General Review

Consensus methods: Review of original methods and their main alternatives used in public health

F. Bourrée a, P. Michel a,b,* L.R. Salmi c,d

a EA 3672, laboratoire santé, travail, environnement, ISPED, université Victor-Segalen Bordeaux 2, Bordeaux, France
b Comité de coordination de l’évaluation clinique et de la qualité en Aquitaine, hôpital Xavier-Arnozan, 33604 Pessac cedex, France
c Service d’épidémiologie et biostatistique, centre de recherche Inserm (U987), ISPED, université Victor-Segalen Bordeaux 2, Bordeaux, France
d Service d’information médicale, CHU de Bordeaux, Bordeaux, France

Received 15 February 2008; accepted 6 October 2008
Available online 21 November 2008

Abstract

Background. – Consensus-based studies are increasingly used as decision-making methods, for they have a lower production cost than other methods (observation, experimentation, modeling) and provide results more rapidly. The objective of this paper is to describe the principles and methods of the four main methods – Delphi, nominal group, consensus development conference and RAND/UCLA – their use as reported in peer-reviewed publications and validation studies published in the healthcare literature.

Methods. – A bibliographic search was performed in PubMed/MEDLINE, banque de données santé publique (BDSP), The Cochrane Library, Pascal and Francis. Keywords, headings and qualifiers corresponding to a list of terms and expressions related to the consensus methods were searched for in the thesauri and used in the literature search. A search with the same terms and expressions was performed on Internet using the website Google Scholar.

Results. – All methods, precisely described in the literature, are based on common basic principles such as definition of the subject, selection of experts and direct or remote interaction processes. They sometimes use quantitative assessment for ranking items. Numerous variants of these methods have been described. Few validation studies have been implemented. Not implementing these basic principles and failing to describe the methods used to reach the consensus were both frequent reasons raising suspicion regarding the validity of consensus methods.

Conclusion. – When it is applied to a new domain with important consequences in terms of decision-making, a consensus method should first be validated.

© 2008 Elsevier Masson SAS. All rights reserved.

Résumé

Position du problème. – Par leur coût de production de connaissances souvent inférieur à celui d’autres méthodes (observation, expérimentation, modélisation) et la plus grande rapidité d’obtention des résultats, les études fondées sur des méthodes de consensus constituent des méthodes d’aide à la décision favorisées. L’objectif de cet article est de présenter les principes et la méthode originale des quatre principales méthodes de consensus : Delphi, groupe nominal, conférence de consensus et RAND/UCLA, de décrire leur utilisation au travers des publications et de présenter les travaux de validation de ces méthodes dans le domaine de la santé.

Méthodes. – Une recherche bibliographique a été effectuée dans PubMed/MEDLINE, banque de données santé publique (BDSP), The Cochrane Library, Pascal et Francis. Une liste de termes et expressions relatifs aux méthodes a été établie, les descripteurs correspondants ont été recherchés dans les thésauri et ont servi à réaliser la recherche. La même recherche a été effectuée sur Internet via le moteur de recherche Google Scholar.

* Corresponding author.
E-mail address: philippe.michel@ccceca.asso.fr (P. Michel).

0398-7620/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.respe.2008.10.005
Résumé: — Toutes ces méthodes, précisément décrites dans la littérature, comprennent des principes fondamentaux communs tels que la définition du sujet, la sélection des experts et un processus d’interactions interpersonnelles directes ou à distance. La cotation, pour hiérarchiser les propositions, est parfois utilisée. Il existe de nombreuses variantes aux définitions originelles de ces méthodes mais peu de travaux de validation. Les variantes qui ne respectent pas les principes fondamentaux des méthodes, ou l’absence de description précise du processus d’obtention du consensus dans les publications, contribuent aux interrogations souvent exprimées sur la validité de ces méthodes.

Conclusion: — Lorsqu’elles sont appliquées à un domaine nouveau avec des retombées importantes en termes d’aide à la décision, ces méthodes devraient faire l’objet d’une validation préalable.

© 2008 Elsevier Masson SAS. All rights reserved.

