Expert Judgment Quantification

Muhammad Amer and Tugrul Daim

Abstract Expert judgments are used when there are no objective data is available. It is critical to solicit these judgments accurately for decision makers. This chapter reviews methods and issues around the expert judgment quantification.

1 Expert Judgment Methods

Expert judgment is the data given by an expert in response to a question or problem. Judgment is defined as an inferential cognitive process by which an individual draws conclusions about unknown quantities or qualities on the basis of available information [1]. Meyer and Booker describe that expert judgment consists of information and data obtained from the qualified individuals that can be used to solve problem or make decisions in various fields and domains [2]. Keeney and von Winterfeldt describe expert judgement as an expression of opinion, based on knowledge and experience, that experts make in responding to a problem [3]. In the literature expert judgment is also called expert opinion, subjective judgment, expert forecast, best estimate, and educated guess.

The following are some applications of expert judgment [2]:

- Determining probability of an event and assess impact of a change
- Determining present state of knowledge in a field
- Predicting performance of a product or process
- Determining validity of assumptions

M. Amer · T. Daim (✉)
Portland State University, Portland, USA
e-mail: tugrul@emp.pdx.edu

M. Amer
e-mail: amer1992@gmail.com

T. Daim et al. (eds.), *Research and Technology Management in the Electricity Industry*, Green Energy and Technology,
DOI: 10.1007/978-1-4471-5097-8_3,

- Selecting input and response variables for a chosen model
- Providing the elements needed for decision making in the presence of several options.

Expert judgment is often obtained and considered a very reliable option available when faced with an uncertain future and having a lack of historical data [2, 4]. Generally decision making in these situations with high degree of uncertainty about a future environment give rise to two specific needs [5]:

- The need for a methodology to capture the reliable consensus of opinion of a large and diverse group of experts; and
- The need to develop models of future environments, which would permit various policy alternatives and their consequences to be investigated.

Roger Cooke states the following five principles in an attempt to formulate guidelines for obtaining expert judgment [6]:

- Reproducibility: It should be possible to review and if necessary to reproduce all the results.
- Accountability: The source of expert judgment must be identified (not necessarily by name, but certainly by professional background and level of expertise).
- Empirical Control: The results must in principle be susceptible to empirical control.
- Neutrality: The method for combining and evaluating expert opinion should encourage experts to state their true opinions.
- Fairness: All experts should be treated equally.

The comparison of the various expert judgement methods reveals a set of generic phases/steps which are used to a greater or lesser extent in each method depending on its objectives. Following are the generic phases of an expert judgement method [7]:

- Definition of elicitation objectives
- Identification and selection of the experts
- Preparation of questionnaire, instruments, training session etc.
- Process of obtaining the expert opinion
- Analysis and aggregation of expert judgments
- Synthesis.

There are various means to obtain expert judgment and according to Börjeson et al. usually workshops, Delphi method, and surveys are conducted to obtain expert opinion for the development of scenarios and roadmaps [8]. Technology roadmaps are always developed based upon expert judgment obtained through workshops, expert panels, Delphi studies, and surveys [9]. Therefore, use of expert panels is a widely used approach for technology planning and roadmapping and it has been used in several Ph.D. dissertations [10, 11].

Cooke provides a brief historical overview of expert judgement methods and argues that systematic use of expert judgement for decision making was developed at the RAND Corporation in the United States after World War II [6]. The first two methods using expert judgement, developed by the RAND Corporation, were the Delphi method and Scenario Analysis.

2 Scenario Analysis

Herman Kahn is considered as the father of scenario analysis approach and he developed this approach at RAND Corporation [6, 12, 13]. Kahn and Wiener define scenarios *"as hypothetical sequences of events constructed for the purpose of focusing attention on causal processes and decision-points"* [14]. Scenario planning has increasingly been applied as a useful tool for the improvement of decision-making process and dealing with uncertainty, by considering number of possible future environments [15]. Scenarios are alternative, plausible and consistent images of the future and highlight the large-scale forces that push the future in different directions [12]. Scenarios are useful whenever the problem is complex, uncertain and has long-term effects [16]. So scenarios significantly enhance the ability to deal with uncertainty and increase the usefulness of overall decision making process [17, 18].

There has been significant growth in the use of scenario planning, especially in the decade up to the year 2010 [17, 19]. Scenario planning has been extensively used at the corporate level, and in many cases it has been applied at the national level [13, 20, 21]. The scenario building process also contributes towards organizational learning [22]. Shell was one of the first companies to use scenarios at the corporate level, and usage of scenarios helped the company to cope with the oil shock and other uncertain events in the 1970s [13, 23, 24]. Scenarios are considered a valuable tool that helps organizations to prepare for possible eventualities, and makes them more flexible and more innovative [25]. Empirical research conducted by Linneman and Klein indicate that after the first oil crisis in the early 1970s, the number of U.S. companies using scenario planning techniques doubled [26, 27].

There are numerous techniques for scenario development ranging from intuitive based approaches to purely quantitative approaches [28]. The literature on scenario planning indicates that scenarios mean different things for different users, and often scenarios are developed for various purposes [8]. On the basis of perspective, scenarios are classified into descriptive and normative scenarios [20]. The descriptive scenarios are extrapolative and the normative scenarios are goal directed. Scenarios are also classified on the basis of scenario topic, breadth of the scenario scope, focus of action, and level of aggregation [16].

Scenario planning approach has been widely used in the energy and renewable energy sector. Scenario planning helps to analyze emerging issues in a complex energy system [29]. It has been applied to forecast energy resources [30], energy foresight and long-term energy planning at the national level [31], analysis of future primary energy demand at national level [32, 33], improvement of energy efficiency and reduction of energy consumption in commercial buildings [34], development of hydrogen energy infrastructure [35–37], deployment [38] and integration of renewable energy [39, 40] and renewable energy portfolio planning [41–44].

3 The Delphi Method

The Delphi method was developed at the RAND Corporation in the 1950s as a spin-off of an Air Force sponsored research project, "Project Delphi" [6, 45]. The Delphi method is a popular technique for forecasting and an aid in decision making based on the experts opinions. The need to elicit and synthesize expert opinion inspired the development of the Delphi technique [46]. Linstone and Turoff define "Delphi as a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem" [47] Delphi is a systematic and interactive method relying heavily on expert panels [48]. Delphi is based upon the principle that judgments and opinions from a structured group of experts are better and more accurate than those from an individual or unstructured group [45, 49]. This method has gained popularity among research managers, policy analysts and corporate planners and has been used extensively in various fields. The Delphi method is applied to technology forecasting and many types of policy analyses in various fields and domains [48].

The Delphi method is based on a structured process for collecting and distilling knowledge from an expert panel through a series of questionnaires combined with controlled opinion feedback. A series of sequential questionnaires or rounds are conducted with controlled feedback in order to gain the consensus of opinion of a group of experts [47]. Feedback and opinion of group members is summarized, combined and given back to experts after the first Delphi round and they are asked the same questions again [47, 48]. This process is repeated until a general consensus in the outcome is obtained or results are stabilized. Research on number of Delphi rounds indicates that of the most changes occur in the transition from the first to the second round when members of expert panel change their judgment and generally four rounds are sufficient to reach consensus [50]. This systematic control ensures objectivity to the outcome of a Delphi study and provides a sharing of responsibility which releases the participants from group inhibition [51].

The Delphi Method also reduces the impact of the powerful members in the group by establishing anonymous group communication and avoids imposition of their point of views on other group members [47]. Torrance state that power or status of the group members influences the decisions and less powerful members demonstrate an unwillingness to disagree with the most powerful member, even if they have the correct solution and this may adversely affect quality of the decisions [52]. Moreover, in highly structured cultures individuals may refrain from expressing their opinions freely, so Delphi can be used as a useful approach to overcome cultural barriers [11]. The Delphi method is also used as a useful communication tool to generate a debate [53]. The following three characteristics of Delphi method distinguish it from conventional face to face interaction [48, 49, 54]:

- Anonymity: Each expert gives his answers to the questions in an independent and anonymous way without any undue social pressures and group members do not know who else is in the group. This gives an opportunity to the experts to freely express their opinion on the basis of merit alone.
- Iteration with controlled feedback: The process is reiterated until a degree of consensus is reached or results are stabilized. Iteration allows the experts to change their opinions. Controlled feedback takes place after every Delphi round, during which each group member is informed of the opinions of the other group members. Thus, participants are encouraged to review their answers in light of the combined judgment of all participants.
- Statistical group response: The set of responses (combined group judgment) is then sent back to the experts and they are asked if they wish to revise their initial feedback. It includes statistical information of the group response such as the mean or median and the extent of the spread of members' opinions.

In the first Delphi round, members of the expert panel are asked questions related to the subject matter under consideration. Moderator collects their judgments and provides feedback of the first round to the experts. For the second round, the experts are asked to either adjust their estimates or provide justification of their rationale if they differ from the majority judgment. Due to controlled feedback from the previous rounds, sometimes experts tend to achieve a consensus of opinion [55]. However, this process is iterated until consensus is generated or results are stabilized between two rounds. Linstone and Turoff emphasize that the number of rounds should be based on when stability in the responses is attained, not when consensus is achieved [56]. Chi squared test has been proposed in the literature to determine the stability of results from an expert panel [57].

The Delphi method has undergone substantial evolution and diversification. Online Delphi approach can play an important role and allows conducting study in shorter time period, ensuring anonymous contributions, summarizing results quickly, and has been used in multiple applications [58–60]. Linstone and Turoff state that in future Delphi will be used as an online tool [56]. In this approach Delphi panelists have access to a web-based questionnaire. Use of computers and internet have enhanced the original concept and make it possible for any group member to participate from anywhere in the world [58]. Geist made a comparison of traditional paper–pencil version of the Delphi method with web-based, computerized, and real-time version and described that web-based Delphi can overcome some shortcomings of the traditional approach [61]. Literature review also suggests some variants of the Delphi technique based on the purpose of the project like the classical Delphi, the policy Delphi and the decision Delphi [62, 63].

3.1 Advantages of the Delphi Method

The Delphi method helps to achieve consensus in a given area of uncertainty or lack of empirical evidence [64, 65]. Participants of a Delphi study bring their extensive knowledge and vast experience to the decision making process [65]. The Delphi technique is very useful in situations where individual judgments must be tapped and combined to overcome incomplete state of knowledge [64]. The controlled feedback between the Delphi rounds stimulate new ideas and it is also motivating for the participants [66].

