

Good practice for guidance development – review of consensus methods

Jacek Siwiec  <https://orcid.org/0000-0003-4058-7005>

Magdalena Konieczna

Magdalena Koperny  <https://orcid.org/0000-0002-3352-0972>

Department of Health Care Benefits, Agency for Health Technology Assessment and Tariff System

Address for correspondence: Jacek Siwiec, Wydział Świadczeń Opieki Zdrowotnej, Agencja Oceny Technologii Medycznych i Taryfikacji, ul. Przeskok 2, 00-032 Warszawa, sekretariat@aotm.gov.pl

Abstract

Introduction. The process of synthesizing scientific evidence and translating elaborate scientific research findings into recommendations for clinical practice is one of the most complex and significant initiatives in health care. In connection with the increasing number of recommendations, previously unformalised elements such as the process of consensus decision-making require developing a methodological approach. The use of consensus methods serves to reduce the impact of psychosocial factors in the formulation of recommendations and to increase the degree of process transparency. The aim of the publication is to present formal methods of achieving consensus used in the process of developing clinical practice guidelines.

Materials and methods. The review focuses on results of reports analyzing the procedure and methods of developing clinical recommendations, with particular emphasis on the identification of consensus methods.

Results. Many international and national organizations and institutions develop recommendations or adapt guidelines developed by others. They use both informal and formal instruments to deal with divergent expert opinions. The most popular formal methods of achieving consensus identified in the review are the Delphi method, the nominal group technique, the RAND/UCLA method, the consensus conference, and the combinations of individual approaches. Formal methods have been shown to lead to less biased and more reliable recommendations. Regardless of the method used, the guidelines should clearly define the quorum and document the process of agreeing a common standpoint on recommendations, guidelines and recommendations.

Conclusions. Clinical guidelines have become an important tool influencing clinical practice. The participation of many experts representing the opinions and interests of different groups makes it necessary to apply a methodological and structured approach so that all participants have the opportunity to voice their opinion and to ensure process transparency, deal with misunderstandings and achieve a consensual position. The consensus methods allow to provide a wide range of stakeholders clinicians, policy makers, patients – with agreed rules of conduct in a given topic. The formal consensus methods and recommendations based on these methods combine scientific evidence with the practice and experience of experts.

Key words: clinical practice guidelines, standards, recommendations, formal consensus methods

Słowa kluczowe: wytyczne praktyki klinicznej, standardy, rekomendacje, metody formalnego konsensusu



AGENCJA OCENY TECHNOLOGII MEDYCZNYCH I TARYFIKACJI



Fundusze Europejskie
Pomoc Techniczna



Rzeczpospolita
Polska

Unia Europejska
Fundusz Spójności



Introduction

The development of clinical practice guidelines is one of the most complex and significant health care initiatives

affecting clinical practice and the functioning of the system as a whole. Over the past decades, the number of guidelines developed by various governmental and private institutions and organisations around the world has

grown rapidly. Many local, national and international communities are now in the process of identifying relevant clinical areas that require reviewing scientific evidence and making recommendations that they believe should help clinicians to care for their patients, as well as provide information on possible diagnostic and treatment options for patients [1].

The U.S. Institute of Medicine (IOM) defines clinical practice guidelines as 'positions that contain recommendations to optimise patient care, based on a systematic review of evidence and an assessment of the benefits and harm of alternative care options' [2]. Based on this definition, two essential elements can be distinguished in the structure of the guidelines:

- a systematic review of scientific evidence from studies relating to a clearly formulated clinical question, focusing on the strength of the evidence on which clinical decision-making is based; and
- a set of recommendations, including both evidence-based judgements and valuing the benefits and disadvantages of alternative patient care options relating to what was found in the review.

