Factors extraction for Clinical Research educational model and their Relation

Ayako OHSHIRO†*, Shinichiro UEDA†

†Department of Clinical Research and Quality Management Graduate School of Medicine University of the Ryukyus
Nishihara, 0kinawa 903-0215, Japan
*Correspondence

Summary
Clinical research is used in improving treatment methods or examining diagnostic guidelines. Different skills from clinical practice training are required for clinical research training, but such an educational model is not currently standardized. In this research, we extracted factors using morphological analysis from the survey data carried out for the clinical research education program, and obtained relevance with traditional educational model factors.

Key words:
Factor Analysis, Morphological analysis, Clinical Research Education model

1. Introduction

Appropriate clinical research training program for physicians and other medical staff is apparently necessary to conduct clinical research, which solves clinical questions occurring in daily clinical practice and is essential for high-quality care of patients. While the educational model of clinical training proposed by Kern [1] is used in the clinical training field, that of clinical research has not been standardized. Therefore, as measures to cope with research misconduct and conflicts of interest management that have occurred one after the other, guidelines such as the “International Guidelines for Ethical Review of Epidemiological Studies” [2] and “International Ethical Guidelines for Biomedical Research involving Human Subject” [3] from CIOMS (Council for International Organizations of Medical Science) have been introduced. Even in Japan, as the “Clinical Trials Act” law for clinical research and “Ethical Guidelines on research in medical science for human beings” came into force, an improvement in the quality of clinical research is required more than ever. Hence, the necessity of standardization of the educational model used in clinical research is considered to be increasing. In this research, we have generated clinical research educational programs [4] consisting of multiple courses, ranging from basic to applied clinical research, for medical personnel. The fundamental course consists of a workshop of learning procedures to organize a research plan for a simulated theme, or a hands-on seminar of learning analysis methods to research data using statistical software. The standard course consists of workshops that actually conduct clinical research for medical personnel with real clinical questions in clinical practice. The advanced course is aimed at ensuring the acceptance of a journal paper. As a next step, we tried to visualize clinical research educational program for standardization of clinical research education by using data from surveys of program students. Specifically, we defined multiple factors beforehand and requested each student to score these factors; the relationship between them was calculated by using factor analysis with many parameters. For the definition of the factors, we referred to the steps “**Ana**lysis, **Des**ign, **Dev**elopment, **Imp**lementation, **Eva**luation” of the ADDIE model [5]. As a result of the study, low values were calculated for **Dev** among 5 factors, indicating its possibility of being excluded in the generation of the clinical research education model. Furthermore, we concluded that stabilization of each factor class can be expected by adding new factors, while dependency and complexity among factors became clear. Moreover, by classifying all factors into two groups, "Before beginning" and "After starting" clinical research, the contents of guidance before and after clinical research started were indicated by factors [6][7].

In response to these results, we focused on the application of text mining to survey data of free description formulas that we have conducted for students to acquire new factors from the educational program activity data. Morphological analysis [8], which is one of the mining methods used, is a technique wherein document data is divided into words containing the minimum unit and the frequency of occurrence of each word is output; it is widely used in the feature extraction of documents in the field of language processing. It is possible to extract what is required for clinical research education from a students’ viewpoint, by mining with the free description field provided in the evaluation questionnaire of this program. Furthermore, it is expected that addition of new factors and deletion of unnecessary factors would enable improvement in accuracy and simplification of the educational model. In this paper, we report the contents of addition and deletion of newly acquired factors using text mining, and the relation, obtained by exploratory factor analysis, with dependency relationship among factors.

Manuscript received March 5, 2019
Manuscript revised March 20, 2019

2. Process of extraction for the construction of a clinical research education model

In order to extract the factors for necessary generating a clinical research education model, free description data collected from the students after the program was used. In this chapter, we describe the process.

First, by applying morphological analysis to collected data, we extract the words that students consider important for clinical research education. In this case, since words in minimum units are output, by excluding words that are not directly related to educational models such as adverbs, particle auxiliary verbs, the extraction of factors essential to the model can be expected. In addition to adding the extracted new factor to the existing factor group, and, at the same time, excluding factors which are considered unnecessary, a better practical factor selection is made. In particular, scored data regarding new factor groups is obtained from program students. Moreover, factors with low scores and factors with high factor loading are excluded by applying factor analysis. Finally, the relationship between factors in a new factor group is considered. The actual procedure is shown Fig. 1.