The Delphi process facilitates the experts to participate in a group communication process asynchronously at times and places convenient to them, which is another key benefit [56]. Absence of an obligation to meet in person improves the feasibility of the Delphi, significantly lowers its cost [67], and allows participation from diverse geographic locations.

Rowe et al. suggest that the structured approach and participant anonymity offered by Delphi approach leads to a process gain [49]. Whereas, other methods of obtaining expert judgment or consensus like committees are considered to be prone to the biasing effects of personality traits, seniority, status, and domination by powerful individuals [55, 65]. In face to face interactive groups there is a tendency among low-status members to "go along" with the opinions of high-status members despite of contrary feelings [52]. In contrast, the Delphi method can overcome these negative effects and in a Delphi study consensus reflects a normative rather than informational influence or tendency to follow the leader [65].

Brockhoff, Riggs, Larreche and Moinpour found that a Delphi procedure produces superior predictions and accurate forecasts than a normal interacting group or a face-to-face committee meeting [68–70]. Rowe and Wright also systematically reviewed the empirical studies looking at the effectiveness of the Delphi technique and found that the Delphi method outperforms other structured group procedures by providing more accurate assessments or judgments [45].

The Delphi technique is also very useful in situations when there no adequate models exists to develop a statistical prediction and Coates says that Delphi is the last resort in these situations [71].

3.2 Disadvantages of the Delphi Method

It takes long time to conduct a Delphi study due to nature of the process [55] and extensive time commitment is required.

Output of a Delphi process reflects the best opinion of the experts [66] so it is critical to choose appropriate experts. Sometimes selection of expert panel is problematic.

It is possible that in pursuing the consensus among experts may lead to diminish some of the best opinion and the study may only generate a set of bland statements representing the lowest common denominator [72].

It is also argued that anonymity in a Delphi study may lead to lack of accountability of views expressed and encourage hasty decisions [72]. However, the sequential Delphi processes may positively discourage such action.

Sackman points out in his critical review of the Delphi method that it is difficult to determine reliability and scientific validation of the findings [72]. It is because Delphi studies are based upon intuitive judgments, collection of half formed ideas from the experts therefore, one cannot judge it on the same basis as a concrete measurement [66].

3.3 Application

The initial application of the Delphi method was in the area of national defense and after that it has been extensively used in a wide variety of applications [47]. Martino states that it remains one of the most popular methods for technology forecasting [73]. Delphi is considered a promising technique in future roadmapping [10, 74] and scenario planning activities [31, 75, 76]. The literature review highlights widespread use of the Delphi method for policy analysis, healthcare, education, finance, management, marketing, human resources, manufacturing, information systems, transportation, engineering, national foresight planning, urban planning, energy foresight, environment, budget allocations, service planning, analysis of professional characteristics and competencies, and curriculum development [21, 31, 47, 53, 74, 77, 78]. Delphi method has been a useful tool for solving problems in energy sector for developing energy roadmaps and energy foresight projects [31, 79–86].

Delphi method has also been used by many countries; there are examples of various Delphi studies conducted at national level in Germany [87], Japan [74, 88], France [83], Turkey [89], Thailand [90], India [91], Poland [31], Finland [82], Korea [92] and Austria [93].

4 Multi Criteria Decision Analysis

Multi-criteria decision analysis (MCDA) methods are considered an appropriate and useful decision making tool for multi-dimensional, intricate and complex decision problems. MCDA is also suitable for conflicting evaluations consisting of multiple aspects and helps the decision makers to find a way to make rational compromises. MCDA methodology requires identification of criteria, sub-criteria, and alternatives related to a goal, followed by assigning numerical measures to evaluate importance of criteria and sub-criteria and finally the alternatives are prioritized and ranked [94]. Experts are used to assign these numerical measures in order to prioritize the available options [95]. There are many MCDA methods highlighted in the literature including: Analytical Hierarchy Process (AHP), Multi-Attribute Global Inference of Quality (MAGIQ), Goal Programming, Simple Multi-Attribute Rating Technique (SMART), SWING, SIMOS and Technique for order preference by similarity to ideal solution (TOPSIS), Preference ratios in multi-attribute evaluation (PRIME), weighted sum and weighted product methods etc. [94, 96, 97]. Wang et al. conducted a detailed literature review and thorough analysis of various MCDA methods and concluded that AHP is the most popular and comprehensive MCDA technique [96].

Analytic hierarchy process (AHP) is a widely used MCDA method and considered a very effective and powerful technique. AHP approach was developed by Thomas L. Saaty in the 1970 s and it has been used by decision-makers in diverse applications to resolve decision problems. It is a paired comparisons technique. International scientific community has accepted AHP as a robust and flexible decision making technique, useful for complex decision problems [98]. AHP is primarily used for the resolution of choice problems in a multi-criteria environment [99]. AHP technique allows decision maker to decompose the complex decision problem in a logical manner into many small but related sub-problems in the form of levels of a hierarchy [100]. AHP technique also allows the decision makers to incorporate both quantitative and qualitative judgments into a decision problem [101]. Evaluation of weighting and scoring can be objective if actual data is available related to criteria and alternatives; otherwise subjective data (expert judgment) obtained by pairwise comparisons through expert panel is used.

In general, AHP methodology provides a comprehensive and rational framework for structuring a decision problem. AHP technique has the following three fundamental concepts [100]:

- Structure complex decision problem as a hierarchy of goal, criteria, sub-criteria and alternatives, with goal at the top of the hierarchy, criteria and sub-criteria at lower levels and alternatives at the bottom of the hierarchy
- Pair-wise comparison of elements (criteria and alternatives) at each level of the AHP model with respect to each criterion on the preceding level. Through pairwise comparison the ratio-scaled importance of each alternative is calculated
- Synthesizing the judgments over the different levels of the hierarchy.

Pairwise comparisons are used to prioritize and rank the criteria and alternatives for decision making. Satty recommends to use 1–9 scale measurements and eigenvector approach [100]. In contrast to this Kocaoglu recommends constant sum approach by allocating 100 points between each pair [102]. Constant sum method using 100 points is considered better than 1-9 scale measurements approach because user can state their judgments without limiting to nine point scale [11]. Through these three steps, AHP technique estimates the impact of each alternative on the overall mission or goal of the decision hierarchy. This approach also helps the decision-makers to compare conflicting criteria and subsequently prioritize and rank the alternatives.

AHP method employs a consistency test to screen out inconsistent judgments by any expert and this is also considered as an advantage of using AHP. It is important that the decision-makers should be consistent in their preference ratings expressed by pairwise comparisons. Saaty recommended that consistency ratio (CR) should be less than 0.10 and mentioned that CR greater than 0.10 indicates serious inconsistencies and in that case AHP may not provide meaningful results [100].

4.1 Application

AHP has been extensively applied to a wide variety of decision problems in various domains including project selection and evaluation, measuring business performance, technology evaluation and selection, technology policy, energy policy, new product screening, portfolio management, customer requirement structuring, arms control, transport systems, agriculture sector, real estate investment, conflict resolution, quality management, public policy, and healthcare [90, 94, 98, 99, 101, 103, 104]. In the energy sector especially for renewable energy technologies, AHP approach has been used for energy policy formulation, energy planning, power plant selection, power plant site selection, prioritizing emerging renewable energy technologies, energy resource allocation, energy resource assessment, integrated resource planning, renewable energy exploitation, controlling greenhouse gas (GHG) emissions, energy conservation, and developing energy management systems [94, 98, 103, 105–121].

5 Other Expert Judgment Methods

There are several other methods based on eliciting the expert knowledge and judgment through a group of experts. Some methods used in foresight studies and based on the use of expert knowledge are mentioned in this section.

5.1 Nominal Group Technique

Delbecq and Van deVen developed the Nominal Group Technique (NGT) in 1968 and it is a structured decision making method for working toward consensus [64]. In this method, every participant of the expert panel gives their views and ideas for the solution and these ideas are prioritized using a ranking process [64, 122]. Its major strength is that opinions of everyone are taken into account, so every team member has an equal voice in sharing ideas. During the ranking process, the duplicate solutions are eliminated and every participant ranks the ideas as 1st, 2nd, 3rd, 4th, and so on. This output of NGT is a prioritized list of ideas generated by a group of experts. It is critical to carefully select the members of an expert panel because the value of the NGT is based on their knowledge and expertise.

NGT has gained considerable recognition and it has been widely applied in health, social services, education, industry, and government organizations [123]. The NGT process consists of the following six steps after defining the problem [64, 123]:

- Brainstorm and generate ideas in writing
- Round-robin feedback from group members to record each idea in a terse phrase
- Discussion of each recorded idea for clarification and evaluation
- Individual voting to prioritize ideas by anonymous rating
- Brief discussion of the preliminary vote and
- Final individual voting through rank ordering or rating followed by group discussion and group decision.

NGT is a useful approach especially in the following situations [64, 122, 123]:

- When discussion is dominated by some individuals of the expert panel and it may prohibit participation or creativity of the other members
- When some members are reluctant to suggest ideas and participate due to apprehension of being criticized or any other reason
- When group members think better in silence
- When some group members are new and less experienced than others or there is difference in their social status like manager and staff
- When the issue is controversial or there is heated conflict
- When it is desired to generate a lot of ideas
- When it is required to prioritize a few alternatives for further examination.

Major advantage of NGT is that it provides balanced participation of every member of the expert panel in the process and final result. NGT groups perform better than other interacting groups in accuracy and better use of group resources because all members participate and it results in better decisions [124]. NGT is a simple technique and usually it takes less than a day to complete the entire process. It is also less costly than other group methods.

Major disadvantage of NGT is that it is overly mechanical, simplified and lacks flexibility. It is focused on a single purpose and single topic. Only individual brainstorming is done and cross-fertilization of ideas is constrained. NGT minimizes discussion and does not allow for the full development of ideas. Therefore, it is less stimulating group process than the other methods. It is also quite possible that opinions may not converge and consensus is not achieved in the voting process. Due to these reasons, it has been recommended to combine NGT with other group techniques to overcome these limitations [125].