Keywords: Consensus; Delphi technique; Nominal group; Qualitative research; Consensus development conference

Mots clés : Consensus ; Méthode Delphi ; Groupe nominal ; Méthodes qualitatives ; Conférence de consensus

1. Background

Knowledge-production in public health is founded on four approaches: experimentation, observation (whether or not controlling for the determinants of the intervention evaluated), modeling and consensus. A consensus is defined as a general agreement, implicit or explicitly expressed, among the members of a group [1]. Consensus corresponds to an agreement on something, but does not necessarily mean total agreement on everything, i.e., unanimity. Consensus is intended to allow differences to coexist, not to do away with them. As a knowledge-production or decision-making method, consensus emphasizes the importance of each participant and can express a result that is not formally opposed. Only formalized methods, grouped under the name of consensus methods, can limit the biases stemming from the confrontation of actors who have come together to generate or synthesize knowledge.

In public health, these methods are most often used to make decisions, generate ideas or establish rank in situations in which scientifically founded information is lacking or when there is too much contradictory information [2]. Even though these methods are founded on the involvement of a group of participants defined as experts, consensus methods are not reserved only for consulting the scientific authorities of the domain concerned [3–5]. The expert is any person with good practical, legal or administrative knowledge of a precise subject and with sufficient legitimacy to express a representative opinion of the group of actors to which he or she belongs.

The main objectives of this study were to identify the principles describing the four consensus methods used in healthcare (Delphi, nominal group, consensus development conference and RAND/UCLA), to describe the domains in which each of the methods is applied as well as the variants used through articles presenting applications of the methods.

2. Method

2.1. Research strategy

A systematic bibliographic search was done using the PubMed/MEDLINE, banque de données santé publique (BDSP), The Cochrane Library, Pascal, and Francis databases, with no restriction as to date. These databases covered the theme of the subject studied. The articles selected were published either in French or English. Articles that did not center on public health were not retained. We defined the terms and expressions corresponding to the subject of our study as “Delphi”, “nominal group”, “consensus conference” and “RAND UCLA”. Based on the thesauri used in each of the databases, we determined the descriptors corresponding to the terms and expressions defined in our study. In these thesauri, we found no descriptors for the expressions “nominal group” and “RAND UCLA”. The expression “consensus conference” was defined as a type of article in the PubMed/MEDLINE database: consensus development conference.

The articles identified were selected if these terms or expressions appeared in the title or abstract. This restriction allowed us to target the articles describing the consensus methods used and to exclude articles presenting the results of studies that had used a consensus method without detailing it sufficiently in the publication. The same search was carried out in the different databases and the repeats were removed. An Internet search was also done using the Google Scholar search engine with the same terms and expressions. The references cited in the articles selected were studied to identify additional similar articles.

2.2. Inclusion criteria

This study selected articles based on two distinct types of publication. All the articles presenting a detailed description of the method and the principles of the consensus methods were included (methodological articles). Of the articles whose main objective was to present an application and the results of the consensus method used in a particular field (applied methodology articles), only those that could be accessed for free, i.e., the full articles made available without payment, were studied and illustrated the different domains where the different variants of the methods studied were applied. When necessary, particularly for the description of the variants of consensus methods, the articles were acquired even if they were not freely accessible.

2.3. Analysis of the articles

The relevance of the articles was evaluated by the authors following the study’s objectives based on how well the article described the method compared to the reference documents. Given the objectives outlined above, we did not take into
account the research results themselves but only the description and use of consensus methods.

2.4. Data analysis and presentation of the results

In agreement with the objectives and the type of data, the results are presented textually. The methodological articles allowed us to present the principles and the methodological steps of the consensus methods. The applied methodology articles illustrated the domains in which these methods were used as well as the variants used. Only a few applied methodology articles identified for each of the consensus methods are referenced as examples in this article.

3. Results

3.1. The Delphi method

3.1.1. History

The Delphi method was named after the Delphi oracle in which the Greeks dialogued with the gods through the prophetess Pythia [6]. The Delphi method was developed by the RAND Corporation (a contraction of the expression “research and development”), a nonprofit organization in California whose mission was to assist decision makers in different domains by applying research and analysis tools. The method was initially applied within the Delphi Project, which aimed to describe the long-term development trends in science and technology (scientific breakthroughs, population control, progress in space, prevention of war, and weapons systems) [7].