The primary objective of developing clinical recommendations, especially clinical guidelines, is to base clinical practice on the most up-to-date and highest quality scientific reports contained in clinical trials treated as scientific facts. In the paradigm of evidence-based medicine, guidelines should be the most important source of regulatory recommendations, and to some extent standardise current clinical practice, positively influencing the improvement and provision of high quality medical services [3]. Clinical practice guidelines also play an important role as a health policy tool. In a situation of unrestricted needs and limited resources of the health system, they are a guideline to facilitate the allocation of resources by financing only clinically efficient and cost-effective technologies. At the same time, they indicate to those who have clinical practice (doctors, nurses and other medical personnel) the paths of patient management. More and more often, the subject of the guidelines are no longer only recommendations concerning clinical practice, but also recommendations formulated with regard to the process of multidisciplinary patient care and organisation of health care as a whole and the quality of this care.¹

Group decision-making is a cognitive process based on cooperation. With regard to clinical practice, a team of experts develops both clinical practice guideline and consensus statement. Although both types of documents provide guidance to physicians, there are significant differences between them. The development of clinical practice guidelines results in recommendations indicating the validity (or otherwise) of a specific intervention and determining the strength of the recommendation, based on available scientific evidence and other factors. Clinical practice guidelines emerging in the HTA paradigm are always based, in the first place, on clinical reports, i.e. conclusions from clinical trials, systematic reviews, meta-analyses, which are quantified in terms of quality and timeliness for clinical practice. It is worth

noting that consensus methods are usually used when scientific evidence is limited or their quality is questionable in relation to methodological determinations.

The method of formal consensus can be used both in the process of developing clinical practice guidelines and expert positions. Narrow consensus means a wide agreement among the members of a group. In a broader sense, it is also the theory and practice of reaching an agreement, i.e. the process of reaching consensus. The essence of consensus is that the group seeks the best solutions, acceptable to all persons in the group through discussion, exchange of views and opinions. Consensus is more than a vote for or against, it is a form of reaching agreement through negotiation. It is also not synonymous with compromise, since the aim of consensus is to find solutions that are acceptable to all those involved in the process, rather than being achieved through concessions [4].

The purpose of this article is to present formal consensus methods and their role in the development of clinical practice guidelines.

■ Key elements of the guidance process – GRADE development methodology

The subject of clinical practice guidelines is inextricably linked to the issue of their quality. Clinicians, patients and other stakeholders are confronted with a growing number of different and often conflicting guidelines. In 2000, a working group called Grading of Recommendations Assessment, Development and Evaluation (GRADE) was set up to develop a uniform scheme for the development of reliable guidelines. The GRADE methodology has already been adopted by organizations such as the World Health Organization (WHO), American College of Chest Physicians, American Thoracic Society, European Respiratory Society, Endocrine Society, Society of Critical Care Medicine, American College of Physicians, Canadian Cardiovascular Society, National Institute for Health and Clinical Excellence and Scottish Intercollegiate Guidelines Network. The GRADE tool offers a transparent and structured process for developing and presenting systematic reviews and guidelines. On the basis of the GRADE methodology, general stages will be presented to help identify areas for consensus.

The most important steps in the development of guidelines defined by the GRADE working group include [5]:

1. Defining the clinical question and collecting available scientific evidence (data).
2. Assessing the quality of the scientific evidence gathered.
3. Formulating and classifying recommendations.

The process of developing clinical guidelines begins with the definition of the clinical question in terms of: population, alternative management strategies (interventions and comparators) and all relevant patient outcomes (endpoints as indicators of change or lack of change in the patient's condition). It is recommended to use

the PICO/PICOC principle for formulation of the clinical question. For the purpose of the guidelines, the endpoints are classified as ‘critical’ and ‘relevant’ [5]. A systematic literature search according to the clinical question leads to the inclusion of relevant studies. The authors of the guidelines then use data from individual eligible studies to obtain the best estimate of the impact of the assessed conduct on each patient-relevant outcome, together with uncertainties associated with this estimate.

In the GRADE approach, randomised controlled trials (RCTs) are considered to be the highest quality scientific evidence, while observational studies are considered to be lower quality evidence supporting the estimation of the effects of interventions. Five factors may lead to a decrease in the quality of evaluated evidence [6]. These are:

- risk of systematic error;
- inconsistency of test results;
- uncertainty as to whether the results can be related to the clinical situation;
- imprecise estimation of effects;
- publishing error.

In turn, three factors may lead to a higher final rating [6]. These include:

- a strong link between the intervention and the endpoint;
- demonstration of dose-dependent effect;
- disturbing factors increase certainty as to the estimated effect.