5.2 Focus Groups

Focus groups are generally used for idea generation [126]. In this technique a group of experts focus on a topic and they are asked about their views, perceptions, opinions, beliefs and attitudes towards a product, service, concept, idea or advertisement [126]. Questions are asked in an interactive group setting and the experts are free to talk with other group members in an open environment. Generally focus groups do not produce an actual technology forecast but may be useful in generating an insight and list of items that may be used in conjunction with another technique. This method usually requires some group interaction prior to the creation of a list of ideas [122].

5.3 Brainstorming

Brainstorming is a popular and widely used tool for doing creative tasks in organizations such as developing products, redesigning business systems, and improving manufacturing processes [127]. Its main objective is to elicit ideas from a group of people [128]. Brainstorming brings new ideas on how to tackle a particular problem in a freethinking atmosphere and presents a wide range of ideas and solutions. Participants are encouraged to freely articulate their ideas followed by more rigorous discussion in order to stimulate creativity and thinking “out of the box”, to let dissident viewpoints enter into discussion at an early stage [128]. Brainstorming technique also supports the future studies but does not produce an actual technology forecast. Effective brainstorming sessions consists of 7–12 participants.

5.4 Mindmapping

This technique is also sometimes used in foresight with the brainstorming and other group discussion methods [128]. Experts are asked about the relationships between a large number of factors and highlight the forces driving or shaping a course of development. It allows for a quick charting of a group’s ideas into logical groupings and connections between them. Mindmapping technique can be used in the course of brainstorming for ideas, and can help establish a skeletal framework. Its output is typically a chart or set of charts, outlining key issues and the linkages between them; and this can be used for communication purposes, scenario construction, or in many other ways [128].

6 Expert Judgment Methods used in Nuclear Studies

The following expert judgment methods are developed and mainly used in studies related to nuclear power plants to assess safety of nuclear power plants, facilitate the safety related decision making, and get estimates of subjective probabilities for unknown parameters and uncertain events [129–131]. Cojazzi et al. analyzed these expert judgment methods in a benchmark exercise [130], and these are summarized in Table 1.

- NUREG
- KEEJAM
- STUK-VTT
- CTN-UPM
- GRS
- NNC
- SEJ-TUD.

Although these expert judgment methods presnted in Table 1 are very structured but these are developed and used to study nuclear safety issues. These methods are commonly used in technology foresight studies.

7 Formation of Expert Panel and Selecting the Experts

It is very important to carefully select the members of the expert panel because the quality of all expert judgment is directly based upon their knowledge, capability, and experience. Expert judgment is used to forecast the future utilizing information derived from individuals who have extraordinary familiarity with the subject under consideration [122]. An expert is a person who has the background and knowledge in the subject area and considered qualified to answer the questions [2]. Usually questions are posed on the experts when they cannot be answered by any other means. Members of an expert panel should reflect current knowledge and perception as well as they should be impartial to the research findings [55].

The literature also highlights that well known experts should be selected who are respected among their peers and careful selection of experts increases credibility to the project [132]. Camerer and Johnson describe that an expert is considered an experienced person having some professional or social credentials and know a great deal about their domain [133]. McGraw and Haribson-Briggs state that domain experience, commitment, patience, persistence, ability to communicate ideas and concepts, introspective of own knowledge, honesty, and willingness to prepare for session are the important personal characteristics of the experts [134]. Landeta highlights that selection of suitable experts help to achieve reliability of the study [135]. Rowe, Wright, and Bolger enforce this by stating that the degree of panelist expertise is a key influencing factor on the accuracy of the group

Table 1 Formal expert judgment methods [129–131]

Expert judgement technique	Developed by	Characteristics
NUREG	Nuclear Regulatory Commission (NRC), USA 1987–1990	NUREG is a highly structured method used to evaluate the state of knowledge of events, variables or processes, when there is incomplete data or inadequate models. It includes training of the experts, review discussions, individual elicitations, composition and aggregation of the opinions and review by experts In this approach, the domain experts write reports on the issue and their final estimates are elicited individually after expert's discussions and then averaged on equal weight basis
KEEJAM	Joint Research Center (JRC)—European Commission and Italian Universities in 1997	It is knowledge based method and employs Knowledge Engineering techniques. It includes explicit modeling of the knowledge and problem solving procedure of the domain expert This approach provides structured and disciplined support to the knowledge engineer in eliciting the knowledge and reasoning strategies of the experts, building consistent knowledge models, and applying these models to the solution of expert judgement task
CTN-UMP	University of Polytechnics, Madrid, Spain in 1997	This method is developed on the basis of the NUREG Method. It estimates subjective probabilities for unknown parameter and uncertain events and consists of nine steps. Its main difference from NUREG Method is regarding the way to aggregate evaluation of the experts.
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit (GRS), Germany in 1985	This Method has been developed to quantify the state of knowledge in elements of a decomposition of the question and to propagate it though this decomposition in order to arrive at a quantitative uncertainty statement for the answer. It aggregates the judgments at lower levels and propagates them through the decomposition to arrive at a quantitative expression of the resulting state of knowledge at the model output level

(continued)

Table 1 (continued)

Expert judgement technique	Developed by	Characteristics
STUK-VTT	STUK Radiation and Nuclear Safety Authority), Finland in 1997	The method is originally developed for quantitative risk or reliability assessments in engineering and economical analyses, where remarkable uncertainties are present The method is based on probabilistic representation of uncertainties. The predictions obtained from experts are expressed as probability distributions. Then, the combination of these assessments is based on hierarchical Bayes model. Due to this property, it is also possible to deal with experts who are not familiar with the concepts of probability
NNC	UK, 1996	NNC is a quality based method, leaning on quality assurance methods and problem solving processes; this approach is based on individual estimates. It involves a multi-disciplinary team, defined as a set of individuals with different but complementary skills
SEJ-TUD	Delft University of Technology, The Netherlands in 2000	This is a European Guide for expert judgment in uncertainty analysis. It deals with procedures to perform an expert judgment study with the aim of achieving uncertainty distributions for an uncertainty analysis. This method consists of 15 steps

judgment [49]. Quality and validity of the elicitation process is further improved when the experts feel that they are knowledgeable and well-informed [136].

In the context of scenario planning, van der Heijden state that experts are remarkable people who have some knowledge of the related field or industry, and are acute observers of the environment [18]. Scenarios expert panel members are expected to be fairly knowledgeable of the socio-economic contexts of the region [128]. Schaller highlights the importance of competent experts to ensure quality of technology roadmaps [137]. More qualitative approach has to put a strong emphasis on the selection of suitable experts [16]. Therefore, the experts should be selected based on their experience and knowledge in the relevant area as well as their ability to provide a fair and objective viewpoint.

Shanteau describes the following important characteristics of an expert [138]:

- Extensive and up-to-date knowledge
- Good perceptual abilities so that the expert can extract information and make good judgment
- Ability to sense what is irrelevant when making decisions
- Ability to simplify complex problems
- Ability to clearly communicate their expertise and persuade others
- Strong sense of responsibility and a willingness to stand behind their recommendations
- Ability to work under stressful conditions
- Ability to propose creative solutions to the decision problems
- Exhibit self-confidence in their decision making.

Usually the experts with different backgrounds are brought into the expert panels. It is critical to ensure diversity in the expert panel so that the problem under consideration is thoroughly analyzed from many viewpoints [132]. Murphy et al. argue that diversity in an expert panel leads to better performance because it helps consideration of different perspectives and a wider range of alternatives [65]. In support of this argument, Rowe and Wright suggest that diverse background of the experts brings diversity to a panel [45], and group situations may inhibit creativity and bring possibility of resolving ambiguous and conflicting issues [49]. Members of the expert panel bring their knowledge and experience to the decision making processes [65]. Linstone and Turoff suggest that diversity of viewpoints will help to generate interest and involvement among the participants of an expert panel [47]. Diversity of experts' experience is considered as an important asset to the success of a research project [128]. In addition to technical qualifications and expertise, the expert panel members should be creative thinkers, who can bring diverse viewpoints, work well in groups, and freely express their views and opinions [128].

Delbecq et al. cite that heterogeneous groups consisting of participants with widely varying personalities and different perspectives on an issue, produce acceptable solutions of higher quality than homogeneous groups [64]. Diverse backgrounds of expert panel members also help to assure that any bias from any member would have little impact on the overall outcome of the study [11]. Gathering diverse experts in the panel will minimize the influence of a single individual. Moreover, in order to avoid biases that may be caused by personal influence, it is recommended that members in the panel are not given information about the other members [11]. Ascher state that there is a human tendency to stick with the status-quo and not to look outside their comfort zone when considering the future and utilization of multiple experts with different viewpoint helps to overcome this human tendency [139]. Thus, diversity among the members of an expert panel plays a vital role and a group of diverse experts will lead to better quality of the judgment [132, 139].

It is critical to have appropriate number of experts in an expert panel. Mitchell state that the expert panel must have at least 8–10 members [140]. Whereas, Meyer and Booker recommend to have around five to nine experts in a panel and state that having less than five experts in a panel will reduce the chances of providing adequate diversity of information to make credible inferences [132]. Some researchers suggest that more participants are better because their combined opinions will increase the reliability of composite judgment [65]. However, large number of participants in an expert panel will result in difficulties to coordinate with them and analyze their feedback. Therefore, it is important to balance the size of the group appropriately. Research indicates that for Delphi studies a group of 11–15 experts is preferred for achieving high correlation [48, 64, 141]. It has been empirically proved that panel reliability increases with increase in panel size and 11–15 is considered an optimum size of an expert panel [48]. In a widely cited article Powell emphasizes that success of a Delphi study clearly rests on the combined expertise of the members of the expert panel and highlights the importance of appropriate panel size [53]. The number of experts also depends on the scope of the problem, objective of the study, and resources available [64, 132] and sometimes a larger group may be useful if the study seeks to increase group support or understanding rather than decision making or gaining information.

Issues related to logistics are also important concerns for the selection of experts such as: willingness of expert to join the panel, willingness to devote their time for the study, and permission from their employer to participate in the research [132]. It is very important that expert panel members are willing and able to make a useful contribution [47, 53]. Thus, there is a tradeoff between finding the appropriate experts who have the expertise and organizational position; and finding panelists who have sufficient time to participate in the complete study.