3.1.2. Objectives and basic principles

The objective of most of the applications of the Delphi method is to shed light on zones of uncertainty based on the knowledge of experts to aid in decision-making [8]. The Delphi method is an exercise in group communication, bringing together and synthesizing the knowledge of a group of geographically scattered participants who never meet [9]. The participants provide individual input through successive questionnaires of close-ended questions, designed by the study’s organizers, to construct an agreement scale [10]. The fundamental characteristics are [11]:

- anonymity: the participants do not know the identity of the others so that dominance, authority or affiliation phenomena can be avoided;
- iteration with controlled feedback: the experts are consulted several times. Before each round, they receive the results obtained at the preceding round so as to consider their opinion in light of what the other participants have said;
- quantified analysis of the group responses: members with opinions at the extremes (totally in agreement or totally in disagreement) are asked to provide additional arguments to the other group members.

3.1.3. Steps in the consensus process

The method follows four basic steps [12].

3.1.3.1. Definition of the objective. This is a fundamental step as much for writing the questionnaires as in choosing the experts [13]. This phase consists in rigorously and precisely defining the object (defined as the problem to examine and the broad questions related to the problem), which will serve as the foundation for the study so that the experts are not pulled into a process whose topic evolves as they go along.

3.1.3.2. Choosing the experts. To prevent the process from being challenged in the future, the experts are chosen for their knowledge of the object and as representatives of a variety of end-users. The minimum number of participants needed for the results to be valid depends on the study’s object: a group of four participants is sometimes sufficient [14], but groups numbering 10 to 15 participants are the general rule.

3.1.3.3. Writing the questionnaire. The questions must be well targeted, precise and formulated to solicit close-ended answers. In this case, each participant responds to each question using a rating scale classically ranging from 1 to 9, from “total disagreement” to “total absence of proof” or a “formal contraindication” (rated 1) to “total agreement” or “formal proof” or “formal indication” (rated 9).

3.1.3.4. Administering the questionnaire and processing the results. The starting questionnaire is administered to the experts by mail (postal or e-mail). The same questionnaire will serve as the main thread throughout the exercise; it will be enriched at each round by the results and comments generated by the preceding round. At the second round, the experts receive the results of the first round and again express an opinion on the questionnaire, knowing the responses of the other members of the group. The participants can maintain their response or modify it; if their new response strongly differs from the group mean, they must explain their response. This round can be reiterated as many times as necessary to reach a convergence of the responses or identify the discordances. The information returned to the participants includes the median consensual opinions and the deviation of the opinions around this median; this deviation can be interpreted using the explanations and comments collected from the experts.

3.1.4. How the method is applied in the literature

We found 2051 articles in the databases with the term “Delphi” (descriptors identified by the MeSH Database, “Delphi Technique”; “Delphi Study”; “Delphi Technic” and by the Thesaurus Santé Publique version 4 “Méthode Delphi”). Of these 2051 articles, 1986 were excluded because they did not provide sufficiently precise information on how consensus was obtained or could not be accessed for free. Of the 65 articles retained, nine were methodological articles [2,6–13] and 56 were applied methodology articles.

The Delphi method was used in knowledge-production and decision-making in a variety of domains (strategy orientation in public health [15], health-related education [16], prevention priorities [17], definition of professional practices and their
improvement [18], quality of care and medical practices evaluation [19]) or in epidemiology and clinical research [20].

Of the 56 articles on applied methodology, 27 used a variant of the Delphi method (“modified Delphi”). The steps that were adapted were, for the most part, limiting the number of questionnaire rounds to two from the start [21], including a meeting of participants (such as the mini-Delphi, which proposes a real-time application [22]), and varying the objectives to meet (arrive at consensus, arrive at the broadest possible range of opinions possible, come to a decision with the participants despite their diverse interests, or succeed in creating common and objectives, as in the Imen-Delphi method [23,24]).

3.2. The nominal group method

3.2.1. History

The nominal group technique was created by two American researchers, André Delbecq and Andrew Van de Ven, at the end of the 1960s [25]. The method was initially recommended for problems requiring the genesis or prioritization of information [26].

3.2.2. Objectives and basic principles

This method was used to analyze a problem, explore a knowledge domain, globally summarize a question, plan and set up activities, assess an intervention, establish priorities, or prioritize actions. Some recommended using the nominal group technique when participants could be gathered quickly and urgent problems or approaching deadlines required immediate responses [2].