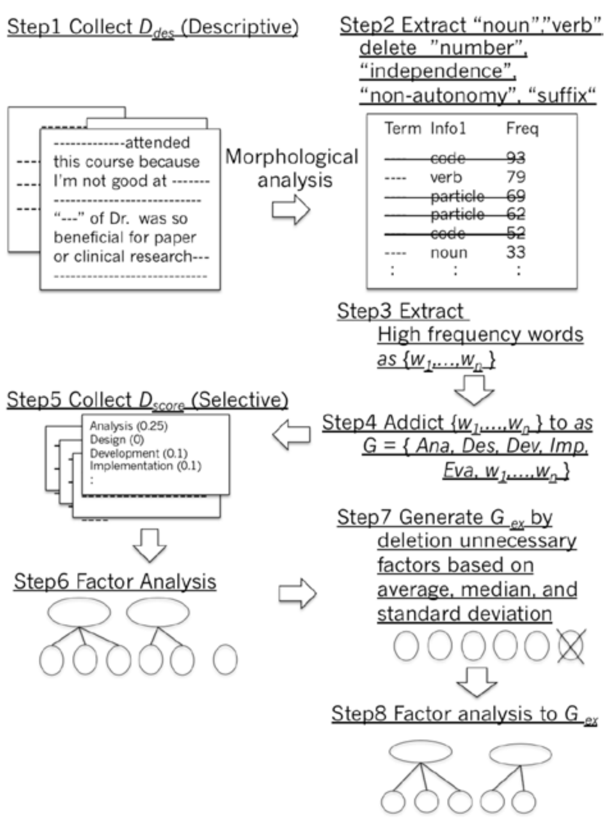

Fig. 1 Flow of extraction of clinical research education model factor and relation

Step1 Obtain free description data by evaluating students in a clinical research education programs, and define the collected data as D_{des}

Step2 To apply the result of morphological analysis to D_{des}, extract "noun" and "verb" and delete "number," "independence," "non-autonomy," and "suffix."

Step3 For the result obtained in step2, words which appear frequently are extracted and set as $\{w_1, ..., w_n\}$

Step4 Let G be a new factor group added $\{w_1, ..., w_n\}$ to {**Ana**lysis, **Des**ign, **Dev**elopment, **Imp**lementation, **Eva**luation}

Step5 Request scoring for "factors that are felt to be useful in clinical research skill acquisition" in the evaluation questionnaire, distributed to the program's students, for each factor of G, and name the collected scored data D_{score}

Step6 Apply exploratory factor analysis to D_{score} and observe the complexity of each factor and the dependency relationship with other factors

Step 7 As a result of step 6, a new factor group, defined as G_{ex}, is formed; in G_{ex}, factors with high factor loading and, in particular, small average values (A), median (M), and standard deviation (S) are deleted.

Step 8 Apply exploratory factor analysis to G_{ex} and observe the relationship between each factor.

3. Experimental Data

Details of the data actually used as D_{des} and D_{score} are shown below.

D_{des}：Descriptive data of program students from 2015 to 2017 (bearer formula N = 80)

D_{score}： Selective answer data (N = 33) of the program attendee in fiscal 2018

4. Experimental result and discussion

The results of the experiment conducted according to the "Flow of extraction of clinical research education model factor and relation" described above will be shown.

First, the results of step2 (i.e., applied morphological analysis of D_{des}) and the sorted results in descending order of appearance frequency are shown in Table 1.

Table1: Terms with high frequency appearance

Term	Info1	Frequency
Discussion	Noun	24
Group	Noun	18
Teacher	Noun	18
Research	Noun	15
Lecture	Noun	15
Presentation	Noun	11
Study	Noun	10
Attend	Noun	9
Myself	Noun	9
Benefit	Noun	8

As high frequency terms, "Discussion," "Group," "Teacher," "Research," and "Lecture,"" were the output. In some courses of the clinical research education program, there are students' presentations and discussion time to report their research progress, as well as the teacher's lecture. Therefore, it can be concluded that many opined that the lecture (introduction) by the lecturer, group discussion, and presentation were beneficial in promoting research. From the above, w_1 = Discussion, w_2= Introduction (Lecture) were selected as new factors, and we defined the new factor group G as follows.

$$G = \{Ana, Des, Dev, Imp, Eva, Dis, Int\}$$

Next, in order to exclude unnecessary factors for model generation, the features and the relationships of factors belonging to G were obtained using exploratory factor analysis and statistical value. For application of exploratory factor analysis, the number of common factors and their eigenvalues regarding D_{score} are shown in Fig. 2. The horizontal axis shows the number of common factors and the vertical axis shows eigenvalues.