The following criteria have been proposed in the literature for the identification and selection of experts and formation of the expert panels [49, 142–144]

- Experience in the subject/field under consideration
- Reputation in the subject/field under consideration
- Interest and willingness to participate in the study
- Availability for the project
- Publications in the field of interest
- Experts should represent a great diversity within the relevant discipline
- Familiarity with uncertainty concepts

- Balanced viewpoint in a group to compensate for individual biases on the outcome
- Absence of evident conflicts among the panel members
- Absence of forceful dominators by position and personality.

Meyer and Booker state that the expert panels are frequently criticized because sometimes answers of the experts are skewed and it generally happens if the majority of the experts are selected from one place or one organization [132]. Therefore, it is important to select a balance group of experts from the government, universities, research institutes, regulatory agencies, and various segments of industry. It will ensure that members in the expert panel have diverse backgrounds and they cover all segments of an industry or sector.

8 Inconsistencies in Expert Judgments

Consistency is the degree to which an individual is consistent in his/her own judgment. Inconsistency describes a situation where expert judgments change over time. In this section it is described how to verify consistency of responses obtained from a group of experts and consistency of responses of individual expert.

8.1 Consistency of Expert Panel

Consistency of responses between successive rounds of a Delphi study is also referred as stability [57, 145]. If most members in an expert panel choose the same reply in two consecutive rounds then it is considered that consistency or stability has been achieved. Kastein et al. state that consistency of responses of panelists over consecutive Delphi rounds reflects high reliability [146]. Dajani, Sincoff, and Talley recommend to conduct Chi square (χ^2) test to measure whether stability of group response has been achieved or not. Responses of two successive rounds obtained from a group of experts are required to conduct this test. The following two hypotheses are tested to determine whether Delphi rounds and response categories are independent [145]:

H_0: The Delphi rounds are independent of the responses obtained in them.
H_1: The Delphi rounds are not independent of the responses obtained in them.

If the null hypothesis (H_0) is supported by an appropriate statistical test, it can be concluded that there is no dependency relationship between round and responses. If, on the other hand, the alternative hypothesis (H_1) is found to be true, a dependency could be concluded and the response frequencies between rounds could be construed as different. If the null hypothesis is found to be true, it indicates that group stability has been achieved between the Delphi rounds, and the study may be terminated. However, if H_1 is found to be true then we conclude the opposite and it is required to pursue the study in further rounds.

In order to conduct Chi square (χ^2) the observed and expected frequencies of occurrence of particular types of responses in each of the two rounds being tested must be determined first. The required Chi square statistic for testing the above hypotheses can then be calculated from the following Eq. 1 [145]:

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{1}$$

where,

O_{ij} Observed frequency for response interval j in round i

E_{ij} Expected frequency for response interval j in round $i + 1$

i Delphi round i

j Response interval j

The observed frequencies are readily available from the raw Delphi data, and the expected frequencies are determined under the assumption that the null hypothesis is true. So the expected frequency that would occur in each cell if the null hypothesis was true would be the average of the two frequencies observed in the two consecutive rounds.

8.2 Consistency of Individual Expert

Chaffin and Talley argue that it is more important to establish individual consistency than determining group consistency [57]. They extended the work done by Dajani et al. and empirically demonstrated that individual stability does imply group stability; whereas, group stability does not necessarily imply individual stability [57]. So Chaffin and Talley proved that the individual stability test provides more information than the group stability test for measuring consistency of responses between successive rounds of a Delphi study and recommended to use individual stability test whenever possible. They refer stability to the consistency of responses between successive rounds of a Delphi study [57]. The Eq. 2 is used to calculate individual stability in two consecutive Delphi rounds [57]:

$$\chi^2 = \sum_{k=1}^{n} \sum_{j=1}^{m} \frac{(O_{jk} - E_{jk})^2}{E_{jk}} \tag{2}$$

where O_{jk} and E_{jk} are the observed and expected frequency indicating the number of respondents who voted for the *j*th response interval in the *i*th round but voted for the *k*th response interval in round $i + 1$.

m and n are number of non-zero response intervals in the round *i* and round $i + 1$

In order test individual stability in Delphi studies, we need to determine whether there is a significant difference between individual responses in two consecutive rounds using the same Chi square test. The following hypotheses are tested again:

H_0 Individual responses of rounds *i* and $i + 1$ are independent.
H_1 Individual responses of rounds *i* and $i + 1$ are not independent.

8.3 Index of Individual Stability

Chaffin and Talley also suggest to use the "index of predictive association" to measure the extent of individual stability between two consecutive Delphi rounds [57]. The index of predictive association is first proposed by Guttman and later on further developed by Goodman and Kruskal, so it is also called Goodman–Kruskal index of predictive association [147]. It is used to measure the extent of association between two attributes. In case of expert judgment it is used to measure the ability to predict response of round $i + 1$ from those of the *i*th round. The index gives the proportional reduction in the probability of error in predicting the responses of round $i + 1$ given that the responses of the *i*th round have been specified. Chaffin and Talley further state that if value of the index is zero; then there is no predictive association and if the index is 1.00, there is complete predictive association [57].

The Eq. 3 is proposed to measure the degree of individual stability between two consecutive Delphi rounds [57]:

$$\mathrm{I} = \frac{\sum max\ O_{jk} - max\ O_{.k}}{n - max\ O_{.k}} \tag{3}$$

where,

O_{jk} Number (or frequency) of respondents who voted for the *j*th response interval in the *i*th round but voted for the *k* th response interval in round $i + 1$

$max\ O_{jk}$ Largest frequency for the *j*th response interval at the *i*th round

$max\ O$ Largest total frequency among the *k*th response intervals at round $i + 1$

n total observed frequencies

A high value of index indicates that there is relatively high predictive association and individual stability has been achieved. If the value of index is 0, then there is no predictive association. Its value range between 0 and 1.

Thus, it is concluded that Chi square test indicates stability, but does not indicate about the degree or extent of association; whereas, the index of predictive association also measures the degree of association (or consistency) and it can provide a valuable adjunct to the Chi square test.

9 Disagreement Among Experts

Disagreement is the extent to which members in an expert panel are in difference to each other in their judgments. There is a misperception and some people assume that experts will always reach same the conclusion if they are given the same data and if the experts conclude differently, they consider that judgment is questionable [2]. However, this concept is misleading for expert judgment due to two reasons [2]. First reason is that experts do not possess the same knowledge even if they possess same background information and data of a problem. Due to different background and professional experience each expert differs in their expertise and knowledge. If we assume that experts have same knowledge even then they will use it in different ways and may come up with different judgment. Second reason is that usually expert judgment is obtained in uncertain situations, where no clear standards or well developed theories exist. Therefore, the expert judgment may leads to some disagreement among their opinions.

The experts may disagree because they think differently about a problem especially when confronted with a multidisciplinary problem having scientific complexity and uncertainty, it is quite possible that competent, honest and disinterested experts may arrive at different conclusions [148]. Difference among the expert brings different perspectives on the problem under consideration and research indicates that combining different answers from the experts brings better chance of covering the right solution [2].

Torrance argues that the more effective groups are characterized by greater participation and wider divergence of expressed judgment [52]. Disagreement within the group will increase the range of judgments considered in making a decision which increases the accuracy of decisions [52]. Shanteau state that sometime due to disagreements, the experts increase their understanding of the subject [138].

Mumpower and Stewart describe that the following factors may cause differences in expert judgment [148]:

- Different problem definitions
- Poor quality or missing feedback

- Poor quality, missing, or different information available
- Causal texture of the environment.

In Delphi study, difference of opinions also occurs due to level of expertise, selection procedure of respondents, clarity of questions, complexity of issue under consideration, and criteria for iteration [57, 145].

9.1 Levels of Agreement

Dajani, Sincoff and Talley state five levels of agreement in Delphi studies: consensus, majority, bipolarity, plurality, and disagreement, and these level are explained in Table 2 [145]:

It is recommended to terminate the study when the consensus or majority is achieved with stability [145]. For the bipolarity, it is recommended to determine the nature of the stability among the two bipolar groups and terminate or rewrite the particular question. When the plurality occurs and stability is established, it is recommended to terminate the study or administer a new round of questions if stability is not established. When the disagreement occurs and stability is achieved for a given question, the decision must be made as to whether to terminate or rephrase the question statement [145].

9.2 Measure of Disagreement

Graham, Regehr and Wright propose to use Cronbach's alpha (α) to measure and determine consensus among members of an expert panel [67]. They define the concept of consensus within a group of experts as a condition of homogeneity or consistency of opinion among the panelists [67]. Cronbach's alpha is a statistical index used to quantify the reliability of a summation of entities which are panelists

Table 2 Levels of agreement in Delphi studies [145]

Levels of agreement	Description	Example (in a Delphi study with 20 participants)
Consensus	Occurs when unanimity is achieved concerning any given issue	Unanimity among all 20 participants
Majority	Occurs when more than 50 % of the respondents exhibit consistency	With 11–19 participants responding the same
Bipolarity	Bipolarity occurs when respondents are equally divided over an issue	With a 10–10 split on an issue
Plurality	Occurs when a larger portion of the respondents (but less than 50 %) reach agreement	With the largest subgroup of respondents between 2 and 9
Disagreement	Occurs when each respondent maintain views independent of each other	Every respondent in a different subgroup

in this case [149]. Cronbach's alpha is usually used to measure internal consistency and its value varies between 0 and 1. Its value in the range of 0.7–0.8 are considered satisfactory, indicating acceptable level of consensus among the experts [149]. However, for the clinical applications, much higher value of Cronbach's alpha is desired and 0.90 or greater value is considered appropriate [67, 149]. Thus, a higher value of Cronbach's alpha, closer to 1 suggests consensus and agreement in the expert panel. Graham, Regehr and Wright argue that Cronbach's alpha is a better measure than the Spearman-Brown formula [67]. Cronbach's alpha estimates the reliability of the sum of panelists responses using the Eq. 4 [67, 149]:

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum \sigma_{yi}^2}{\sigma_x^2}\right) \tag{4}$$

where,

K number of panelists

σ_{yi}^2 The variances of each individual panelist responses

σ_x^2 The variance of the sum of responses for each individual panelist.

Graham et al. further state that the Cronbach's alpha is used to measure agreement and consensus among the Delphi panelists [67].

