリ・ケアと統合医療. 日本統合医療学会誌 2008; 1(2): 28-34.
- 17) 鶴岡浩樹, 鶴岡優子: 相補代替医療の医学教育. 治療 増刊号 2007; 89(3): 709-715.

