Building tools to enable an automatic analysis of Delphi processes' results in health settings

Sarah dos Santos Magalhães
Mestrado Integrado em Engenharia Biomédica
Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract: Health services have become basic in everyday consumer life. For that reason, it is extremely important to include decision-makers' and stakeholders' perspectives in health decision. Participative processes as the Delphi technique, have been increasingly used to include their perspectives in these decisions. However, the analysis of these processes requires a lot of manual work by the user regarding the treatment of the participants' answers and Delphi outputs. Therefore, it is crucial to develop automatic tools to digest information and present outputs generated by Delphi processes in health settings. This thesis proposes the development of a novel and innovative Decision Support System (DSS) to treat Delphi participants' responses and to provide statistical outputs of the features that describe these processes. This DSS will incorporate the analysis of three types of Delphi processes commonly used in healthcare – Delphi for selection of indicators, Delphi for weighting judgments and Delphi for shaping value functions. A review of the literature focusing on Delphi processes and the main techniques that have been used to perform a complete analysis of its features was performed. Also, a framework for DSS was developed following a design and then the *DelphiAnalysis* implementation was performed in Microsoft Excel. The tool was tested with data from a real healthcare project and compared with the available results of the project. Additionally, a webpage guide was developed to help Delphi users along with a questionnaire to collect the opinion of Delphi experts about the tool and the webpage. The results obtained when testing the tool proved the efficiency of the DSS as it can provide the statistical measures accurately and without errors, in a short period. Moreover, the results are provided in table formats that grant user-friendly outcomes. Delphi experts provided positive feedback regarding the DSS and the webpage created and provide some suggestions to improve the *DelphiAnalysis* DSS in future work.

Keywords: Health decision; Delphi processes; Statistical outputs; Decision Support System; Microsoft Excel

1. Introduction

Healthcare is a multidimensional field that requires a lot of research to provide quality services to the target population [1]. Technological development and medical improvements brought higher patient expectations, increasing the demand for good health services. Therefore, resources need to be allocated efficiently to attend the necessities of the society [2]. However, this is not easy regarding the variety of aspects that must be taken into account in healthcare - the involvement of stakeholders and/or decision-makers, the fact that evaluators in health are also the decision-makers and that their individual opinions can bring conflicts

[2]. To improve health quality it is necessary to identify the problems and implement qualitative and quantitative methods involving multiple criteria [3]. Multicriteria Decision Analysis is intended to help decision-makers evaluating different options to reach the best decision, based on the most appropriate evidence [4], [5]. Regarding the qualitative methods used to improve health services that consider stakeholders' and decision-makers' (DM) perspectives, interviews, observation and analysis of documents stand out as collecting data techniques. However, some problems related with these techniques, as pressure and fear by the participants, have been reported [3], [6]. To solve some of these problems, quantitative techniques, as consensus approaches, are required. The most common consensus methods are the nominal group technique, consensus conference and Delphi processes [7]. The Delphi technique is a communication method that allows participants to express their opinions about a subject, anonymously [8]. On each round, a survey is filled by the panelists and at the end of each round, feedback about the responses of the entire panel is provided so that the respondents can adapt their answers according to the rest of the opinions, if they wish so [9], [10]. These processes have stood out in healthcare for their ability to avoid conflicts, pressure and personal bias [8]. Some web-based platforms have emerged to help performing a Delphi and to promote consensus among stakeholders from different geographies. However, these platforms only facilitate the application of the process, but do not perform statistical analysis of the features that describe Delphi processes. Therefore, a lot of manual work by the researcher is still needed, taking some extra valuable time. The goal of this thesis is to offer valuable information by developing a novel Decision Support System (DSS) to help decision analysts and health DM to access Delphi results, in this case, three types of Delphi commonly used in healthcare - for selection of indicators, for weighting judgments and for shaping value functions -, and make analyses of the main features that describe these processes, in a more expedite way.

2. Context

Healthcare quality is defined as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and care consistent with current professional knowledge" [1]. Consensus methods are a way to include professional knowledge on health decisions dealing with conflicts in scientific evidence and overcoming the disadvantages in decision-making of groups or committees [7] . These methods attempt to assess the extent of agreement about a subject, solving disagreements at the same time [7]. The Delphi technique is one of the most useful consensus technique that provides an opportunity for experts to communicate their opinions anonymously and avoiding domination of any expert [1]. Moreover, this method eradicates geographical barriers, saving time, money and inconvenience [1], [2]. Delphi processes have then, the potential to provide useful information in healthcare and facilitate wider group participation being, for that reason, a

powerful and promising research tool to apply in many health fields.