In a Delphi study involving 456 experts, Celiktas and Kocar used standard deviation to measure the variability or dispersion of expert judgment of the participants [59]. A low standard deviation indicates that the data points tend to be very close to the same value (the mean), while a high standard deviation indicates that the data are spread out over a large range of values. Thus, it is an appropriate approach for large number of experts.

In another Delphi study Terrados, Almonacid, and Pérez-Higueras recommended to use Variation Factor (v) in order to gauge disagreement and it can be calculated using Eq. 5 [150]:

$$\text{Variation Factor} = v = \frac{\sigma}{\mu} \tag{5}$$

where, σ is the sample standard deviation and μ is the sample arithmetic mean. In their study, Terrados et al. found that the consensus among the experts increased in the second Delphi round and value of variation factor decreased [150].

Kastein et al. used the intraclass correlation coefficient (ICC) to evaluate reliability of a Delphi project conducted to develop evaluation criteria for the performance of family physicians in the Netherlands [146]. They state that reliability of Delphi can be evaluated in a more accurate and effective way by means of ICC when numerical ratings of respondents are available and these ratings are normally distributed [146]. ICC quantifies the reliability of the results of an instrument, its value varies between 0 and 1 indicating degree of agreement among the experts and higher value indicates a higher level of agreement among the experts [151, 152].

9.3 Reducing Disagreement Among Delphi Participants

There is a crucial difference between Delphi and a traditional panel, where consensus is desired and may even be forced in some cases [56]. Coates emphasizes that "*the Delphi is not in reporting high reliability consensus data, but rather in alerting the participants to the complexity of issues, by forcing, cajoling, urging, luring them to think, by having them challenge their assumptions*" [153].

It has been observed that greater concern is given toward agreement among the experts in studies related to healthcare, medical diagnostics, and risk and safety assessment. However, for scenario development, technology foresight, and technology roadmapping it is encouraged to have diverse input from the experts. In this case the objective of the work is to explore a variety of potential futures in order to allow the stakeholders, to prepare for this variety and contribute to shape the desired outcome.

Rohrbaugh describes feedback of outcome after every Delphi round is the compelling force toward reducing the disagreement among a group of experts [1]. Kastein et al. recommend that through a standardized expert selection process, group size, background information, proper design of the questionnaires, and the provision of feedback; disagreement can be reduced among the experts [146]. Thus, through careful selection of the expert panel members, providing adequate background information, making clear questionnaires without any ambiguity, conducting multiple rounds, and giving proper feedback can significantly reduce the disagreement among the experts.

9.4 Disagreement Among FCM Expert Panel

Sometimes causal maps and Fuzzy Cognitive Maps (FCM) are created for scenario planning. There are certain techniques to measure disagreement among the members of FCM expert panel. Multiple FCMs can be combined together to develop an integrated FCM. The integrated FCM captures and summarizes the judgment of all expert panel members in one map for further analysis [154]. The combined FCM is considered stronger and robust than an individual FCM because it possess information obtained from multiplicity of sources [155].

During this process of obtaining causal maps from multiple experts, it is quite possible that experts will differ in content of the map, weight or direction of causal link when addressing the same problem. Taber and Siegel proposed a method to combine multiple FCMs based on calculating the expert credibility weights of every expert. They conducted a study to explore the estimation of expert credibility weights in FCM and proposed a methodology [155, 156]. Their approach is based on calculating Hamming distance (bit difference) between inferences from various experts. The following two assumptions are made for developing this method:

- Concurrence of an expert with the others implies a high level of expertise
- The maps contain a sizeable measure of expertise.

The combined FCM (CFCM) is calculated using the Eq. 6 [155–157]:

$$CFCM = \sum_{i=1}^{NE} FCM_i \, . \, W_i \tag{6}$$

where,

NE Number of experts
W_i Credibility weight of expert i
FCM_i FCM matrix of expert i.

The step by step procedure to calculate the expert credibility weight is described below [155, 156]:

1. Generate 500 hundred random stimuli i.e. input vectors, initial condition vectors. For n number of concepts, the input vector is 1 *by* n.
2. Excite each FCM matrix from these input vectors and get inference. For n number of concepts, the input vector is 1 *by* n, the FCM matrix is n *by* n, and the output vector is 1 *by* n. The FCM adjacency matrix is excited through an input vector by multiplying input vector with the FCM matrix. The new output vector (first inference) is again multiplied with the FCM matrix to get the second inference vector. This process is further repeated until fourth inference is obtained.

Hamming distance between the inferences generated by expert i and k is calculated and stored in a matrix. Hamming distance between two vectors is the number of bit difference, for example for the Hamming distance between two vectors (010100) and (011101) is 2. Hamming matrix ($H_{q,i,k}$) is symmetric with zero diagonal. One Hamming matrix is developed for each stimulus i.e. input vector.

3. Next step is to make a sum these 500 Hamming matrices according to the Eq. 7:

$$M_{i,k} = \sum_{q=1}^{NQ} \frac{H_{q,i,k}}{NQ} \tag{7}$$

where NQ is the number of questions and $M_{i,k}$ indicates the average distance between expert i and k over all questions. This is also a symmetric matrix with zero diagonal.

4. Then calculate the total Hamming distance of each expert. Hamming distance of expert i can be obtained from the Eq. 8:

$$h_i = \frac{\sum_{k=1} M_{i,k}}{(NE - 1)} \quad \text{for } 1 \le i, k \le NE \tag{8}$$

where NE is the number of experts and h_i is the total Hamming distance of expert i. This is the sum of all columns of row i of the collapsed matrix. The smaller value of h_i indicates that response of expert i is closer to the mean response.

5. Last step is to calculate credibility for each FCM. So the weight of expert *i* is calculated using Eq. 9.

$$W_i = 1 - \frac{h_i}{\sum_{i=1} h_i} \qquad \text{for } 1 \le i \le NE \tag{9}$$

where W_i is the *credibility* for expert *i* or credibility for *FCMi* proposed by expert *i*. Then combined FCM is calculated using the Eq. 6.

It is pertinent to mention that this method is based on the principle that the expert who differs and disagrees from the majority of experts will be given a lesser expert credibility weight while combining the judgment of several experts. So there is an important question that should we give a lesser weight to expert who disagrees from the majority, when combining the judgment of several experts?

Scenario planning literature highlights the importance of identifying the weak signals and future surprises. If we apply the approach proposed by Taber and Siegel, it will further suppress any weak signals if highlighted by a few experts. Thus, the expert credibility weight approach is suitable for healthcare sector where it is critical to achieve consensus; such as development of a diagnostic criteria for a disease. However, it is not a suitable approach for combining multiple FCMs developed for scenario planning.

Another method is commonly employed to combine multiple FCMs and it takes an average of the expert opinions expressed in all FCMs to generate integrated FCM [154, 158, 159]. This approach does not measure disagreement among the experts, but is recommended for combing multiple FCMs for the purpose of scenario planning [160].

10 Consistency, Reliability and Validity of the Delphi Method

It is important that reliable and accurate methods are used in a research. Reliability measures that a particular technique is applied repeatedly to same object and yields the same results each time [161]. Whereas, research validity means that we are actually measuring what we intend to measure or it is the extent to which a measure actually measures a trait [161].

The results of Delphi studies are generally considered to be reliable and various studies have demonstrated the validity and accuracy of the Delphi technique [49, 141, 162]. Reliability and accuracy of Delphi can be evaluated by comparing two or more Delphi studies on the same subject [54, 163]. Ono and Wedemeyer assessed the accuracy and validity of the forecasts derived from a classic Delphi study exercise [162]. They compared expert judgment of 24 trends and 17 events in the communication field with the forecasts made by utilizing the Delphi technique 16 years earlier and results indicated that earlier forecasts were significantly correlated with the present day trend assessment [162]. Some researchers consider this study by Ono and Wedemeyer as a valid proof of the validity of the Delphi method [141, 164].

Martino also provided evidence of consistency and accuracy of expert opinions through Delphi approach [163]. He investigated the consistency of Delphi studies on a same subject by different expert panels and made comparisons of published Delphi forecasts [163]. The results of these Delphi studies were compared and standard deviation of the differences were computed. It was found that for long range forecasts, the standard deviation was ranging from approx. 3.5 to 2.5 years which is considered a high degree of consistency [163]. Thus, based on this analysis and comparison, Martino concluded that the forecasts produced by the Delphi method are quite consistent and different panels tend to produce similar results [163]. He further states that it is unlikely that two panels of equally competent experts will produce significantly different forecasts.

The Delphi method is also criticized by some researchers due to lack of clear evidence of validity and reliability of the output of this approach [54, 72]. However, it is also not very clear in the literature as how to establish reliability and validity of the Delphi study [66, 165]. Landeta emphasizes that by effectively carrying out the Delphi process including careful selection of the experts, formulation of questions, and processing of data, significantly helps to achieve high levels of reliability and validity for a study [135]. Hill et al. and van Zolingen et al. also state that reliability is principally achieved through the standardization of research procedures [143, 166].

The literature review highlights that results of Delphi studies convergence after successive Delphi rounds and the general trend is toward more valid judgments over iterations [167, 168]. Thus, the feedback after every Delphi round increases the convergence of the individual judgments and overall validity of the individual expert is improved due to this information exchange [49]. Dalkey, Brown, and Cochran also found that the feedback improves the accuracy of the group estimates [169]. Moreover, due to anonymity, experts change their opinions due to objective and rational reasons without any pressure.

Reliability is usually defined in terms of the precision of measurement instruments and dependability of measurement across different replications [143]. There cannot be good scientific results without reliability [143]. Woudenberg state that factors such as the skills of the group leader, motivation of the participants, and quality of the instructions and questionnaires affect the validity of Delphi approach

[54]. Moreover factors like the number of experts and their expertise level are also very important [166].

A study was conducted to investigate the relationship between expert panel size and panel reliability and it was found that the mean correlation between the median and the true answer increases with increasing number of experts [48]. Correlation between actual measurements and expert opinions exceeds 0.70 when the panel size is 11 and no further change occurs after the panel size exceeds 15 [48]. Therefore, we can conclude that 11–15 is an optimum size of an expert panel.