The principle is to gather information by classing a series of items or questions during a meeting, which is led by a coordinator [27], either a professional in the domain of the subject at hand [25] or someone with no expertise in the domain [28]. The interactions take place essentially between the coordinator and each of the group’s members to avoid dominance phenomena. Five to nine people participate, but it is possible to include as many as 15 if the subject to discuss is not overly complex. If there are more participants, it is preferable to divide them into several groups.

3.2.3. Steps in the consensus process

3.2.3.1. Silently generating ideas. The coordinator distributes a sheet of paper to the participants on which the question is written. After a few minutes, each person brainstorms silently and notes as many ideas as possible for five minutes.

3.2.3.2. Stating ideas. The coordinator asks each participant in turn to communicate the first response written on his or her sheet of paper. All the responses are written on a board. There are as many rounds as responses. Each participant communicates a single idea per round. If someone considers that one of his ideas has already been expressed by another participant, he passes to the next one. During this second step, no critiques should be brought up and discussion and comments should be limited. The ideas become the property of the group and the coordinator numbers the ideas as they are written on the board.

3.2.3.3. Discussing and individually presenting ideas. Once the ideas have been written on the board, they are clarified in comparison to the other ideas, verifying that all the participants attribute the same meaning to the items and that the logic underlying each opinion is clearly understood by all. The group is responsible for clarifying the ideas and not the person who suggested it. It is important to divide the time as equally as possible among the different items.

3.2.3.4. Ranking the ideas. The group discussion is followed by individual rating of the relative importance of the ideas, with the objective of retaining a number of ideas defined beforehand (for example, five). Each participant attributes five points to the idea judged to be the most important of the five and one point to the least important. The ideas that receive the largest number of points become the group’s priorities.

3.2.3.5. Compiling the results. The coordinator counts the points obtained by each item. Two steps can be added (particularly when the number of participants is high): a discussion of the results of the vote or a final ranking to decide between the items that had tied.

3.2.4. How the method is applied in the literature

Our research identified 1243 articles (term used, “nominal group”). Among these articles, 57 met the inclusion criteria and were identified as applied methodology articles. Six additional articles were identified as methodological articles detailing the method’s principles, the technique, and the objectives [1,2,25–28]. The 57 articles were included as studies of the method applied using the expression “modified nominal group.” The domains of application were the definition of research priorities [29,30], medical education [31] and evaluation of medical practices [32].

Variants were used in 29 of the 57 articles. For example, the method was limited to the last three steps – discussing and individually presenting ideas, voting and ranking the ideas, and compiling the results – with the preceding steps remotely carried out beforehand by the coordinator. In another case, before the meeting, the participants used the literature review to contribute to the discussion on the different points [33,34].

3.3. The consensus development conference

3.3.1. History

The consensus development conference method was used in the healthcare sector in the United States in the 1970s and was developed by the National Institutes of Health (NIH). In 1977, the NIH’s first consensus development conference took place on breast cancer screening. It was part of a consensus research program responding “to the US Congress’s dual concern of setting up a formalized system to evaluate innovations and seeing the NIH undertake the transfer of research results to
medical practitioners, giving it a new responsibility in the validation and dissemination of medical innovations.” The principle was based on bringing together a selected panel that listened to experts present the available data in a public forum. An assessment of these innovations was provided by the panel, which wrote a report on the spot and provided guidelines that were then widely disseminated. Medical expert groups attempt to define by consensus the proper use of new therapeutics drugs. These experts provide guidelines designed to be taken up by the entire profession for the proper use of a medical innovation [35].

3.3.2. Objectives and basic principles

A consensus development conference aims to bring out the points of agreement and divergence within the community concerned relative to a healthcare intervention, whether it be a diagnostic procedure, a therapeutic strategy, or aspects related to the organization of the healthcare system [36]. Like any medical guideline, its objectives are to establish a synthesis of the current knowledge and aid in making decisions.

The consensus development conference consists in a panel writing guidelines after a public presentation of expert reports that synthesize the current knowledge on a precise topic. The public presentation is inspired by the scientific conference, democratic debate and the judicial model.

3.3.3. Steps in the consensus process

As early as 1990, the Agence nationale pour le développement de l’évaluation médicale (ANDEM; French National Agency for Medical Development and Evaluation) described the consensus development conference method in a professional guide [37], then the Haute Autorité de santé (HAS; French Health Authority) proposed a methodological basis for their use in France [38].