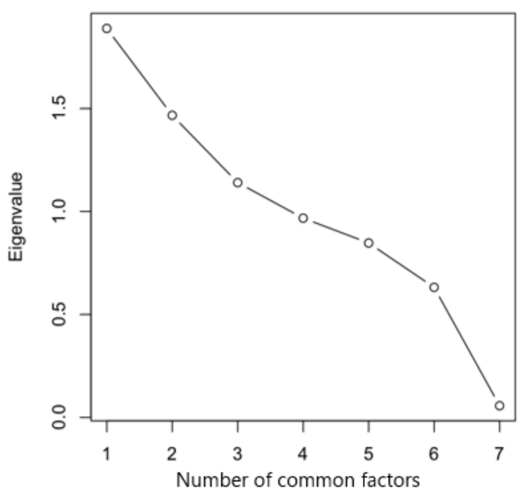

Fig. 2 Common factors and eigenvalues

It shows that the eigenvalue attenuation has a slow curve, and it is difficult to clearly determine the number of common factors. So, we observe that each classification result in the case of common factor numbers = 2~3. As a calculation rule, the maximum likelihood method for initial solution and varimax rotation were selected, and the cases where the number of common factors is 2~3, are shown as Tables 2 and 3 by use of the output factor load amount denoting F_i, and uniqueness denoting U. Here, each factor is classified as a common factor, F, having the maximum factor load amount.

Table 2 Factor analysis result of G (number of factors = 3)

	F_1	F_2	F_3	U
Ana	0.446	-0.360	-0.241	0.613
Des	0.756	-0.196	-0.052	0.387
Dev	0.202	-0.021	0.016	**0.959**
Imp	0.161	-0.026	0.984	0.005
Eva	0.099	0.002	0.167	**0.962**
Dis	-0.933	-0.211	-0.283	0.005
Int	-0.037	0.994	-0.074	0.005

Table 3 Factor analysis result of G (number of factors = 2)

	F_1	F_2	U
Ana	0.380	-0.920	0.005
Des	0.650	0.070	0.576
Dev	0.170	0.180	**0.937**
Imp	0.110	0.360	0.865
Eva	0.110	0.290	**0.902**
Dis	-0.990	-0.110	0.005
Int	0.100	0.430	0.804

Tables 2 and Table 3 show that the pair of *Ana* (Analysis) and *Des* (Design), and the pair of *Dis* (Discussion) and *Int* (Introduction) are belonging to the same common factor, even when there is a different number of common factors. Furthermore, it is found that the factor loading amount of *Dev* (Development) and *Eva* (Evaluation) were particularly small, compared to the other factors in the common factor, and large uniqueness. Finally, to determine the unnecessary factors to be excluded, the values of average (A), median (M), and standard deviation (S) of F for each factor are shown in Table 4.

Table 4 Average value (A), median (M), and standard deviation (S) of each factor

	A	M	S
Ana	0.242	0.250	0.164
Des	0.241	0.250	0.180
Dev	**0.024**	**0.000**	**0.079**
Imp	0.082	0.000	0.124
Eva	**0.073**	**0.000**	**0.110**
Dis	0.253	0.200	0.235
Int	0.079	0.000	0.171

Table 4 shows that *Dev* and *Eva* had lower average values and median values, and also had lower standard deviation values than other factors. We can conclude that students recognize these as less important in learning clinical research. From the above, let us assume that $G_{ex} = \{Ana, Des, Eva, Dis, Int\}$ is a factor group excluding two factors of *Dev* and *Imp* from the factor group G. Additionally, each number of common factors and its eigenvalues about D_{score}, which are scored by an education program student about "factors that are felt useful in clinical research skill acquisition" belonging to G_{ex}, are shown as Fig.3.

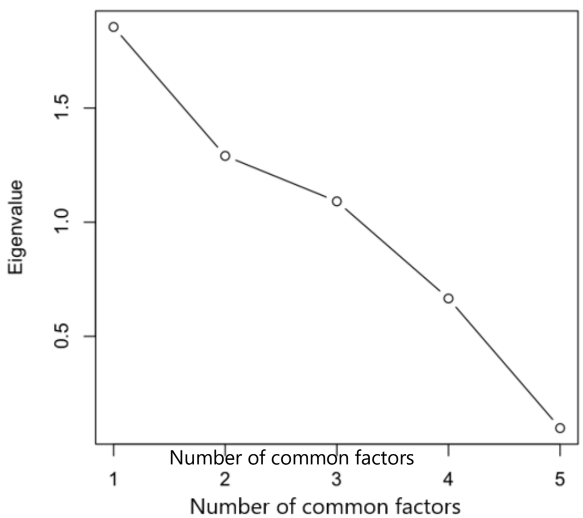

Fig. 3 Number of common factors and eigenvalues

As in Fig. 2, since the eigenvalues attenuation has a slow curve, each classification result in the case of common factor number = 2~3 is also shown here. Tables 2 and 5 show that, in the case of common factor number = 3, it is not affected by factor deletion because there is no change in uniqueness denoting *U*, although there is a change in factor load amount *F* of all factors.