During Delphi study long time intervals between one round and the next round could distort the research and dishearten the participants; thus, shortening this turnaround time periods improves reliability of the research [170]. Therefore, online Delphi approach is considered a better option because it shortens the overall time period of the study [58, 59].

Landeta et al., Hill et al. and van Zolingen et al. highlight that the reliability of expert judgment study depends upon the following factors [143, 166, 170]:

- Quality and expertise of the members of an expert panel
- Time elapsing between Delphi rounds
- Proper administration of the questionnaire and the feedback
- Ensure clear and standardized instructions without any ambiguity
- Clarity of questions
- Consensus/convergence of opinions
- Stability of the results between consecutive rounds.

As mentioned earlier that despite some criticism, it is evident from research that Delphi technique yields better predictions and accurate forecasts than another normal interacting group or a face-to-face committee meeting [68–70]. Landeta analyzes the published literature on Delphi approach over time and argue that its extensive use in scientific publications and doctoral theses indicates the validity of this method [135, 170]. It further highlights that the scientific community considers it as a valid technique for obtaining and processing subjective information from the experts, whether used alone or in combination with some other technique [165, 170]. Landeta also cites that large number of dissertations have used the Delphi method and according to data from ProQuest database, it has been used in 1668 dissertations during the time period from 1970 to 2004 [135]. It highlights the importance, usefulness and validity of Delphi technique for research.

11 Selection of Appropriate Research Method

It is critical to choose an appropriate technique for a particular application in a research project [171, 172]. Levary and Han highlight that selection of a technological forecasting method depends on factors such as: stage of technology development, degree of similarity between proposed and existing technologies, number of forecasting variables, extent of data availability, data validity, and technological uncertainty [172]. Chambers et al. highlight that selection of a research method also depends on factors like the context of the research, relevance and availability of historical data, degree of desirable accuracy, and the time available for making the analysis [171].

Based on an in depth analysis of nine case studies, Levary and Han conclude that Delphi method and scenario writing approaches are suitable in situations where there is no or low level of similarity between the proposed technology and existing technologies, a medium number of variables affect technological development, less data available, and low or medium degree of data validity requirements [172]. They further elaborate that in these circumstances it will be a reasonable choice to use a method based upon obtaining information from an expert panel and employ Delphi method and/or scenario planning approach [172].

Rowe, Wright, and Bolger argue that Delphi allows the experts to make meaningful judgments, particularly in cases where a variety of factors (economic, technical, political and so on) affect the problem under consideration and it gives an opportunity to each expert to derive benefits from others experts having diverse knowledge and background [49]. It is very useful approach especially when no historical data exists for judgment [49]. The Delphi method is also very useful when it is difficult to bring experts together due to time or cost constraints [47]. Linstone and Turoff state that sometimes it is necessary to benefit from subjective judgments on a collective basis because due to the peculiarity of a problem it is difficult to address it with a precise analytical technique [47].

Delphi is a powerful tool to engage stakeholders and it is an appropriate approach when there are many stakeholders because it is not possible to gather everyone in one place [61]. Woudenberg arranged the expert judgment methods in order of increasing accuracy based on a comprehensive literature review and found that the Delphi method is the most accurate technique among various expert judgment methods [54]. Various other researchers have also indicated that Delphi approach generate accurate and reliable judgment than other techniques [68–70, 163].

References

1. Rohrbaugh J (1979) Improving the quality of group judgment: social judgment analysis and the Delphi technique. Organ Behav Hum Perform 24:73–92
2. Meyer MA, Booker JM (1991) Eliciting and analyzing expert judgment: a practical guide. Academic Press Limited, London
3. Keeney RL, von Winterfeldt D (1989) On the uses of expert judgment on complex technical problems. IEEE Trans Eng Manage 36:83–86
4. Durance P, Godet M (2010) Scenario building: uses and abuses. Technol Forecast Soc Chang 77:1488–1492
5. Bradfield R, Wright G, Burt G, Cairns G, Van Der Heijden K (2005) The origins and evolution of scenario techniques in long range business planning. Futures 37:795–812
6. Cooke RM (1991) Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, New York

7. Simola K, Mengolini A, Bolado-Lavin R (2005) Formal expert judgment: an overview. European Commission, Directorate General Joint Research Centre (DG JRC), Institute for Energy. http://ie.jrc.ec.europa.eu/publications/scientific_publications/2005/EUR21772EN.pdf. Accessed 18 March 2011
8. Börjeson L, Höjer M, Dreborg K-H, Ekvall T, Finnveden G (2006) Scenario types and techniques: towards a user's guide. Futures 38:723–739
9. Amer M, Daim TU (2010) Application of technology roadmaps for renewable energy sector. Technol Forecast Soc Chang 77:1355–1370
10. Gerdsri N (2005) An analytical approach on building a technology development envelope (TDE) for roadmapping of emerging technologies. Ph.D. Dissertation, Systems Science: Engineering Management, Portland State University, Portland
11. Gerdsri P (2009) A systematic approach to developing national technology policy and strategy for emerging technologies. Ph.D. Dissertation, Engineering and Technology Management (ETM) Department, Portland State University, Portland
12. Schwartz P (1996) The Art of the Long View: Planning for the Future in an Uncertain World. Currency Doubleday, New York
13. Chermack TJ, Lynham SA, Ruona WEA (2001) A review of scenario planning literature. Futures Res Q 17:7–31
14. Kahn H, Wiener AJ (1967) The year 2000: a framework for speculation on the next thirty-three years. The Macmillan, New York
15. Varum CA, Melo C (2007) Strategic planning in an uncertain business environment: the diffusion of scenario planning. In: Conference Factores de Competitividade, Competitiveness Factors: A Portuguese Perspective, Aveiro
16. Mietzner D, Reger G (2005) Advantages and disadvantages of scenario approaches for strategic foresight. Int J Technol Intell Planning 1:220–239
17. Varum CA, Melo C (2010) Directions in scenario planning literature: a review of the past decades. Futures 42:355–369
18. Van Der Heijden K (1996) Scenarios: the art of strategic conversation. Wiley, Chichester
19. Rigby D, Bilodeau B (2007) Selecting management tools wisely. Harv Bus Rev 85:20–22
20. Porter AL, Roper AT, Mason TW, Rossini FA, Banks J (1991) Forecasting and management of technology, 1st edn. Wiley, New York
21. Saritas O, Oner MA (2004) Systemic analysis of UK foresight results: joint application of integrated management model and roadmapping. Technol Forecast Soc Chang 71:27–65
22. Chermack TJ, Lynham SA, van der Merwe L (2006) Exploring the relationship between scenario planning and perceptions of learning organization characteristics. Futures 38:767–777
23. Schoemaker PJH, van der Heijden CAJM (1992) Integrating scenarios into strategic planning at royal Dutch/Shell. Strategy Leadersh 20:41–46
24. Joseph CF (2000) Scenario planning. Technol Forecast Soc Chang 65:115–123
25. Hiltunen E (2009) Scenarios: process and outcome. J Futures Stud 13:151–152
26. Linneman RE, Klein HE (1979) The use of multiple scenarios by U.S. industrial companies. Long Range Plan 12:83–90
27. Linneman RE, Klein HE (1983) The use of multiple scenarios by U.S. industrial companies: a comparison study, 1977–1981. Long Range Plan 16:94–101
28. Amer M, Daim TU, Jetter A (2013) A review of scenario planning. Futures 46:23–40
29. Blomgren H, Jonsson P, Lagergren F (2011) Getting back to scenario planning: strategic action in the future of energy Europe. In: 8th international conference on the European energy market (EEM), Zagreb, pp 792–801
30. Silberglitt R, Hove A, Shulman P (2003) Analysis of the US energy scenarios: meta-scenarios, pathways, and policy implications. Technol Forecast Soc Chang 70:297–315
31. Czaplicka-Kolarz K, Stanczyk K, Kapusta K (2009) Technology foresight for a vision of energy sector development in Poland till 2030. Delphi survey as an element of technology foresighting. Technol Forecast Soc Chang 76:327–338

32. Di W, Rui N, Hai-ying S (2011) Scenario analysis of China's primary energy demand and CO2 emissions based on IPAT model. Energy Procedia 5:365–369
33. Keles D, Möst D, Fichtner W (2011) The development of the German energy market until 2030—a critical survey of selected scenarios. Energy Policy 39:812–825
34. Zhou N, Lin J (2008) The reality and future scenarios of commercial building energy consumption in China. Energy Build 40:2121–2127
35. Winebrake JJ, Creswick BP (2003) The future of hydrogen fueling systems for transportation: an application of perspective-based scenario analysis using the analytic hierarchy process. Technol Forecast Soc Chang 70:359–384
36. Sørensen B, Hauge Petersen A, Juhl C, Ravn H, Søndergren C, Simonsen P, Jørgensen K, Henrik Nielsen L, Larsen HV, Erik Morthorst P, Schleisner L, Sørensen F, Engberg Pedersen T (2004) Hydrogen as an energy carrier: scenarios for future use of hydrogen in the Danish energy system. Int J Hydrogen Energy 29:23–32
37. Wietschel M, Hasenauer U, de Groot A (2006) Development of European hydrogen infrastructure scenarios–CO2 reduction potential and infrastructure investment. Energy Policy 34:1284–1298
38. Contaldi M, Gracceva F, Mattucci A (2008) Hydrogen perspectives in Italy: analysis of possible deployment scenarios. Int J Hydrogen Energy 33:1630–1642
39. Antoine B, Goran K, Neven D (2008) Energy scenarios for Malta. Int J Hydrogen Energy 33:4235–4246
40. Weisser D (2004) Costing electricity supply scenarios: a case study of promoting renewable energy technologies on Rodriguez, Mauritius. Renewable Energy 29:1319–1347
41. Chen T-Y, Yu OS, Hsu GJY, Hsu F-M, Sung W-N (2009) Renewable energy technology portfolio planning with scenario analysis: a case study for Taiwan. Energy Policy 37:2900–2906
42. Madlener R, Kowalski K, Stagl S (2007) New ways for the integrated appraisal of national energy scenarios: the case of renewable energy use in Austria. Energy Policy 35:6060–6074
43. Varho V, Tapio P (2005) Wind power in Finland up to the year 2025—'soft' scenarios based on expert views. Energy Policy 33:1930–1947
44. Morales JM, Mínguez R, Conejo AJ (2010) A methodology to generate statistically dependent wind speed scenarios. Appl Energy 87:843–855
45. Rowe G, Wright G (1999) The Delphi technique as a forecasting tool: issues and analysis. Int J Forecast 15:353–375
46. Raubitschek R (1988) Multiple scenario analysis and business planning. In: Lamb R, Shrivastava P (eds) Advances in strategic management, vol 5. JAI Press Inc., London
47. Linstone HA, Turoff M (1975) The Delphi method: techniques and applications. Addison-Wesley, Reading
48. Martino JP (1983) Technological forecasting for decision making, 2nd edn. North-Holland, New York
49. Rowe G, Wright G, Bolger F (1991) Delphi: a reevaluation of research and theory. Technol Forecast Soc Chang 39:235–251
50. Erffmeyer RC, Erffmeyer ES, Lane IM (1986) The Delphi technique: an empirical evaluation of the optimal number of rounds. Group and Organ Manage 11:120–128
51. Lindeman CA (1975) Delphi survey of priorities in clinical nursing research. Nurs Res 24:434–441
52. Torrance EP (1957) Group decision-making and disagreement. Soc Forces 35:314–318
53. Powell C (2003) The Delphi technique: myths and realities. J Adv Nurs 41:376–382
54. Woudenberg F (1991) An evaluation of Delphi. Technol Forecast Soc Chang 40:131–150
55. Jairath N, Weinstein J (1994) The Delphi methodology: a useful administrative approach. Can J Nurs Adm 7:29–42
56. Linstone HA, Turoff M (2011) Delphi: a brief look backward and forward. Technol Forecast Soc Chang 78:1712–1719
57. Chaffin WW, Talley WK (1980) Individual stability in Delphi studies. Technol Forecast Soc Chang 16:67–73