The promoter, often a learned society or a public institution, defines the topic and provides the financial means to hold the consensus development conference. The organization committee assumes the responsibility of the entire process and guarantees that it proceeds as it should. It recruits the panel, appoints the president, prepares and organizes the intermediary meetings, recruits the qualified experts, implements the communication plan for media coverage, and finally, organizes the public conference. The panel is multidisciplinary and multiprofessional, made up of eight to 16 people, chosen among physicians, researchers, and nonphysician healthcare personnel, methodologists, representatives of the ethics, economic and legislative domains, and representatives of the general public (patient support groups, the media, etc.). A literature review is provided by a bibliographic group. The panel of experts is composed of 12 to 15 people (scientists, engineers, jurists, economists, sociologists, etc.) chosen by the organization committee. Their role is to respond to the jury’s questions, which they have received beforehand, during the public conference. The public conference is the grand finale of the process. Led by a chairperson, this conference develops in three steps, in public and with the participation of the public: the experts respond to the panel’s questions, the panel provides the responses to the experts, and additional questions are posed by the public. Then the panel meets in camera to write up the final document, which gives special importance to seeking consensus on the questions debated, it establishes the guidelines as independently and objectively as possible, distinguishing what stems from scientific fact, presumption and past practice. After public reading of the final document, the experts can intervene to correct any imprecision or factual error, but do not have the right to influence the opinions expressed.

3.3.4. How the method is applied in the literature

The consensus development conference is the most frequently used method in the healthcare sector: 2238 articles were identified in our study (PubMed/MEDLINE: Limits; Type of Article, Consensus Development Conference; Consensus Development Conference, NIH; descriptors identified by the MeSH Database: Consensus Development Conference; Consensus Development Conference, NIH; NIH Consensus Development Conference; Consensus Development Conferences as Topic; Conferences, Consensus Development as Topic; Consensus Development Conferences, NIH as Topic; NIH Consensus Development Conferences as Topic; descriptors identified by Thesaurus Santé Publique version 4: Consensus development conference). One hundred thirty-five articles responded to the inclusion criteria. Only three articles presented the method and its principles [35,37,38]. Nearly all of the articles presented the results of consensus development conferences, for the most part used by healthcare institutions to define guidelines for professional practices [39,40] or to update the knowledge in a particular domain [41,42].

Changes and variations in the implementation of consensus development conferences were identified [43–46]. The Danish parliament in Denmark sought to differentiate the evaluation of technological choices from the American model. One of the objectives was to reduce the discrepancy between the experts, the public authorities and the general population. This meant designing methods that could take into account the preoccupations, the daily experience and the views of the citizens as well as the skills of the experts, while taking into account constraints encountered by politicians. The topics treated with this variant, called the citizens’ conferences [47], are broader: they must be both a matter of public interest and the subject of controversy in the scientific community as well as in society in general (e.g., air pollution, genetic therapy, etc.). Other methods have come along such as PubliForum, the Swiss version of citizens’ conferences.

3.4. The RAND/UCLA Appropriateness Method and formalized consensus

3.4.1. History

Developed in the United States in the 1980s, mainly by the RAND Corporation and the University of California at Los Angeles [48], the RAND/UCLA Appropriateness Method is highly codified and widely used in public health. The formalized consensus method, published by the French National Authority for Health, was derived from this method...
The French version differs from the original method mainly in that it incorporates a review committee and a group meeting to rank the ideas coming out of the review committee’s consultation.

### 3.4.2. Objectives and basic principles

The RAND/UCLA Appropriateness Method associates aspects of the Delphi and nominal group methods [1]. The objective of the RAND/UCLA Appropriateness Method is to clearly bring out the points of concordance, the foundations of the future guidelines, and the points of discordance or indecision, which may be used to orient later research work.

The role of the experts is predominant since the analysis, literature review and the writing of proposals for guidelines for each field of application are the responsibility of the experts. Repeated quantitative assessment is used by all the experts to encourage relevance, objectivity and homogeneity when ranking the proposals. Individual listing of the proposals for guidelines has the advantage of allowing minority opinions to be expressed, encouraging communication and interactivity with the group, and avoiding the effect of dominance on the part of an expert.