3. Literature Review

The Delphi technique is a structured communication method whose first goal was to obtain consensus about an important topic, and it is based on the idea that group opinion is more trustworthy than an individual belief. A panel of experts are questioned about their opinion through series of surveys with controlled opinion feedback between rounds [9]. The method avoids negative face-to-face interactions, stimulating individual sharing of opinions without pressure or fear [11].

Taking into account that there are no universal guidelines on the use of this method, it suffered a lot of modifications over the years, what reflects its wide flexibility and led to the appearance of many types of Delphi techniques - Conventional, Modified, Real-Time, Policy and others -, which vary in the first goal and some characteristics: [8], [12]. Although the existence of several subtypes, anonymity, iteration, controlled feedback and statistical group response are characteristics that underlie the technique and are implicated in almost all these sub types of Delphi [8]. Anonymity is used to address different panel opinions without matching them with their identification [13], which grants the respondents not to feel pressured by other participants and admits that any response has equal weighting for the analysis [8]. Iteration is given through successive rounds allowing participants to adjust their judgments in consecutive rounds [9]. Between rounds, the researcher takes care of the members' responses providing quantitative or qualitative data as controlled feedback [8], [14]. Lastly, the statistical group response expresses quantitative measures of the judgments of the entire panel as an average opinion [8], [13]. These characteristics provide advantages to the method as flexibility and simplicity to support specific studies, keeping the participants motivated [13], [15]. Some drawbacks also exist - experts' bias, tendency to eliminate extreme positions to achieve consensus, lack of empirical rules, the definition of consensus, size of the panel and time-consuming problems when dealing with complex issues – which can be overcome or minimized if a good planification is performed before implementing the method [13], [16]. First, it is important to identify the nature of the issue being studied and understand if the Delphi approach is adequate to deal with it [17]. Then, the availability of resources, the definition and establishment of the necessary level of consensus should be studied along with the analysis of time and costs inherent to the execution of the process [15]. The extension of each round, size of the panel and language used in the surveys should be planed as well to avoid misinterpretations [18]. Also, the experts must have knowledge, communication skills, experience, and enough time to cooperate [16]. For each specific study, the researcher needs to perform a protocol with detailed steps to pursue to obtain the most accurate results, saving time and money as much as possible [16]. This planification is a task prone to error as it requires a lot of concentration and time [10]. One way to avoid errors, automatizing this task, is to use

a DSS, i.e. a computer-based system that combines data from various sources, supporting choice by assisting the decision-makers. A proper DSS design requires the identification of the nature of the target decision problem and deal with all the issues identified in the proposed framework (Figure 1) that was adapted from an existent one proposed by Miah et. al [19].

4. Methodology

Knowing all the requirements to perform a Delphi process, it is important to establish the main steps needed until the development and implementation of the DSS. A workflow was elaborated to show the information that need to be gathered in order to be able to implement the best DSS for analysing the proposed types of Delphi.

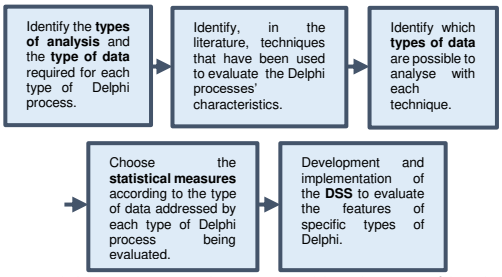

Figure 1 - Block diagram representing the steps made until the implementation of the tool.

According to the block diagram, the first task is to identify the types of data used in each one of the Delphi processes that will be incorporated in the tool and to find the techniques that have been used to perform a complete analysis of the main features that describe these specific types of Delphi studies. After, it is important to understand which techniques deal with the types of data addressed in these Delphi processes and choose the best ones to implement in the DSS that can provide a complete and diverse analysis of them.