58. Gordon T, Pease A (2006) RT Delphi: an efficient, "round-less" almost real time Delphi method. Technol Forecast Soc Chang 73:321–333
59. Celiktas MS, Kocar G (2010) From potential forecast to foresight of Turkey's renewable energy with Delphi approach. Energy 35:1973–1980
60. Steinert M (2009) A dissensus based online Delphi approach: an explorative research tool. Technol Forecast Soc Chang 76:291–300
61. Geist MR (2010) Using the Delphi method to engage stakeholders: a comparison of two studies. Eval Program Plann 33:147–154
62. Rauch W (1979) The decision Delphi. Technol Forecast Soc Chang 15:159–169
63. Turoff M (1970) The design of a policy Delphi. Technol Forecast Soc Chang 2:149–171
64. Delbecq AL, Van de Ven AH, Gustafson DH (1975) Group techniques for program planning. A guide to nominal group and Delphi processes. Scott, Foresman and Company, Glenview
65. Murphy M, Black N, Lamping D, McKee CM, Sanderson CFB, Askham J (1998) Consensus development methods, and their use in clinical guideline development. Health Technol Assess 2:1–65
66. Pill J (1971) The Delphi method: substance, context, a critique and an annotated bibliography. Socio-Econ Plan Sci 5:57–71
67. Graham B, Regehr G, Wright JG (2003) Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol 56:1150–1156
68. Brockhoff K (1975) The performance of forecasting groups in computer dialogue and face to face discussions. In: Linstone HA, Turoff M (eds) The Delphi method: techniques and applications. Addison-Wesley, Reading
69. Riggs WE (1983) The Delphi technique: an experimental evaluation. Technol Forecast Soc Chang 23:89–94
70. Larreche JC, Moinpour R (1983) Managerial judgment in marketing: the concept of expertise. J Mark Res 20:110–121
71. Coates JF (1975) In defense of Delphi: a review of Delphi assessment, expert opinion, forecasting, and group process by H. Sackman. Technol Forecast Soc Chang 7:193–194
72. Sackman H (1974) Delphi critique: expert opinion, forecasting, and group process. Lexington Books, Lexington
73. Martino JP (2003) A review of selected recent advances in technological forecasting. Technol Forecast Soc Chang 70:719–733
74. Breiner S, Cuhls K, Grupp H (1994) Technology foresight using a Delphi approach: a Japanese German study. R&D Manage 24:141–153
75. Biloslavo R, Dolinek S (2008) Scenario planning for climate strategies development by integrating group Delphi, AHP and dynamic fuzzy cognitive maps. In: Portland International Center for Management of Engineering and Technology (PICMET), Cape Town, South Africa, pp 1103–1111
76. Huss WR, Honton EJ (1987) Scenario planning—what style should you use? Long Range Plan 20:21–29
77. Bengisu M, Nekhili R (2006) Forecasting emerging technologies with the aid of science and technology databases. Technol Forecast Soc Chang 73:835–844
78. Gupta UG, Clarke RE (1996) Theory and applications of the Delphi technique: a bibliography (1975–1994). Technol Forecast Soc Chang 53:185–211
79. Iniyan S, Sumathy K (2000) An optimal renewable energy model for various end-uses. Energy 25:563–575
80. Iniyan S, Sumathy K (2003) The application of a Delphi technique in the linear programming optimization of future renewable energy options for India. Biomass Bioenergy 24:39–50
81. Pätäri S (2010) Industry- and company-level factors influencing the development of the forest energy business—insights from a Delphi study. Technol Forecast Soc Chang 77:94–109

82. Rikkonen P, Tapio P (2009) Future prospects of alternative agro-based bioenergy use in Finland–Constructing scenarios with quantitative and qualitative Delphi data. Technol Forecast Soc Chang 76:978–990
83. Ronde P (2003) Delphi analysis of national specificities in selected innovative areas in Germany and France. Technol Forecast Soc Chang 70:419–448
84. Sharma DP, Nair PSC, Balasubramanian R (2003) Analytical search of problems and prospects of power sector through Delphi study: case study of Kerala state, India. Energy Policy 31:1245–1255
85. Suganthi L, Williams A (2000) Renewable energy in India—a modelling study for 2020–2021. Energy Policy 28:1095–1109
86. Utgikar VP, Scott JP (2006) Energy forecasting: predictions, reality and analysis of causes of error. Energy Policy 34:3087–3092
87. Blind K, Cuhls K, Grupp H (2001) Personal attitudes in the assessment of the future of science and technology: a factor analysis approach. Technol Forecast Soc Chang 68:131–149
88. Cuhls K, Kuwahara T (1994) Outlook of Japanese and German future technology: comparing technology forecast surveys. Phisica-Verlag, Heidelberg
89. Amer M, Daim T (2012) Technology and science policies in transitional economy: a case of Turkey. Sci Technol Soc 17:297–321
90. Gerdsri P Kocaoglu D (2009) A systematic approach to developing national technology policy and strategy for emerging technologies: a case study of nanotechnology for Thailand's agriculture industry. In: Portland International Center for Management of Engineering and Technology (PICMET), Portland, pp 447–461
91. Chakravarti AK, Vasanta B, Krishnan ASA, Dubash RK (1998) Modified Delphi methodology for technology forecasting case study of electronics and information technology in India. Technol Forecast Soc Chang 58:155–165
92. Shin T (1998) Using Delphi for a long-range technology forecasting, and assessing directions of future R&D activities the Korean exercise. Technol Forecast Soc Chang 58:125–154
93. Tichy G (2004) The over-optimism among experts in assessment and foresight. Technol Forecast Soc Chang 71:341–363
94. Daim T, Yates D, Peng Y, Jimenez B (2009) Technology assessment for clean energy technologies: the case of the Pacific northwest. Technol Soc 31:232–243
95. Saaty TL (2003) Decision-making with the AHP: why is the principal eigenvector necessary. Eur J Oper Res 145:85–91
96. Wang J–J, Jing Y–Y, Zhang C-F, Zhao J-H (2009) Review on multi-criteria decision analysis aid in sustainable energy decision-making. Renew Sustain Energy Rev 13:2263–2278
97. Salo A, Gustafsson T, Ramanathan R (2003) Multicriteria methods for technology foresight. J Forecast 22:235–255
98. Elkarmi F, Mustafa I (1993) Increasing the utilization of solar energy technologies (SET) in Jordan analytic hierarchy process. Energy Policy 21:978–982
99. Forman EH, Gass SI (2001) The analytic hierarchy process—an exposition. Operations Research 49:469–486
100. Saaty TL (1980) The analytic hierarchy process: planning, priority setting, resource allocation. McGraw Hill International, New York
101. Gerdsri N, Kocaoglu DF (2007) Applying the analytic hierarchy process (AHP) to build a strategic framework for technology roadmapping. Math Comput Model 46:1071–1080
102. Kocaoglu DF (1983) A participative approach to program evaluation. IEEE Trans Eng Manage 30:112–118
103. Kablan MM (2004) Decision support for energy conservation promotion: an analytic hierarchy process approach. Energy Policy 32:1151–1158
104. Ahsan MK, Bartlema J (2004) Monitoring healthcare performance by analytic hierarchy process: a developing-country perspective. Int Trans Oper Res 11:465–478