### 3.4.3. Steps in the consensus process

This method follows different steps [50]:

- the promoter chooses the topic and sets the method in motion;
- the organization committee defines the topic and the main fields to cover (fewer than 10) following a literature review on the subject and designates the experts;
- during the first meeting, the group of experts validates the choice of the different fields to cover and divides up the fields between the subgroups of experts;
- each subgroup of experts works independently and carries out the analysis, literature review, writes an expository paper and a list of proposals for guidelines in the field for which he or she is responsible for the overall document that will later be submitted to the entire group;
- during the second meeting, the group of experts puts together the overall document. This means examining the wording of the recommendation proposals, which should be clear, unambiguous and comprehensible, removing any redundancies, and verifying the overall coherence and absence of major oversights. In principle, at this step, there is no discussion on the content nor on the relevance of the proposals for guidelines.

The overall document is sent to all the experts and is rated. The rating rules are predefined and determine both the agreement (or disagreement) and the degree of convergence of the expert opinions. The responses to each question or proposal are analyzed, taking into account first the median, then the deviation of the rates on a scale from 1 to 9. Three zones are defined in function of the place of the median calculated out of all the rates: 1 to 3 correspond to the disagreement zone, 4 to 6 to the indecision zone and 7 to 9 to the agreement zone. For each question, the degree of convergence of the group’s opinions is assessed by its position on the scale surrounded by the minimal and maximal values. Agreement (or disagreement) is said to be strong if the interval is within the boundaries. If the interval encroaches on a boundary, the agreement (or disagreement) is said to be weak (e.g., interval from 1 to 4 or from 6 to 8) [51]. In the RAND/UCLA Appropriateness Method, an extreme high value and an extreme low value are eliminated.

For the second round of rating, the experts know the group’s responses to each proposal (median, distribution and range, along with their own rating) and the arguments of each subgroup. At each round, the expert is free to change or retain his or her opinion (according to the Delphi method’s principle).

The group of experts is brought together for the third time. The objective is to identify the proposals for guidelines for which there is agreement, disagreement or indecision so as to reach a final text of guidelines. The results of the second round are presented. The proposals for which there is weak agreement, disagreement or indecision within the group are discussed based on the arguments of the subgroup concerned. At the end of this discussion, the items are rated individually by each expert a third time, with each free to change his or her mind or remain with the same position. This last round of rating is the basis of the formalization of the guidelines on which the final written document is based. The points of disagreement or uncertainty are identified and used notably to define future research perspectives.

The guidelines are written by the organization committee based on the results of the last round of rating and the arguments cited.

The national healthcare organizations (NIH in the United States, HAS in France) have developed other consensus methods founded on the Delphi, nominal group and RAND/UCLA Appropriateness Method. The formalized consensus method [49] brings in different actors: the promoter, the pilot group, the rating group and the review group, which gives an opinion on the content and structure of the guidelines retained, in particular on their applicability, acceptability and readability.

The formalized consensus method is carried out in four phases:

- a preparation phase, which aims to define the topic, carry out the literature review and produce a preliminary series of items (pilot group);
- a rating phase (rating group): first individual rating of the proposals followed by a group meeting to discuss the results and then a second rating using a rating scale;
- an evaluation phase (a review group);
- a text finalization phase (pilot group).

### 3.4.4. How the method is applied in the literature

Sixty-eight articles were identified (the expression “RAND/ UCLA”), 19 of which responded to the inclusion criteria. Two articles were identified as methodological articles [48,50]. The RAND/UCLA Appropriateness Method is used in public health, notably to define professional practices [52,53] and evaluate new healthcare technologies [54]. Thirty-four articles covered formalized consensus (using the expression “formalized
consensus”), four of which responded to the inclusion criteria. A single article was a methodological article [49]. The formalized consensus method was mainly used to evaluate healthcare technologies [55] and for professional practices guidelines [56].

4. Validity of the consensus methods

The number of studies aiming to validate the consensus methods, compared to the other decision-making assistance approaches, was extremely low relative to the frequency with which they are used. At the beginning of the 1970s, studies demonstrated greater validity of group consensus compared to an individual elaboration of knowledge [57]. Some studies compared the consensus groups to unstructured groups operating on direct interactions. A psychometric study conducted by Delbecq et al. on 420 participants [2] demonstrated that the Delphi technique and the nominal group method were clearly superior to the unstructured group and that the nominal group method was slightly better than the Delphi method in terms of the number of ideas expressed and the satisfaction of the participants in the group decision-making process. The 420 participants were assigned into 20 groups of seven members for each of the three techniques. They came from the same milieu, a midwestern university, and they had equal knowledge on the topic. The heterogeneous groups were composed of students, professors and administrators. The two effectiveness criteria were the number of ideas expressed and participant satisfaction.

