Large or Small Panel in the Delphi Study?
Application of Bootstrap Technique

Somayeh Alizadeh[1], Shahnam Sedigh Maroufi[2], Zohreh Sohrabi[3],
Azam Norouzi[4], Reza Jafaei Dalooei[5], Ghobad Ramezani[6]

[1]Center for Educational Research in Medical Sciences (CERMS), Department of Medical Education, School of Medicine, Iran University of Medical Sciences, Tehran, Iran. [2]Center for Educational Research in Medical Sciences (CERMS), Department of Anaesthesia, School of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran. [3]Center for Educational Research in Medical Sciences (CERMS), Department of Medical Education, School of Medicine, Iran University of Medical Sciences, Tehran, Iran. [4]Center for Educational Research in Medical Sciences (CERMS), Department of Medical Education, School of Medicine, Iran University of Medical Sciences, Tehran, Iran. [5]Department of Medical Education, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran [6]Center for Educational Research in Medical Sciences (CERMS), Department of Medical Education, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

ABSTRACT

BACKGROUND
Despite consensus group methods such as Delphi technique commonly used in medical education, there is a lack of standardization in methodology definitions and reporting. Size of the panel is one of fundamental questions of the researcher in a Delphi study. The present study was conducted to determine adequacy of the panel size in the Delphi round.

METHODS
In this study, to determine adequacy of the panel size, the data stability was investigated in Delphi round. The Delphi study was conducted to determine competencies of educational leaders in Iran's medical sciences universities. Original data was collected from 33 experts, who were selected purposefully. Computer-generated samples of 1000 and 2000 were used for resampling. Data analysis was performed using bootstrap technique.

RESULTS
Results of the bootstrap technique showed that, in each of the three data groups (original and resampling data), 10 common items were in priority of 1 to 10. Ethical, commitment, punctuality, full recognition of their duties, strategic thinking, believing in values and morals, self-confidence, team working, trusteeship, motivation, maintaining morale, and supporting the faculty and staff were selected as 10 common competencies.

CONCLUSIONS
Stability results of three data types showed that, selected panel size was sufficient to conduct this research. In Delphi studies, regardless of emphasis on small or large panel size, a panel of appropriate size should be selected. Using statistical methods, adequacy of size of the study panel can also be ensured.

KEY WORDS
Expert Panel Size, Delphi Study, Bootstrap.

BACKGROUND

In medical education, consensus methods are used to determine new component of the curriculum, developing items for evaluation of tools, describing competencies, and developing educational resources. Consensus methods are systematic tools for measuring and building consensus. Despite a widespread use, a few numbers of the sources describe the use of these methods in detail, thus there is also instability in their use.[1]

Consensus methods include different techniques, but the Delphi technique and nominal group are among the most commonly used ones.[2] Delphi was first introduced by RAND Corporation in the United States in 1950 and was developed in 1970 as a comprehensive tool[3] This method facilitates communication between a group of specialists to encounter a complex problem.[4] In this technique, consensus is achieved using a structured blind process with repetition and presenting individual-based feedbacks.[5] One of important points in using different techniques is the issue of reliability and accuracy. According to some assumptions, researchers claim that the Delphi technique provides the evidence in relation to content and face validity. Results obtained from judgment team rather than a person seems to be more credible. In addition, the possibility that several individuals make wrong decision rather than only one individual is less.[6,7] Despite the evidence regarding validity of this technique, there are two basic issues while adopting the Delphi method. The first issue concerns design of the questionnaire. Selection of measurement scale and number of the rounds are main points in design of the questionnaire. The second issue is selecting the panel of experts.[6] Group of experts participating in the Delphi method is called "panel."[2,6] In the panel selection, the tips for the panel consists of size, panel member features, and the response rate.[6] There is a complete agreement on how to choose the participants. They are not randomly but purposively selected to enter the research to apply their knowledge and experience with respect to the problem.[6-9] The point about participants is size of the expert panel. Although it is sufficient to reach an acceptable size of 10 experts, there is no agreed standard for number of the participants.[10-13]

It has been pointed out that purpose of the study and the resources available to participate influence on selection of the participants, but there is no standard method to calculate number of specialists needed to achieve stable data.[15] The use of larger panel increases variety of specialists, other issues are highlighted with the increase in the number of panel members and with the increase in the number of participants, however, the probability of sample loss and reduction in the percentage of respondents' response increases.[2] In addition to using stability check, accuracy of the results from the panel can also be ensured. Stability is not usually reported in Delphi studies.[7] In this study, bootstrap was used to determine stability and consequently panel size adequacy. This method was developed by Efron in 1979 and has found widespread use in the field of applied statistics.[15] Bootstrap is a Monte Carlo-type data augmentation method utilizing resampling through replacement that can be used with observed data. This research is a part of a large project aimed at identifying competencies of educational leaders in Iran's medical sciences universities, using a modified Delphi technique.

