PHAR-QA QUALITY ASSURANCE - ERASMUS PROJECT

II. GAUGING EFFECTIVENESS OF QUALITY ASSURANCE ACTIVITIES IN PHAR-QA "KICK-OFF" MEETING

IRINA PRASACU[1], CONSTANTIN MIRCIOIU[1], VALENTINA ANUȚA[1], FLAVIAN RĂDULESCU[1], ROXANA SANDULOVICI[1], DUMITRU LUPULEASA[1], JEFFREY ATKINSON[2], BART ROMBAUT[3], CRISTINA RAIS[1]-, IONELA BELU[4]

[1]Faculty of Pharmacy, University of Medicine and Pharmacy "Carol Davila", 6 TraianVuia, 020956, Bucharest, Romania
[2]Emeritus Professor Lorraine University, Pharmacolor Consultants Nancy (PCN), 12 rue de Versigny, 54600 Villers, France,
[3]President, European Association of Faculties of Pharmacy, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, 1090 Brussels, Belgium
[4]Faculty of Pharmacy University of Medicine and Pharmacy, Craiova
** corresponding author: cristina_rais@yahoo.com*

Abstract

The PHAR-QA 'kick-off' meeting targeted the debut of discussions about the development of standard operation procedures for quality assurance in European pharmacy education and training. Questionnaires regarding participants (members of PHARMINE consortium and other guests) opinion concerning organization, quality assurance of the curriculum, evaluation of the academic achievements, goals and as well as other developments of project, were distributed. The analysis was a qualitative one, partner 9 (University of Medicine and Pharmacy "Carol Davila" being responsible for the quality plan in the project) used statistical parametric and statistical nonparametric methods with the purpose of gauging the effectiveness of quality systems. Verification of all feedbacks and observations lead to the conclusion that the best solution is the application of a mosaic of statistical tests combined with phenomenological analysis of the education process.

Rezumat

Întâlnirea care a dat startul proiectului PHAR-QA a urmărit începerea discuţiilor referitoare la dezvoltarea unor proceduri standard de operare referitoare la asigurarea calităţii în învăţământul şi pregătirea farmaceutică europeană. Au fost distribuite chestionare conţinând răspunsurile participanţilor (membri ai consorţiului PHARMINE şi alţi invitaţi) cu privire la organizare, asigurarea calităţii curriculei universitare, evaluarea rezultatelor şi obiectivelor propuse precum şi alte evoluţii ale proiectului. Analiza a fost una calitativă, partenerul 9, Universitatea de Medicină şi Farmacie "Carol Davila", fiind responsabilă cu planul de calitate în proiect folosindu-se metode statistice parametrice şi non parametrice în scopul evaluării eficacităţii calităţii sistemelor.

Verificarea tuturor *feedback*-urilor şi observaţiilor a dus la concluzia că cea mai bună soluţie este aplicarea unui ansamblu de teste statistice, combinat cu analiza fenomenologică a procesului de învăţământ.

Keywords: quality assurance, gauging effectiveness, Likert scales, statistical methodology.

Introduction

An initiative of the European Association of Faculties of Pharmacy (EAFP) was to create a consortium of universities from European Union member states or other European countries which, in association with other partner associations representing community (PGEU), hospital (EAHP) and industrial pharmacy (EIPG), together with the European Pharmacy Students' Association (EPSA) was to harmonize the pharmaceutical education and training activities.

The result was the project PHARMINE run in the 2008 – 2012 period. As a natural prolongation of PHARMINE, last year, a successful application assured the financial support for a new *Lifelong Learning Centralized Erasmus* project [16].

Quality assurance in pharmacy education and training in Europe (PHAR- QA) more specifically oriented towards a consensus for a common European pattern for the quality assurance systems over the Europe [13, 14, 15].

PHAR-QA project includes 5 work programmes (WP): WP 1 *Management*; WP 2 *Implementation*; WP 3 *Quality Assurance* (Quality Plan); WP 4 *Dissemination*; WP 5 *Exploitation of results*.

WP3 is jointly coordinated by partner 1 (Vrije Universiteit Brussel – Prof. Bart Rombaut), partner 2 (Pharmacolor Consultants Nancy – Prof. Jeffrey Atkinson) and partner 9 ("Carol Davila" University of Medicine and Pharmacy – Prof. Constantin Mircioiu) [15].

In the frame of both projects it was created a special group dedicated to the quality assurance of project development and finalization in accordance with the essential objectives of the project [1].

A questionnaire based on the quality criteria of the International Pharmaceutical Federation and the Accreditation Council for Pharmacy Education (USA) [17] was sent out to European faculties. Replies were obtained from 28 countries. Just above half have a working QA system.

QA scores were high concerning matters such as curriculum and training, students' representation and promotion of professional behaviour; etc. QA scores were low concerning matters such as evaluation of achievement of mission and goals. This suggests that a QA system based on competence is required [13].