Other studies have also shown the effectiveness of the Delphi and nominal group methods [26,58,59]. No studies were found that investigated the validity of the Delphi and nominal group methods in the healthcare domain, contrary to the studies on the RAND/UCLA Appropriateness Method [54,60]. The last method was the subject of one article that reported the reliability of the definition of the appropriateness of medical treatment by groups of doctors applying this method independently [61]; another put forward the uncertainties concerning the validity of the method when decisions concerned patient management [62]. A last study concluded that, despite its shortcomings, the method made it possible to define clinical guidelines [63]. The example of digestive endoscopy was a particularly well-conducted study [64,65].

5. Discussion

All the methods were precisely described in the literature and included the common principles such as the definition of the subject, the selection of experts and a process of direct or remote interpersonal interactions. Rating was sometimes used to prioritize the proposals. A number of variants to the original definitions were found but there were few validation studies.

To meet the objectives set out for the present study, an exhaustive review of the literature did not seem necessary and for practical reasons, only the articles that could be accessed for free were reviewed. These were sufficient to cover the known variants and probably the different types of application of the consensus methods. Only one order had to be made for articles describing the Imen-Delphi method. An independent double reading was not useful given the objectives of the study.

The modalities for implementing the consensus methods were precisely described in the original articles or the reference documents. The different steps involved in the methods are presented in Table 1. Even if their use should be adapted to the context, objectives and local conditions, their main principles should be respected. Yet, certain examples in the literature show that the methods can be watered down to such an extent that the procedures have no more in common with the methods than their name. Could it be that the current literature dilutes these consensus methods even more, given that most of the consensus results are not published? Even though we did not conduct an exhaustive review of the literature, the few articles cited illustrate that the consensus methods are used at a cost of sometimes substantial distortions of the original method, even though they directly take advantage of them. This observation poses the question of the validity of the results.

| Table 1 |
|------------------------|------------------------|------------------------|------------------------|
| **Main steps in consensus methods.** | **Delphi method** | **Nominal group method** | **Consensus development method** | **RAND/UCLA method** |
| 1 | Define topic | Define topic | Define topic | Define topic |
| 2 | Select experts | Select experts | Logistical organization of method | Select experts |
| 3 | Write questionnaire | Organize meeting | Select experts and panel | Assign topics to cover between expert groups |
| 4 | Validate questionnaire | Silent brainstorming | Review literature | First group meeting, review literature and define items |
| 5 | First round | List ideas | Public conference: expert, panel, public question/response | Second group meeting, write overall questionnaire |
| 6 | Analyze group responses | Expressing individual ideas and discussing | Panel deliberates in camera | Individuals ranking items |
| 7 | Second tour | Prioritization of ideas | Write guidelines | Third meeting discussing ratings |
| 8 | Analyze group responses | Compile ideas and validate | Communicate guidelines | Establish consensus report on items |
| 9 | Third round | Communicate data obtained | | Communicate results |
| 10 | Analyze group responses | | | |
| 11 | Repeat as many times as necessary to stabilize group responses (steps 9–10) | | | |
| 12 | Conclude analysis, obtain consensus | | | |
| 13 | Communicate information obtained | | | |
These heterogeneous practices, which contrast with those that prevail in other methods designed to elaborate knowledge, undoubtedly maintain the suspicions the scientific community has concerning these methods. Several reasons can be given: the lack of rigor in the use of these methods, the absence of a precise description of the process put in action to obtain consensus in these publications and the lack of validation studies. The validity of these methods is indeed rarely studied, although their value seems to be growing. Proposed as legitimate when the other approaches are inappropriate, they could be preferred to other approaches even if these are feasible, because they are less costly and more rapid [66]. To encourage their use, validation studies should be undertaken in the domains where these methods are applied, particularly in healthcare.

6. French version

A French version of this article is available at doi:10.1016/j.respe.2008.09.006

References