Table 5: Factor analysis result of G_{ex} (number of factors = 3)

	F_1	F_2	F_3	U
Ana	0.390	-0.380	-0.310	0.613
Des	0.730	-0.230	-0.170	0.387
Imp	-0.010	-0.030	1.000	0.005
Dis	-0.970	-0.160	-0.140	0.005
Int	0.000	1.000	-0.060	0.005

Table 6: Factor analysis result of G_{ex} (number of factors = 2)

	F1	F2	U
Ana	0.331	-0.360	0.760
Des	0.693	-0.224	0.470
Imp	0.134	-0.092	**0.973**
Dis	-0.986	-0.148	0.005
Int	0.006	0.997	**0.005**

On the other hand, Tables 3 and 6 show that when the number of common factors = 2, *Imp* belonged to a same common factor with *Ana* and *Des* after deleting the unnecessary factor, while it was belonging to a different common factor before deletion. In addition, although *Imp* was not the deletion target in this report, its degree of uniqueness is higher than other factors and it changed by factor deletion. In other words, it is affected by deleting *Dev* and *Eva*. Therefore, it is concluded that stability of *Imp* is not so high.

Finally, the factor analysis results of *G* and G_{ex} in the case of each common factor number were visualized as shown in Fig. 4.

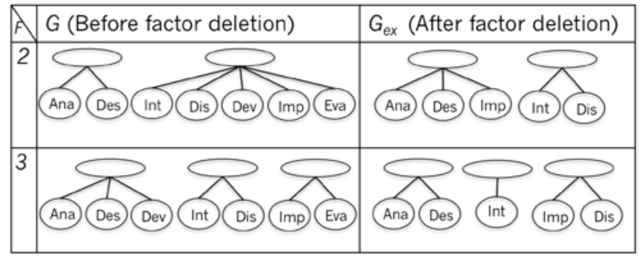

Fig. 4 Classification results on the number of common factors before and after factor deletion

From the results of G_{ex}'s number of common factors = 3, we could interpret that, as a procedure of acquiring clinical research skills, firstly, we should learn (***Int****roduction*) about clinical research, and secondly, we should organize them to solve clinical questions. (***Ana****lysis* ・ ***Des****ign*). After the start of the research, it is possible to interpret factors such as progressing according to the situation (***Imp****lementation* ・ ***Dis****cussion*).

5. Conclusion

In order to modelize the clinical research education program, text mining and factor analysis are applied to the free description formulas data and selective formula data answered by the program students. The factors required for the model were extracted, and we visualized procedures to learn clinical research.

References

[1] David E. Kern, Patricia A. Thomas, "Internet Resources for Curriculum Development in Medical Education", Journal of General Internal Medicine 2004

[2] https://cioms.ch/wp-content/uploads/2017/01/1991_INTERNATIONAL_GUIDELINES.pdf (accessed 2017)

[3] https://cioms.ch/wp-content/uploads/2016/08/International_Ethical_Guidelines_for_Biomedical_Research_Involving_Human_Subjects.pdf (accessed 2017)

[4] Ayako Ohshiro, Shinichiro Ueda, "Intermediate educational activity report of Clinical Research Management human resources development program", Igaku Kyoiku/ Medical Education (Japan) 47 367-370 2016.

[5] Gagne, R. M., Wager, W.W., Golas, K. C. & Keller, J. M, "Principles of Instructional Design (5th ed.). ADDIE model", Thomason Wadsworth. 21-37 2004

[6] Ayako Ohshiro, Shinichiro Ueda, "Analysis of ADDIE factors based on student surveys for generating clinical research education model", The 49th Annual Meeting of the Japan Society for Medical Education 2017

[7] Ayako OHSHIRO, Shinichiro UEDA, "Comparative study of explanatory factor analysis for construction of Clinical Research education model", IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.3, 27-30 2018

[8] Jeremy M. Anglin, George A. Miller and Pamela C. Wakefield, “Monographs of the Society for Research in Child Development”, The Society for Research in Child Development,Vol. 58, No. 10 1993

Ayako OHSHIRO received the B.S. and M.S. degrees from University of the Ryukyus in 2009 and 2011, respectively. She took Ph.D. from University of Ryukyus in 2017 and she has been assistant professor at University of the Ryukyus. Her research interests are data analysis at the field of Clinical Research education.

Shinichiro UEDA received the M.D in 1985 and took Ph.D. from Yokohama City University in 1997 and he has been professor at University of the Ryukyus. His research interests are Clinical Pharmacology and Therapeutics, and Clinical Research education.