METHODS

The present research was conducted to determine adequacy of the panel size by investigating data stability in a Delphi round. The modified Delphi method was used to evaluate competencies of educational leaders and predetermined statements were sent to the participants.

Delphi Questionnaire Design
In a modified Delphi, an open–ended - round of a classical Delphi is replaced with statements developed from existing literature in the field or focus groups or undertaking one-to-one interviews.[7] In this research, a critical review was used to extract the statements. Two independent reviewers were selected to read the papers. The questionnaire was developed based on competencies extracted from the literature. In the questionnaire, competencies were divided into two types (personal, functional).

Delphi Panel
Primary questionnaire was distributed among first group participants, by email. The first group participants included medical and educational management specialists. Agreements and disagreements of the experts with each statement were marked, and they were asked to determine that which competencies need to be trained. The questionnaire was open-ended. The experts had the opportunity to comment on their items in the questionnaire. The second questionnaire was developed based on the comments and results of the expert panel. The second group of the participants was different from the first group. These participants were asked to determine importance of each item based on a scale of 1 to 10. A score of 10 represented high importance. Second group of the participants were the faculty member, and they had managerial experience. The second group involved 50 heads of departments and vice-chancellors selected based on purposive sampling. Ethical approval was obtained from the Research Ethics Committee at Iran University of Medical Sciences. Written informed consent was taken from all the participants.

Bootstrap Technique
In this study, at first, response characteristics of the participants, mean, trimmed mean, standard deviation, and 95% confidence interval were determined. At the next step, responses of the participants were randomly selected through replacement by a computer program based on primary data. Computer-generated samples of 1000 and 2000 were used for resampling. Data analyses were conducted by SPSS software.

RESULTS

There were two groups of participants in this study. From the first group, 8 participants and from the second group, 33 participants completed and returned the questionnaire. In the first group, response rate was equal to 80%. Mean age was equal to 41.5 $\pm$ 3 years and mean managerial experience was equal to 14$\pm$2 years. There were 1(12.5%) female participant and 7 (87.5%) male participants. In the second group, response rate was equal to 66%. Participants included 18 females (55%) and 13 males (45%). Minimum of managerial experience was 1 year and its maximum was 25 years, and mean managerial experience was equal to 8.9 $\pm$ 4.9 years.

Original and Augmented Data Results
Based on raw data results, from personal competency items, the ethical item had the highest mean (9.80$\pm$0.9) and the items of recognizing types of evaluation methods, monitoring and controlling of individuals, groups, organizations, and programs had the lowest mean (7.28$\pm$0.8). From functional competency items, team working had the highest mean (9.48$\pm$0.16) and networking item had the lowest mean. (7.99$\pm$0.14) (table. 1)

Competency Item	Competency Type	Mean	5% Trim.	Mean SD	95% CI	
Ethical	Personal	9.80	9.80	0.09	9.60	9.96
commitment	Personal	9.68	9.68	0.09	9.48	9.84
Punctuality	Personal	9.64	9.64	0.14	9.36	9.88
Full recognition of their duties	Personal	9.56	9.56	9.56	9.20	9.88
strategic thinking	Personal	9.52	9.52	0.17	9.16	9.80
Belief in values and morals	Personal	9.52	9.52	0.14	9.24	9.76
Self Confidence	Personal	9.48	9.48	0.14	9.20	9.76
Team working	Functional	9.48	9.48	0.16	9.16	9.76
trusteeship	Personal	9.44	9.44	0.20	9.00	9.80
Motivate, maintain morale and support the faculty and staff	functional	9.44	9.44	0.17	9.12	9.76
Table 1. Statistical Data for the Original Sample (N: 33)						

Based on results of data augmentation with repetition (1000), the ethical item showed the highest mean (9.88$\pm$0.10) from personal competency, and the team-working item (9.58$\pm$0.16) showed the highest mean from functional competency (table. 2).