105. Yates D, Jimenez BT, Peng Y (2007) Portland general electric (PGE): clean power generation wind project in biglow canyon boardman coal plant. In: PICMET, Portland, pp 2530–2549
106. Akash BA, Mamlook R, Mohsen MS (1999) Multi-criteria selection of electric power plants using analytical hierarchy process. Electric Power Syst Res 52:29–35
107. Lee SK, Yoon YJ, Kim JW (2007) A study on making a long-term improvement in the national energy efficiency and GHG control plans by the AHP approach. Energy Policy 35:2862–2868
108. Lee SK, Mogi G, Kim JW (2008) The competitiveness of Korea as a developer of hydrogen energy technology: the AHP approach. Energy Policy 36:1284–1291
109. Lee SK, Mogi G, Kim JW, Gim BJ (2008) A fuzzy analytic hierarchy process approach for assessing national competitiveness in the hydrogen technology sector. Int J Hydrogen Energy 33:6840–6848
110. Lee AHI, Chen HH, Kang H-Y (2008) Multi-criteria decision making on strategic selection of wind farms. Renewable Energy 34:120–126
111. Aras H, Erdogmus S, Koc E (2004) Multi-criteria selection for a wind observation station location using analytic hierarchy process. Renewable Energy 29:1383–1392
112. Jaber JO, Jaber QM, Sawalha SA, Mohsen MS (2008) Evaluation of conventional and renewable energy sources for space heating in the household sector. Renew Sustain Energy Rev 12:278–289
113. Lee TL, Lin HM, Jeng DS, Hsu TW (2008) Application of fuzzy analytic hierarchy process to assess the potential of offshore wind energy in Taiwan. In: Proceedings of the international offshore and polar engineering conference. Vancouver, pp 461–465
114. Pilavachi PA, Chatzipanagi AI, Spyropoulou AI (2009) Evaluation of hydrogen production methods using the analytic hierarchy process. Int J Hydrogen Energy 34:5294–5303
115. Tsoutsos T, Drandaki M, Frantzeskaki N, Iosifidis E, Kiosses I (2009) Sustainable energy planning by using multi-criteria analysis application in the island of Crete. Energy Policy 37:1587–1600
116. Wijayatunga PDC, Siriwardena K, Fernando WJLS, Shrestha RM, Attalage RA (2006) Strategies to overcome barriers for cleaner generation technologies in small developing power systems: Sri Lanka case study. Energy Convers Manage 47:1179–1191
117. Chatzimouratidis AI, Pilavachi PA (2008) Multicriteria evaluation of power plants impact on the living standard using the analytic hierarchy process. Energy Policy 36:1074–1089
118. Chatzimouratidis AI, Pilavachi PA (2009) Technological, economic and sustainability evaluation of power plants using the analytic hierarchy process. Energy Policy 37:778–787
119. Clarke RR (1997) Validation and legitimation of an analytic hierarchy approach to integrated resource planning for electric utilities. In: Energy conversion engineering conference (IECEC-97), Honolulu, pp 2197–2201
120. Afgan NH, Carvalho MG (2002) Multi-criteria assessment of new and renewable energy power plants. Energy 27:739–755
121. Amer M, Daim TU (2011) Selection of renewable energy technologies for a developing country: a case study of Pakistan. Energy Sustain Dev 15:420–435
122. Millett SM, Honton EJ (1991) A manager's guide to technology forecasting and strategy analysis methods. Battelle Press, Columbus
123. Rohrbaugh J (1981) Improving the quality of group judgment: social judgment analysis and the nominal group technique. Organ Behav Hum Perform 28:272–288
124. Herbert TT, Yost EB (1979) A comparison of decision quality under nominal and interacting consensus group formats: the case of the structured problem. Decis Sci 10:358–370
125. Islam R (2001) Modification of the nominal group technique by using the analytic hierarchy process. In: Multiple Criteria Decision Making in the New Millennium, vol 507, pp 294–303
126. Henderson NR (2009) Managing moderator stress: take a deep breath. You can do this! Mark Res 21:28–29

127. Sutton RI, Hargadon A (1996) Brainstorming groups in context: effectiveness in a product design firm. Adm Sci Q 41:685–718
128. Miles I, Keenan M (2002) Practical guide to regional foresight in the United Kingdom. Directorate-General for Research, European Commission. http://cordis.europa.eu/foresight/cgrf.htm. Accessed 14 March 2011
129. Keeney RL, von Winterfeldt D (1991) Eliciting probabilities from experts in complex technical problems. IEEE Trans Eng Manage 38:191–201
130. Cojazzi G, Fogli D, Grassini G, De Gelder P, Gryffroy D, Bolado R, Hofer E, Virolainen R, Coe IM, Bassanelli A, Puga J, Papazoglou I, Zuchuat O, Cazzoli E, Eyink J, Guida G, Pinola L, Pulkkinen U, Simola K, von Winterfeldt D, Valeri A (2001) Benchmark exercise on expert judgment techniques in PSA Level 2. Nucl Eng Des 209:211–221
131. Cojazzi G, Fogli D (2000) Benchmark exercise on expert judgment techniques in PSA level 2, extended final report. http://ie.jrc.ec.europa.eu/publications/scientific_publications/2005/EUR21772EN.pdf
132. Meyer MA, Booker JM (1991) Selecting and motivating the experts. In: Eliciting and analyzing expert judgment: a practical guide, pp 85–98, 1st edn. Academic Press Limited, London
133. Camerer CF, Johnson EJ (1991) The process-performance paradox in expert judgment. In: Ericsson KA, Smith J (eds) Toward a general theory of expertise: prospects and limits, 1st edn. Cambridge University Press, Cambridge, pp 195–217
134. McGraw KL, Harbison-Briggs K (1989) Knowledge acquisition: principles and guidelines. Prentice-Hall International, New Jersey
135. Landeta J (2006) Current validity of the Delphi method in social sciences. Technol Forecast Soc Chang 73:467–482
136. Jetter AJ (2006) Elicitation-extracting knowledge from experts. In: Jetter AJ, Kraaijenbrink J, Schroder H-H, Wijnhoven F (eds) Knowledge integration: the practice of knowledge management in small and medium enterprises, pp 65–76. Physica-Verlag, Heidelberg
137. Schaller RR (2004) Technology innovation in the semiconductor industry: a case study of the international technology roadmap for semiconductors (ITRS) Ph.D. Dissertation, School of Public Policy, George Mason University, Fairfax
138. Shanteau J (1992) The psychology of experts: an alternative view. In: Wright, G. Bolger F (eds) Expertise and decision support, Plenum Press, pp 11–23. New York
139. Ascher W (1978) Forecasting: An Appraisal for Policymakers and Planners. Johns Hopkins University Press, Baltimore
140. Mitchell VW (1991) The Delphi technique: an exposition and application. Technol Anal Strat Manage 3:333–358
141. Okoli C, Pawlowsk SD (2004) The Delphi method as a research tool: an example, design considerations and applications. Inf Manage 42:15–29
142. Daim TU (1998) Technology evaluation and acquisition strategies and their implications in the U.S. electronics manufacturing industry. Ph.D. Dissertation, System Science: Engineering and Technology Management, Portland State University, Portland
143. Hill KQ, Fowles J (1975) The methodological worth of the Delphi forecasting technique. Technol Forecast Soc Chang 7:179–192
144. Cooke RM, Goossens LHJ (2000) Procedures guide for structured expert judgement. http://ie.jrc.ec.europa.eu/publications/scientific_publications/2005/EUR21772EN.pdf. Accessed 12 May 2011
145. Dajani JS, Sincoff MZ, Talley WK (1979) Stability and agreement criteria for the termination of Delphi studies. Technol Forecast Soc Chang 13:83–90
146. Kastein MR, Jacobs M, van der Hell RH, Luttik K, Touw-Otten FWMM (1993) Delphi, the issue of reliability: a qualitative Delphi study in primary health care in the Netherlands. Technol Forecast Soc Chang 44:315–323
147. Goodman LA, Kruskal WH (1959) Measures of association for cross classifications. II: further discussion and references. J Am Stat Assoc 54:123–163

148. Mumpower JL, Stewart TR (1996) Expert judgement and expert disagreement. Thinking Reasoning 2:191–211
149. Bland JM, Altman DG (1997) Statistics notes: Cronbach's alpha. Br Med J 314:572
150. Terrados J, Almonacid G, Pérez-Higueras P (2009) Proposal for a combined methodology for renewable energy planning. Application to a Spanish region. Renew Sustain Energy Rev 13:2022–2030
151. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428
152. Wuensch KL (2007) Inter-rater agreement. http://www.ecu.edu/psyc/
153. Coates JF (1975) In defense of Delphi: a review of Delphi assessment: expert opinion, forecasting, and group process by H. Sackman. Technol Forecast Soc Chang 7:193–194
154. Kandasamy WBV, Smarandache F (2003) Fuzzy cognitive maps and neutrosophic cognitive maps. Indian institute of technology, Chennai
155. Taber R (1991) Knowledge processing with fuzzy cognitive maps. Expert Syst Appl 2:83–87
156. Taber W, Siegel M (1987) Estimation of experts' weights using fuzzy cognitive maps. In: IEEE International Conference on Neural Networks, pp 319–326
157. Taber R, Yager RR, Helgason CM (2007) Quantization effects on the equilibrium behavior of combined fuzzy cognitive maps. Int J Intell Syst 22:181
158. Hans-Horst S, Jetter AJM (2003) Integrating market and technological knowledge in the fuzzy front end: an FCM-based action support system. Int J Technol Manage 26:517–539
159. Jetter AJM (2003) Educating the guess: strategies, concepts, and tools for the fuzzy front end of product development. In: Portland International Center for Management of Engineering and Technology (PICMET), Portland, pp 261–273
160. Amer M, Jetter AJ, Daim TU (2011) Development of fuzzy cognitive map (FCM) based scenarios for wind energy. Int J Energy Sect Manage 5:564–584
161. Babbie ER (2007) The practice of social research, 12th edn. Wadsworth, Belmont
162. Ono R, Wedemeyer DJ (1994) Assessing the validity of the Delphi technique. Futures 26:289–304
163. Martino JP (1970) The consistency of Delphi forecasts. Futurist 4:63–64
164. Keeney S, Hasson F, McKenna HP (2001) A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud 38:195–200
165. Engels TCE, Powell Kennedy H (2007) Enhancing a Delphi study on family-focused prevention. Technol Forecast Soc Chang 74:433–451
166. van Zolingen SJ, Klaassen CA (2003) Selection processes in a Delphi study about key qualifications in senior secondary vocational education. Technol Forecast Soc Chang 70:317–340
167. Jolson MA, Rossow GL (1971) The Delphi process in marketing decision making. J Mark Res 8:443–448
168. Brown BB, Helmer O (1964) Improving the reliability of estimates obtained from a consensus of experts, pp 1–13. http://www.rand.org/pubs/papers/P2986
169. Dalkey NC, Brown BB, Cochran SW (1970) The Delphi Method IV: effect of percentile feedback and feed-in of relevant facts. http://www.rand.org/pubs/research_memoranda/RM6118
170. Landeta J, Matey J, Ruíz V, Galter J (2008) Results of a Delphi survey in drawing up the input-output tables for Catalonia. Technol Forecast Soc Chang 75:32–56
171. Chambers JC, Mullick SK, Smith DD (1971) How to choose the right forecasting technique. Harv Bus Rev 49:45–74
172. Levary RR, Han D (1995) Choosing a technological forecasting method, Indus Manag 37:14–18