Statistical methods for analysis of data included usual descriptive tests for characterization of the population of results.

Further, comparative analyses of the systems from different countries or group of countries were developed after the final report of the project [5, 16].

In this paper there are presented specific statistical parametric and non-parametric methods [8, 10] for quality analysis of the activities of the achievements of the "kick-off" meeting. A more in depth methodological analysis was applied in order to develop standard operation procedures for analysis of both activities inside the project and global gauging of the effectiveness of quality systems.

Materials and Methods

Questionnaires. Fourteen questions were conceived to evaluate the opinions of participants concerning the quality of the activities developed previously and during the "kick-off" meeting of the consortium. Responses were asked on a five levels scale: strongly disagree, disagree, neutral, agree, and strongly agree.

Statistical evaluation. In order to perform a statistical analysis, responses were quantified by attribution of numbers from 1 to 5. Comparative analysis of the responses referred to individual questions, to individual responders as well as to their means. Methods applied were both parametric and non-parametric.

Distributions of responses were compared using chi-squared tests. Final analysis compared the results obtained with parametric and non-parametric methods.

Results and Discussion

A first analysis was the point wise correlation between *provided* in project and *achieved*. Quantitatively, it was found that the meeting responded to all objectives.

The proposed tool for measuring effectiveness of activities run by the board of the project was the statistical analysis of participants' responses to Likert type questionnaires, introduced in psychiatry some hundred years ago [7] for measuring psychological attitudes in a quantitative way. Commonly met are 5-point scales ranging from "Strongly Disagree" on one end to "Strongly Agree" on the other with "Neither Agree nor Disagree" in the middle; however, some practitioners advocate the use of 7 and 9-point scales which add additional granularity.

First incertitude appeared yet at the collecting data phase, before mathematical and statistical analysis: not all participants responded and not all responses concerned entire set of questions.

Two questions concerned the evolution of the project but actually the project was at the very beginning and some responders had not yet a clear opinion on this subject. The problem was if we have to consider these as missing data or to include them in the count of "neutral" rank.

Then a global analysis concerned a logical analysis of the reliability and significance of each individual response. Some misunderstandings were put in evidence as potential inductors of bias in statistical analysis. In the set of individual responses it was no "2" rank and only one "1". It was considered necessary to understand, by correlation with commentaries, the reason of the "1" response and it was found that the problem was a misunderstanding. Two raters were not previously informed about the program of the meeting but this was the problem of the managers of their working groups and not that of management of the project.

First statistical analysis was based on calculation and comparison of means. If we consider distribution of ranks as normal distributed, it makes sense to compute mean of ranks for each question and for each rater. Means of responses to all questions were situated between 4 and 5.

Inter-raters range of mean values (3.55 – 5) was larger than inter-questions range of mean values (4.1 – 4.64), but their total means were much closed (4.43 and 4.40).

But here appear a lot of "statistical methodology" problems. How Likert type measurement scales should be appropriately used and analysed has been debated for over 50 years. For example one approach [4] is based on the view that 'Likert scales' are ordinal in character (i.e., produce rank order data) and that they, therefore, must be analysed using non-parametric statistics. Kuzon et al. (1996) wrote about the ''seven deadly sins of statistical analysis'' [6]. Sin 1 is using parametric statistics on ordinal data; Sin 2 relates to the assumption of normality and claims that ''before parametric statistical analysis is appropriate… the study sample must be drawn from a normally distributed population''.

But, non-parametric statistics, however, are less sensitive and less powerful than parametric statistics and more recently underlined that what was left unsaid is how much it increases the chance of an erroneous conclusion [9]. This is what statisticians call ''robustness'', the extent to which the test will give the right answer even when assumptions are violated. And if it doesn't increase the chance very much (or not at all), then we can press on. Monte Carlo studies of the F-test have convincingly shown

that the F-test is extremely robust to violations of its assumptions, except for the homogeneity of variance assumption [3].

It is perfectly appropriate, therefore, to sum Likert items and analyse the summations parametrically, both univariately and multivariately [2].

Consequently there are many arguments for parametric analysis in spite of some violations of the assumptions on which are based such analysis. One such violation was that responses were non-symmetrical distributed. Our opinion is that the scale has to be centred on the most probable response. For example excellent – very good – good – acceptable – not sufficient.

Comparison of interquestions and interraters distribution of ranks for detection of outliers

In order to establish outlier items (question or raters) it was applied "Leave One Out (LOO) method", i.e. comparison of an item with the sum of the rest of items by calculation of the correlation coefficient.

Responses to questions were very well correlated and this test didn't indicate the existence of outlier questions.

Further, since ranks can be considered as multinomial distributed random variables, an alternative method for testing hypothesis concerning outlier character of one or more questions, it was applied the χ^2 test for comparison of distribution of responses [11, 12].