Competency Item	Competency Type	Mean	5% Trim	Mean SD	95% CI	
Ethical	Personal	9.88	9.86	0.10	9.26	10.00
Punctuality	Personal	9.71	9.70	0.15	9.40	9.97
commitment	Personal	9.70	9.70	0.10	9.48	9.88
Full recognition of their duties	Personal	9.68	9.65	0.19	9.23	9.97
strategic thinking	Personal	9.62	9.61	0.19	9.19	9.92
Innovation	Personal	9.58	9.48	0.30	8.65	9.88
Team working	functional	9.58	9.55	0.16	9.23	9.83
Belief in values and morals	Personal	9.58	9.57	0.15	9.27	9.84
Trusteeship	Personal	9.54	9.55	0.20	9.14	9.92
Self Confidence	Personal	9.53	9.53	0.15	9.18	9.79
Motivate, maintain morale and support the faculty and staff	Functional	9.53	9.52	0.18	9.10	9.83
Table 2. Statistical Data for the Augmented Samples (N: 1000)						

In terms of the data augmented with repetition (2000), ethical item (9.88$\pm$0.9) and team working (9.58$\pm$0.16) had the highest mean (table. 3).

Competency Item	Competency Type	Mean	5% Trim	Mean SD	95% CI	
Ethical	Personal	9.88	9.86	0.09	9.67	10.00
Punctuality	Personal	9.71	9.71	0.15	9.40	9.97
commitment	Personal	9.70	9.70	0.10	9.48	9.88
Full recognition of their duties	Personal	9.68	9.67	0.18	9.28	9.97
strategic thinking	Personal	9.62	9.61	0.17	9.23	9.90
Team working	functional	9.58	9.56	0.16	9.22	9.83
Belief in values and morals	Personal	9.58	9.57	0.15	9.27	9.83
Innovation	Personal	9.58	9.47	0.32	8.69	9.88
Trusteeship	Personal	9.54	9.53	0.21	9.06	9.89
Self Confidence	Personal	9.53	9.53	0.15	9.22	9.80
Motivate, maintain morale and support the faculty and staff	functional	9.53	9.52	0.17	9.18	9.83
Table 3. Statistical Data for the Augmented Samples (N: 2000)						

Results showed that, in each of the three data groups, 10 common items were in priority of 1 to 10 (table. 4). Means of competencies were stable across three data (original and two augmented) sets. In general, participants̀ response scores remained stable across the three data sets (n: 33 n: 1000, n: 2000).

Sample Type	Competencies
Original n:33	1-Ethical, 2- commitment, 3- Punctuality, 5- Full recognition of their duties, 6- strategic thinking, 7- Belief in values and morals, 7- Self Confidence, 8-Team working, 9- trusteeship 10-Motivate, maintain morale and support the faculty and staff
Augmented n:1000	1-Ethical, 2- Punctuality, 3- commitment, 4- Full recognition of their duties, 5- strategic thinking, 6- innovation, 7- Team working, 7- Belief in values and morals, 8- trusteeship 9-Self Confidence, 10-Motivate, maintain morale and support the faculty and staff
Augmented n:2000	1-Ethical, 2- Punctuality, 3-commitment, 4-Full recognition of their duties, 5- strategic thinking, 6- Team working, 7- Belief in values and morals innovation, 8- trusteeship 9-Self Confidence Motivate, 10- maintain morale and support the faculty and staff
Table 4. Prioritization of Competencies in the Three Sample Groups	

DISCUSSION

Despite consensus group methods such as Delphi technique commonly used in the medical education, there is a lack of standardization in methodology definitions and reporting.[16] There are still several questions about this methodology. One question is whether small expert panel is adequate to perform a Delphi study.[12] This is an important question because size of the panel influences on various issues such as drop- out percentage from the study, and number of Delphi rounds.[6] In this study, adequacy of the panel size was investigated by checking stability of the data in a modified Delphi round. Results showed that characteristics of responses of original samples were stable similar to simulated samples with repetitions of 1000 and 2000. As a result, size of selected panel was sufficient to conduct the research. Akins (2005) also in a study on determination of stability of the Delphi round results from Bootstrap showed that response characteristics of the participants in simulated and actual samples were stable, indicating that size of the panel was appropriate for the study. In this research, final size of a Delphi panel was 32.[12] Strom (2017) 32 participants of 10 different countries entered the study to design a new scale for assessing competencies in EVRA process.