4.1. Types of scales, data, and methods

In statistics, there are different types of scales and data that must be understood to evaluate Delphi processes as they deal with different types of data [20]. Different rules lead to different kinds of scales and measurements. The type of scale depends upon the character on the type of data and the basic operations performed. Quantitative data represents the one that can be measured with numbers and qualitative data is non-numerical one that is usually textual or descriptive [20]. Regarding scales, in statistics there are four types: nominal, ordinal, interval and ratio [20]. Nominal scales contain rules for deciding if two objects are equivalent [21]. The result of a nominal scale is a series of classes that may be given different numeric designations [21]. Ordinal scale arises from the operation of rank-ordering, involving rules to decide if an object is greater or less than another, concerning a specific attribute [21]. In this case, the order of the values is what's important and significant [21]. Interval scales are quantitative scales that define the unit of measurement as higher, equal or less, knowing the exact distances among objects on the attribute [21]. Lastly, ratio scales are the ones where the relations equality, rank-order, equality of intervals and equality of ratios are applied [21]. According to the type of data under analysis, two different types of methods can be distinguished. When the data is interval or ratio-scaled, the statistical measures should be parametric, i.e, the ones that make assumptions about the parameters of the sample distribution [22], normally suitable for Delphi processes with several experts [23]. When dealing with a small number of experts, using categorical

or ordinal data or when the distribution of responses for each item is non-normal, non-parametric methods are chosen [23].

3.2. Statistical measures and data

During May 2019, databases were used to find the best techniques used to describe Delphi processes features. A group of 22 articles were collected and analysed in order to understand which were the best techniques used to analyse the three Delphi processes described above according to the type of data that each technique deals with. The most important measures identified were measures of central tendency - mode, median and mean -, and measures of dispersion - variance, standard deviation and Interquartile range, that should be incorporated in Delphi assessment as they are the most basic ways to have useful information about statistical group responses and to provide feedback to the panel [22]. A lot of additional qualitative and quantitative methods were identified and analysed. However, since the quantitative ones were identified as the most useful to describe Delphi's features, only these ones were deeply analysed and characterized as parametric or non-parametric in order to be chosen to be implemented in the DSS [24].

3.3. Types of Delphi evaluated in the tool

After a detailed analysis of the techniques, it was necessary to choose the best ones to incorporate in the *DelphiAnalysis* DSS, taking into account the types of Delphi evaluated by the tool, which are described above.

1. **Delphi for selection of indicators**

This technique will be applied using a 5-level Likert scale to express the opinion of the experts on each indicator. The Likert scale is an ordinal scale that measures the strength or intensity of agreement or disagreement of the panelists [25]. They will vote using the categories of the scale - *Strongly Disagree*, *Disagree*, *Neither Agree nor Disagree*, *Agree* or *Strongly Agree* - to select or discard indicators under analysis reflecting scientific evidence and the opinions of experts and stakeholders [25].

2. **Delphi for weighting judgments**

Weights are used in indicator aggregation allowing the users to assign different weights to the indicators [26]. This technique will use an ordinal MACBETH to collect qualitative weighting judgments on the indicators, showing how important is to achieve the evolution of its performance range [26]. The 7-level MACBETH semantic scale incorporates the categories *Extreme, Very Strong, Strong, Moderate, Weak, Very Weak* and *Not Important*.

3. **Delphi for shaping value functions**

Here, the objective is to obtain the shape of a value function - Linear, Concave, Convex or S-shaped - that describes the performance range of each indicator. Only the first round will use the 7-level MACBETH scale described in the previous case. The performance range of the indicators is divided into three gaps, and the participants need to vote on how important is to close each one of the gaps, obtaining a shape that describes their opinions through the three votes. In the following rounds, the participants will vote directly on the type of

shape that they think is the best to characterize the performance range.