Calculated values were smaller than threshold ($\chi^2_{4;0.95} = 9.488$) with exception of case for question 2. Consequently second question had responses not correlated with the answers to the rest of questions. This non-correlation appears from the responses of two raters concerning their information before the meeting, as was presented above.

Application of test, based on multinomial distribution, for comparison of distributions was also a lame method since it is recommended that all entries in matrix to be > 5. Wilcoxon order statistics test would be a choice but difficult to apply following too many equal values ("ties").

It is thought also that one can't use parametric tests in this study because the sample size is too small. But nowhere is there any evidence that non-parametric tests are more appropriate than parametric tests when sample sizes get smaller.

Conclusions

All deliverables provided in project were materialized in program of the "kick-off" meeting program and were achieved.

Results of statistical evaluations indicated that all evaluators for all questions gave a high appreciation to the quality of accomplished tasks; means being situated between 4 and 5, i.e. between agree and strongly agree.

An outlier response, was connected with the fact that one participant was not informed sufficiently before meeting about the program of the meeting. Analysis of the situation indicated that his information was the task of the manager of the working group and not of the management of project.

Since there is a never-ending dispute concerning the application of parametric or non-parametric tests in analysis of results on Likert scales the solution is to apply a mosaic of tests and verify all results and also examine potential outliers with complementary tools, information and common sense.

Acknowledgements

This paper was supported by project PHAR-QA, funded by the EU ERASMUS Lifelong Learning Programme (527194-LLP-1-2012-1-BE-ERASMUS-EMCR).

References

1. Atkinson J., Rombaut B., The 2011 PHARMINE report on pharmacy and pharmacy education in the European Union. *Pharmacy Practice (Internet)*, 2011; 9(4): 169-187.
2. Carifio J., Perla R., Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ*., 2008; 42(12): 1150-1152.
3. Glass G.V., Peckham P.D., Sanders J.R., Consequences of failure to meet assumptions underlying the analyses of variance and covariance. *Rev. Educ. Res*., 1972; 42(3): 237-288.
4. Jamieson S., Likert scales: how to (ab)use them. *Med. Educ*., 2004; 38: 1212-1218.
5. Jones A.B., Principles in Quality Assurance, Part 3: Making an Impact. *Qual. Assur. J*., 2009; 12: 132-138.
6. Kuzon W.M.Jr., Urbanchek M.G., McCabe S., The seven deadly sins of statistical analysis. *Ann. Plast. Surg*., 1996; 37(3): 265-272.
7. Likert R., A Technique for the Measurement of Attitudes. *Arch. Psychol*. 1932; 22(140): 55.
8. Mircioiu C., Sandulovici R.C., Statistica aplicată în farmacie şi studii clinice, second edition, Carol Davila Publishing House, Bucharest 2009; 25, 217-222.
9. Norman G., Likert scales levels of measurement and the "laws" of statistics. *Adv. Health Sci. Educ. Theory Pract.*, 2010; 15(5): 625-632.
10. Prasacu I., Miron D.S., Mircioiu C., Seminar de matematici aplicate în farmacie, second edition, Carol Davila Publishing House, Bucharest 2010; 15-16.
11. Rais C., Modelare structurală şi funcţională a proceselor de formare a resursei umane farmaceutice, PhD Thesis, University of Medicine and Pharmacy Carol Davila Bucharest 2011.
12. Rais C., Anuţa V., Rădulescu F., Prasacu I., Lupuleasa D., Atkinson J., Rombaut B., Mircioiu C., Quality assurance PHAR-QA – ERASMUS project. I. Contribution of "Carol Davila" University of Medicine and Pharmacy Bucharest. *Farmacia*, 2014; 62(2): 236-244.
13. PHAR-QA - Quality Assurance in European Pharmacy Education and Training‖. 527194-LLP-1-2012-1-BE-ERASMUS-EMCR. Available online: http:// www.pharmine.org/PHAR-QA (accessed on 22 July 2014).
14. PHAR-QA: Quality assurance in pharmacy education and training in Europe. PHAR-QA DISS press release (English draft) dated 01 November 2012.
15. PHAR-QA: Quality assurance in pharmacy education and training in Europe. PHAR-QA Form project description, version 2; 2012.

16. Pharmine Web Site. Pharmacy Education in Europe. Education & Culture DG. Lifelong Learning Programme. Available online: http://www.pharmine.org/ (accessed on 22 July 2014).
17. U.S. Department of Health and Human Services, Food And Drug Administration, Title 21, Food and Drugs, Chapter I: Part 58 Good Laboratory Practice For Nonclinical Laboratory Studies. Final Rule published in the US Federal Register 52:33768-33782, 4 September 1987. Available online: http://www.fda.gov/RegulatoryInformation/Dockets/FR/default.htm.

Manuscript received: May 2013