Although, percentage of response was acceptable, the effect of drop-out percentage of the study was evident in reported Cronbach's alpha coefficient.[5] Although, purposive sampling is used in Delphi studies, there is still the possibility of lack of cooperation and withdrawal of participants from the study. It is believed that a sample size of 20 tends to retain the members.[17] Shanna (2015) used the modified Delphi method to design the Cardiopulmonary Bypass Separation checklist; however active members of the Society of Cardiovascular Anesthesiologists were equal to 3000. Among which, 90 subjects were in the first round of Delphi, which declined to 71 in the fourth round. [18] Bentley et al (2016) used a modified Delphi to determine level of consensus across Interprofessional Education (IPE) for collaborative practice. Although, 477 stakeholders were identified, the Delphi panel included 56 members.[19] Kenny (2008) points out that whether the questionnaire is long and includes different domains, instead of sending the entire questionnaire to all participants, each area of the questionnaire will be sent to a specialist related to the topic.[7] Homogeneity and heterogeneity of the samples is another discussion about size of required panel. Both concepts require very different sample sizes.[20] If various reference groups are involved in a Delphi study, more subjects are needed. Size of the panel also influences number of Delphi rounds. Three rounds are usually recommended for samples larger than 30 experts.[6] As noted, size of the Delphi panel influences the entire study in different ways. If sample size was too small, the participants may not be considered as a representative of pooled judgments regarding the issue. If the sample size is too large, then response rate may decrease.[6,12] Therefore, in Delphi studies, regardless of emphasis on small or large panel size, a panel of appropriate size should be selected. Using statistical methods, adequacy of size of the study panel can also be ensured.

CONCLUSIONS

In this study, adequacy of the panel size was investigated by checking stability of the data in a modified Delphi round. Results showed that characteristics of responses of original data were stable, similar to simulated data, with repetitions of 1000 and 2000. As a result, size of selected panel was sufficient to conduct the research. There is no agreed standard for the number of the expert panel.[10-13] In Delphi studies, regardless of emphasis on small or large panel size, a panel of appropriate size should be selected. Using statistical methods, adequacy of size of the study panel can also be ensured.

ACKNOWLEDGEMENT

REFERENCES

[1] Humphrey-Murto S, Varpio L, Gonsalves C, et al. Using consensus group methods such as Delphi and Nominal Group in medical education research. Medical Teacher 2017;39(1):14-9.

[2] McMillan SS, King M, Tully MP. How to use the Nominal group and Delphi techniques. International Journal of Clinical Pharmacy 2016;38(3):655-62.

[3] Heiko A. Consensus measurement in Delphi studies: review and implications for future quality assurance. Technological Forecasting and Social Change 2012;79(8):1525-36.

[4] Vernon W. The Delphi technique: a review. International Journal of Therapy and Rehabilitation 2009;16(2):69-76.

[5] Strøm M, Lönn L, Bech B, et al. Assessment of competence in EVAR procedures: a novel rating scale developed by the Delphi technique. European Journal of Vascular and Endovascular Surgery 2017;54(1):34-41.

[6] Giannarou L, Zervas E. Using Delphi technique to build consensus in practice. International Journal of Business Science and Applied Management 2014;9(2):65-82.

[7] Keeney S, McKenna H, Hasson F. The Delphi technique in nursing and health research. Hoboken, NJ: John Wiley & Sons 2010.

[8] Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud 2001;38(2):195-200.

[9] Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. Journal of Information Technology Education 2007;6:1-21.

[10] Eubank BH, Mohtadi NG, Lafave MR, et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. BMC Medical Research Methodology 2016;16:56.

[11] Avella JR. Delphi panels: research design, procedures, advantages and challenges. International Journal of Doctoral Studies 2016;11(1):305-21.

[12] Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Medical Research Methodology 2005;5:37.

[13] Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12:1-8. Art no. 10.

[14] McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016;38(3):655-62.

[15] Efron B, Tibshirani RJ. An introduction to the bootstrap. 1st edn. CRC Press 1994.

[16] Brady SR. Utilizing and adapting the Delphi method for use in qualitative research. International Journal of Qualitative Methods 2015;14(5):1609406915621381.

[17] Mullen PM. Delphi: myths and reality. Journal of health Organization and Management 2003;17(1):37-52.

[18] Ogden SR, Culp Jr WC, Villamaria FJ, et al. Developing a checklist: consensus via a modified Delphi technique. Journal of Cardiothoracic and Vascular Anesthesia 2016;30(4):855-8.

[19] Bentley MW, Kerr R, Powell S. The use of a modified Delphi technique to inform the development of best practice in inter-professional training for collaborative primary healthcare. Journal of Research in Inter-professional Practice and Education 2016;6(1).

[20] Baker J, Lovell K, Harris N. How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques. Nurse Researcher 2006;14(1):59-70.