Ultimately, the quality of the evidence for each result falls into one of four categories on a scale from high to very low. It is important to emphasise that this approach is used to assess the quality of evidence for each outcome (endpoint) in different studies and not for each individual study. If studies report the same results for the same populations, their estimation is combined and the cumulative effect is assessed (meta-analysis). The GRADE system is oriented towards the evaluation of the ‘outcome’ and the quality of these results may vary within a single study [5].

The third key stage in the development of the guidelines involves formulating practical recommendations based on all available information. The authors of the guidelines make the final decision as to which test results are critical and set a target quality qualification for the evidence. The direction and strength of the recommendations from the collected material is then considered. The balance between desired and undesired results and patient preferences determines the direction of the recommendation, while the quality of the evidence additionally influences the strength of the recommendation. The guidance panel must decide whether the desired effects of following the recommendation will outweigh the undesirable effects and the strength of the recommendation reflects the degree of confidence of the group in this assessment. A strong recommendation for an intervention reflects the collective view that the desired effects of an intervention (beneficial health outcomes, reduced burden on staff and patients and cost

savings) clearly outweigh those that are undesirable (damage, greater burden and higher costs). The weak recommendation reflects the collegial view that the desired effects outweigh the undesirable ones, but the panel is not convinced of the compromise reached – either because the key evidence is of poor quality (and therefore the benefits and risks remain uncertain) or because the benefits and negative effects are balanced. Both direction and strength can be modified after considering alternative approaches. The process of making recommendations is a stage where there are strong differences of opinion in interpreting the evidence and deciding on the best alternative course of action. The approach to developing guidelines proposed by GRADE, although structured, in many aspects, such as deciding on the quality of evidence, requires the subjective judgement of the developers, which can inevitably lead to different opinions. The consequence of this is a consensus and a final decision.

Over the years, as the idea of the most comprehensive and multidisciplinary approach to the development of recommendations has been implemented, guidance panels have become more and more numerous and consensus-based decision making issues have become increasingly important in the process. Clinical experts, methodologists, general practitioners, patient representatives and other stakeholder groups present in the recommendation panels are characterised by different skills, knowledge, experience, perspectives, roles and responsibilities in the healthcare system. Developed and proposed by the Guidelines International Network (G-I-N),² an international network of guidance writers, the guidance development guidelines are largely based on recommendations on the role of experts and stakeholders. The following are a selection of key elements proposed by the GIN for the development of the guidelines related to the subject matter of this study, which is the role and methods for reaching consensus in the process of developing clinical management guidelines [7]. These elements relate to the composition of the panel, decision making, review and consultation of stakeholders:

1. Composition of the guidance development team:

The guidance development team should include representatives of various stakeholders such as health professionals, methodologists with the ability to assess and synthesise scientific evidence, economists, experts in the field, patients or other beneficiaries. The composition of the group may result in biased recommendations, and therefore a balanced contribution of all members to the discussion and consensus building should be ensured. The chairperson of the group should play a key role. The size of the guidance-group is also important: large groups may be difficult to manage, while small groups may lack relevant stakeholders.

2. Decision-making process: The guidelines should describe the process of reaching consensus among panel members. This process should be established before the guidelines are developed. It is required in order to select and interpret scientific evidence,

translate the evidence gathered into recommendations, and determine how to deal with missing or insufficient evidence. Although some countries and organisations use formal consensus processes (e.g. nominal group technique, Delphi or formal voting [described further], many organisations use informal processes. Formal methods have been shown to lead to a less biased and more evidence-based consensus than informal methods [8–10]. Regardless of the choice of method, guidelines should clearly state the quorum and document the consensus process.

3. **Peer review³ and consultation with stakeholders:** The guidelines should be reviewed by independent experts before publication. Reviewers may include external experts, representatives of the sponsoring organisation, and members of the public. Reviewers should include experts in the field who are not directly involved in the development of the guidelines, but who have the knowledge and experience of the subject matter of the guidelines. When selecting reviewers, it is important to consider those who are more likely to provide comments based on scientific and clinical knowledge, rather than simply opinions not supported by evidence. Guidelines may also be reviewed by representatives of patient organisations and even by patients themselves or by representatives of the target group covered by the guidelines. The method of reviewing should be described, and the description may include a list of reviewers and the names of the institutions or organisations with which they are associated. A summary of the review should be attached to the content of the recommendation.