3.4. Selecting the statistical measures

All the three types of Delphi that will be analysed by the tool deal with ordinal scales. For that reason, the statistical measures to implement must be non-parametric and so the parametric ones were discarded in the first place. Finally, it was decided to implement the same statistical measures for the evaluation of all Delphi processes. However, in the case of the value functions, some procedures could not be implemented as the method doesn't work with numerical data in all rounds. For ordinal scales as Likert and MACBETH, feedback reporting central tendencies and dispersion it's very beneficial since it allows the participants to see how their answers compare to the group opinion as a whole [26]. The analysis of the statistical measures revealed that standard deviation and mean are not appropriate when dealing with ordinal scale although they are misused many times. Belton *et. al* (2019) [26] recommended providing the median and inter-quartile range for responses made to individual ordinally-measured items and so these methods were chosen to be part of the statistical techniques to implement, along with the mode. Regarding the level of agreement among panelists, the *Kendall's Correlation Coefficient* and *Fleiss' Kappa* were chosen as they are useful when dealing with levels of concordance. Also, the latter considers the agreement obtained by change and can be used with three or more raters [25]. Regarding stability of responses and change of opinions, the *Spearman's Correlation Coefficient* was selected since is appropriate to use with ranked data [27]. This method will be used to study the overall stability of each round and to see the stability of each indicator between two consecutive rounds. The *Wilks' Lambda* test will be used to evaluate the group's opinion variance along with its p-value associated in the first two types of Delphi. This method allows examining whether the responses given by the panel are statistically different across groups [26]. Along with these procedures, the frequency and percentages of votes, percentage of people who changed opinion between rounds and which participants changed their opinion the most are other examples of analysis that will be offered by the tool. In the value function Delphi, the last statistical methods described will not be applied except the central tendency and dispersion measures that will be applied for the first round. For the two following rounds, the percentages of each shape, mode, and bi-mode and if some curve obtained more than 50% of the votes is the information that will be provided in the DSS.

After selecting the statistical measures to implement, it was possible to proceed to the next phase: the *DelphiAnalysis* DSS implementation.

3.5. *DelphiAnalysis* implementation

In order to understand the main points that should be addressed by the DSS, a framework design was developed following an existent one proposed by Miah et. al [19], as presented in Figure 2. After, the implementation of the *DelphiAnalysis* DSS was done in Microsoft Excel as it is a useful platform to deal with statistical measures and allows to provide user-friendly outcomes.

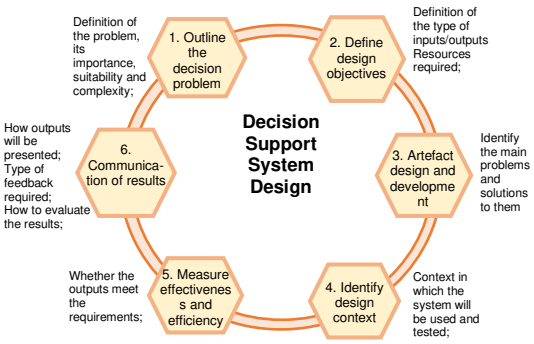

***Figure 2** - Framework design of a DSS (based on [19]).*

The evaluation of each type of Delphi was developed in a different Excel file to avoid the tool to become too heavy and slow. Each Excel has different sections to execute different functions. The first tab is where the user can upload the input data (answers provided by the panel members) and where the data used in the process can be adapted to the data recognized by the tool. The DSS have limitations regarding the number of indicators and participants that the Delphi processes can have - 200 indicators and 350 participants. In the second tab, the tool automatically converts the input categorical responses into numerical data. The results are present in the third tab. First, it presents the results for each round and after the analysis performed between consecutive rounds. To implement the statistical measures, present in the third tab, the steps of how to perform each measure were followed from the literature coding some mathematical programming and using some functions from the Excel. On the fourth tab, some information obtained in the results is disposed in a way that makes it easier to be analysed by the user, concluding about how some important features (level of agreement, the stability of responses and inter-reliability) evolved throughout the process. For the first two types of Delphi, there is another tab called “MANOVA” where the group’s opinion variance can be studied only by applying two formulas. All the Excel files have an additional sheet with extra calculations performed the code needed to implement the statistical measures.

To implement the *Wilks’ Lambda* test and its associated p-value, a *RealStatistics Resource* pack available for Excel, which included additional functions, was downloaded. However, since the application of those formulas depends on the range of the data, these measures were not possible to be automatically calculated by the tool. The user needs to implement the formulas to his data range.

3.6. DSS webpage and questionnaire

A webpage guide was developed to help the user to correctly use the DSS. This guide was developed using the Wix platform, which allows creating professional websites. The webpage incorporates the Excel files of the *DelphiAnalysis* DSS, the explanation of how this work emerged, an introduction about Delphi processes, and some videos, images and explanations of how to correctly use the tool. Also, there is a section where the user can contact us in case of doubts.

Additionally, a questionnaire was developed to collect information about what Delphi experts from the Wisedon company though about the DSS and webpage. The questionnaire aimed to obtain validity for the implemented DSS and to receive suggestions of what could be improved in future works.

4. Results

First of all, it is important to note that the main objective was to develop a tool to help

Delphi' users to automatically obtain statistical measures to analyse the features that describe the three proposed Delphi processes, but not to provide explanations about how to interpret the results obtained.

The implemented DSS was tested using data from a real healthcare project - the EURO-HEALTHY project. Some of the results could be compared with the ones obtained in the real project as frequencies, percentages, central tendency and dispersion measures, and the types of shapes obtained in the third type of Delphi process. These results shown to be accurate since they were the same of the ones obtained in the project, only varying because of the rounding. The other methods were tested using the same data, but they could not be compared. However, the tool provided the expected outputs quickly, without any errors or problems. Small examples that could be performed by hand were tested as well and the DSS seem to work correctly, giving the expected results. The summary page presented in the fourth tab was also displaying the expected results in a useful way to conclude about how the percentages and features that describe Delphi processes evolved during the rounds.

Regarding the answers provided by the Delphi experts in the questionnaire, it was possible to conclude that they found the website and the tool useful and intuitive. The DSS was considered complete and diverse in the type of evaluation that it offers. However, some improvements were suggested as creating the same tool using a better mathematical programming language to so the tool could become faster, with fewer limitations and with a better interface, or providing more notes and step guides in the tool.

5. Discussion

The proposed methodology was defined for three specific cases, where the best statistical measures were chosen according to their type of data and main objectives, following the information existent in the literature. When testing the tool, all the statistical measures are calculated immediately without any errors. Moreover, by comparing the results with the available ones of the EURO-HEALTHY project, it became clear that the *DelphiAnalysis* tool can lead to correct outcomes. It is evident that for the three specific types of Delphi, the tool can be useful in automatically obtaining the statistical measures implemented, saving time and avoiding manual work for the Delphi user.

However, the DSS was implemented in Excel, which limited the types of analysis implemented, as the graphical analysis, which is often important when interpreting the results. Also, it implemented constraints relative to the number of participants, indicators, and rounds which could be avoided if the implementation would be made using another programming language more sophisticated. In general, it can be said that the tool can be useful when dealing with the specific types of Delphi selected as it can provide the most common statistical measures automatically, fast and without mistakes, avoiding the manual working of the researcher and saving time.

Finally, despite all the limitations of the model, it can be seen as an example to follow and improve in the future since it is an

innovative tool that can be useful for Delphi analysts.

6. Conclusion

The main objective of this thesis was to develop a DSS to automatically analyse Delphi processes usually used in healthcare. Therefore, an independent Excel file to analyse each type of Delphi process selected was developed along with a webpage that could give some guidance to the users about how to correctly use the tool. A questionnaire was also developed to collect some Delphi experts' opinions about the developed methodology and the webpage guide created. The work developed in this thesis shown that the first objective was clearly fulfilled and that it was still possible to go further by creating adjacent characteristics.

The proposed DSS addresses the pre-defined objectives as it calculates many statistical measures usually used to analyse the features that describe these types of Delphi processes. This DSS is the first tool that automatically calculates statistical outputs of Delphi processes only with the answers of the panelists. Therefore, it decreases the time needed to evaluate these processes and the user does not need to perform manual work to perform these analysis, which makes it a model to follow in the future. However, being the first tool of this nature developed, it has several dimensions where it can be improved or even continued to be more complete or independent.

6.1. Future Work

The DSS was evaluated only one time, using the data obtained in a real project developed in the healthcare context. In the future, more tests should be done, using different types of data and, preferably, from different contexts. The implementation of more and new statistical measures that can done to complement the analysis offered by the DSS. Similar tools can be created to analyse other types of Delphi or a similar tool can be implemented using another mathematical programming, more sophisticated and adapted to statistics. Regarding the validation of the DSS, it would be great to have more Delphi experts' answering the questionnaire to validate the tool or organize an interview with a group of professionals who could give good advices about it.

7. Acknowledgements

First, I would like to share with you how happy, proud and grateful I am for getting here. I will be eternally grateful for all my experiences, all the people I have met, and all the years spent in IST.

A special thanks to the professors Mónica Oliveira and Ana Vieira for their motivational support, understanding and knowledge sharing that made this work rewarding!

I truly acknowledge to João Bana e Costa, Teresa Rodrigues, Fábio Martins, and Mónica Nóbrega from Decision Eyes for providing all the help I ask them and certainly, for making my days happier during this process.

To those with whom I shared the best and worst moments during the 5 years of faculty and especially those who have been always by my side during the last period: Carolina Barata, Katrin Munzenrieder, Luís Gonçalves, Liliana Lameiras, María Miño, and Sofia Fonseca.

To my father and mother for making all this possible, for the greatest effort they have made but, distinctively, for their love and support. To my brothers, who are always there, no matter what. A special acknowledge to my uncles, Teresa and Orlando, and to my cousin Pedro, for being my second family and to my grandparents who always want to help me.

To the ones who still live in me: Dani, who taught me what it was to have a brother, who taught me how to spread love and smiles, and my grandmother and aunt Florinda, who taught me how to be good. I wish I could still have you with me.

8. References

[1] R. Boulkedid et.al , "Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review," *PLoS One*.

[2] G. Montibeller, "Resource Allocation and Priority Setting in Health Care : A Multi-criteria Decision Analysis Problem of Value ?," vol. 8

[3] C. Pope, "Qualitative methods in research on healthcare quality," *Qual. Saf. Heal. Care*

[4] B. V. S. TJ, *Multiple Criteria Decision Analysis: An Integrated Approach*. Massachussetts: Kluwer Academic Publishers, 2012.

[5] P. Thokala *et al.*, "Multiple criteria decision analysis for health care decision making - An introduction: Report 1 of the ISPOR MCDA"

[6] J. Kitzinger, "Qualitative Research: Introducing focus groups," *Bmj*, vol. 311.

[7] J. Jones, "Consensus methods for medical health," vol. 311.

[8] M. R. Geist, "Using the Delphi method to engage stakeholders: A comparison of two studies," *Eval. Program Plann.*, vol. 33.

[9] S. Thangaratinam and C. W. Redman, "The Delphi technique," *Obstet. Gynaecol.*, vol. 7.

[10] H. A. Linstone and M. Turoff, "Delphi: A brief look backward and forward," *Technol. Forecast. Soc. Change*, vol. 78, 2011.

[11] J. V. Meijering et al. "Quantifying the development of agreement among experts in Delphi studies," *Technol. Forecast. Soc. Cha.*

[12] J. F. Amber, P. M. Gregory, "Using the Delphi Method for Qualitative, Participatory Action Research in Health Leadership," *Int. J. Qual. Methods*, vol. 13.

[13] A. B. Costello and J. W. Osborne, "Practical Assessment, Research & Evaluation," *Pract. assessment, Res. Eval.*, vol. 10.

[14] "The Delphi Technique: characteristics and sequence model," *Couper, M. R. (1984).*

[15] C. Powell, "The Delphi technique: Myths and realities," *J. Adv. Nurs.*, vol. 41.

[16] S. Keeney, F. Hasson, and H. McKenna, "Debates, Criticisms and Limitations of the Delphi," *Delphi Tech. Nurs. Heal. Res.*

[17] S. Keeney, F. Hasson, and H. McKenna, "How to Get Started with the Delphi Technique," *Delphi Tech. Nurs. Heal. Res.*

[18] A. R. Romano, "Malleable Delphi : Delphi Research Technique , its Evolution , and Business Applications," *Int. Reviwe Bus. Res. Pap.*, vol. 6, no. 5, pp. 235–243, 2010.

[19] S. Miah et.al , "A Development-Oriented DSS Evaluation Approach: A Case Demonstration for Conceptual Assessment," *Australas. J. Inf. Syst.*, vol. 17.

[20] "Stevens S. On the theory of scales of measurement.pdf," *Science*, vol. 03.

[21] J. C. Nunnally and I. H. Bernstein, "NunnallyBernstein in Psychometric Theory."

[22] H. A. von der Gracht, "Consensus measurement in Delphi studies," *Technol. Forecast. Soc. Change*, vol. 79.

[23] S. A. Kalaian and R. M. Kasim, "Terminating Sequential Delphi Survey Data Collection - Practical Assessment, Research & Evaluation," *Pract. Assessment, Res. Eval.*

[24] T. Keeley *et al.*, "The use of qualitative methods to inform Delphi surveys in core outcome set development," *Trials*, vol. 17.

[25] H. A. Shah "Which Is the Best Parametric Statistical Method For Analyzing Delphi Data?," *J. Mod. Appl. Stat. Methods*, vol. 8.

[26] I. Juwana et.al "Indicator-based water sustainability assessment - A review,".

[27] I. Belton et.al "Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process,"