In summary, large and diverse panels of experts and reviewers present challenges in the decision-making process, such as ensuring that all participants have a voice and can actively participate in the debate, ensuring transparency, dealing with disagreements, reaching consensus and resolving situations where consensus is not possible. Expert panels often use informal processes to deal with these challenges, but these techniques are vulnerable to the impact of group interactions. Factors such as time pressure, fatigue, lack of knowledge or specialist methods in the field, the dominance of persons with a strong personality and/or reputation threaten the integrity of such consensus. Therefore, the need for a more formal process of dispute resolution, interpretation of evidence and formulation of recommendations was recognised, especially in times of rapid change, with new information coming from ongoing clinical trials. This need was underlined by the criticism of informal methods in terms of conflict of interest and alleged manipulation of academic authors by the pharmaceutical industry [11].

Role of stakeholders

The idea of taking into account the expectations of the stakeholders is in line with the concept of social responsibility, which is widespread also in the area of health care [12–14]. ISO 26000, published in 2010, states that social responsibility is “the responsibility of

an organisation for the impact of its decisions and actions on society and the environment, provided by transparent and ethical behaviour that contributes to sustainable development, including the well-being and health of society, takes into account the expectations of stakeholders, complies with applicable law and is consistent with international standards of conduct, and is integrated into the organisation’s activities and practised in its relationships” [15]. Social responsibility concerns all organisations, whether public, private or non-profit [16]. One of the key manifestations for implementing the concept of social responsibility is the involvement of stakeholders [17]. According to ISO 26000, stakeholders are individuals or groups interested in decisions or actions of an organisation [15]. In the process of creating clinical guidelines, most stakeholder groups include medical professionals (including doctors, nurses, paramedics, physiotherapists), patients, representatives of public and private research organisations, healthcare managers, entrepreneurs and manufacturers [18]. The active participation of representatives from all groups is important to ensure that evidence (or lack of evidence) is adequately discussed when developing recommendations [19]. Where possible, experts and stakeholders involved in the guidance development process should represent separate organisations and interest groups. This prevents bias and is helpful in situations of interpretation of ambiguous data. The stakeholder identification includes all groups whose activities would be covered by the guidelines or who have other legitimate reasons to contribute to the process. Stakeholder identification and sourcing involve considering the clinical subject and the organisation’s methodology [20, 21].

Different stakeholder groups have different expectations and opportunities for impact in a given clinical area and their involvement in the development, evaluation and implementation of guidelines can have a significant influence on health outcomes, such as in infection control [22]. It is possible that the same scientific evidence may lead to different conclusions from a multidisciplinary group. A monodisciplinary group may be systematically biased in favour of technologies and procedures in the implementation of which it has a vital interest [23, 24]. For example, in the Leape study (1992), vascular surgeons preferred the use of carotid endarterectomy more than a mixed group of surgeons and medical specialists [25]. In the following section we will look at the problems related to the functioning of multidisciplinary groups.

Psychosocial factors influencing the guidance process and consensus building

A number of psychosocial factors can influence the process of developing guidelines. To apply consensus methods skilfully, it is worth realising how group processes can result in incorrect or unreliable recommendations, especially in highly diverse multidisciplinary groups. Although at the beginning of the process (usually before or at the first meeting) the roles of individual panel members (e.g. chairperson, patient representative) are

defined, the social development of such a group will follow certain characteristic patterns [26]. The ideal conditions for group decision-making are those that allow the views of all parties to be expressed and considered before a recommendation acceptable to the majority of the panel members is established. Pagliari (2001) points to five main factors related to social impact that can affect the interaction of individuals in small groups and that can cause dysfunction in the decision-making process [26]:

- Conformity – the tendency of individuals to adjust their behaviour to what they perceive as a group standard in a given situation. In the process of developing guidelines, conformity creates problems when identifying areas of disagreement (individual members of the panel do not report a disagreement, or report it after the meeting).
- Compliance is a situation where members are guided in their decisions by direct request (e.g. due to time pressure, panel members are asked to vote according to the voting of the majority). In this situation, there is a risk that official ‘compliance’ leads to a falsified consensus.
- Obedience – circumstances in which an individual’s decision is based on an unequivocal order from a person who is an authority. In the process of developing guidelines, it may have an indirect significance when, for example, sanctions are suggested for those who do not comply with the views expressed by the strong individual.
- Status – the status of panel members (both perceived and actual), where some panel members are normally in positions where they are expected to follow the instructions of others (e.g. nurses and doctors). In such cases, the mere fact that a member of a group belongs to one or other professional group may be sufficient to trigger a trend or pressure even if members have been clearly instructed to express their individual views. The status may also influence the degree of participation in a group discussion (e.g. the proportion of time spent on particular topics).
- Persuasion – can be considered as a process in which the attitudes of individuals or groups are changed as a result of a convincing message.

Although the pressure (influence) of the majority is seen as the main source of social impact in groups, it is possible that an individual member of the group or a minority of the group will significantly influence the viewpoint of the rest of the panel. According to Moscovici (1976), minorities are most effective when they are consistent and committed in their opposition to the majority and are flexible enough to recognise the perspective of other group members [27]. Minority influence can manifest itself through the use of covert divisions within the group, inspiring individuals with ambivalent feelings to change their views.

A commonly observed trend is also that group decisions are more extreme than individuals’ initial inclinations [28]. This phenomenon is defined by group polarisation and can be observed, for example, in the case

of making decisions about judgments in the American judicial system, where jurors after a group discussion are more likely to pass a judgement that is significantly higher (or lower) than the one proposed individually before the debate by its members. One of the attempts to explain this phenomenon is based on the assumption that people are strongly motivated by the desire to be ‘good’ and the desire to be perceived positively by others, so during group discussions they will try to take the most extreme form of this socially defined ‘correct’ position to distinguish themselves positively from other members of the meeting. According to an alternative theory of persuasion, during the exchange of information between the group members, individuals gain additional arguments to support their own original beliefs and mutually reinforce each other’s rightfulness, which in turn leads to the extremes of attitudes, in addition to changing attitudes in ambivalent (undecided) people. To sum up, it should also be emphasized that what is important in this phenomenon is that it is the views of individuals in the group that become stronger (extreme) after the discussion (group interaction) and not the views of all group members, which consequently leads to a situation in which at the end of the meeting these individuals are more opposing to each other than at the beginning. The transition to a more extreme position (the effect of group polarisation) may have negative effects on the decision-making process concerning clinical recommendations [26]. Group polarization will not necessarily be a sign of making more risky decisions; in fact, this change may be related to e.g. support for a more conservative intervention, assuming that the majority of group members initially supported more cautious therapeutic options. According to Pagliari (2001), this may explain the observation that multidisciplinary groups are more conservative than monodisciplinary groups [26].

A more complex problem analysed in relation to the decision-making process is the so-called group-think syndrome. In psychology, this term means a situation in which the group members’ desire for unanimity takes precedence over their motivation to realistically assess alternative courses of action [29]. In groups that are very consistent, value consensus highly and have a very dominant leader, a realistic assessment of the evidence may take a secondary place compared to the desire to maintain good social relations and/or strengthen the decisions that the group has previously made. The dominance of consensus as a condition for accepting recommendations in the guidance process may discourage members from expressing different opinions. As guidelines are usually developed in several meetings, it is possible that tentative decisions taken at an early stage of recommendation development will be defended and strengthened since their rejection may jeopardise the coherence of the group [29].

Another issue is the role of the group leader in the decision-making process and in seeking to define consensus. In the process of socialization and defining group roles, there are two types of leaders, social-emotional leaders and task-oriented leaders. Social-emotional

leaders tend to act in a way that results in maintaining good social relations while task-oriented leaders are more important in terms of effectiveness and ensuring that the group reaches its goals [30]. The managerial style of a group leader may influence the process of deliberations and guidance development, e.g. when an over-controlling and over-directing chairperson prevents all parties from speaking and expressing their views, which in turn inhibits discussion, potentially leading, to unworkable recommendations, for example. An overly charismatic leader, on the other hand, may find it difficult to clearly define objectives and competences resulting in more labour-intensive consensus or ambivalent recommendations. In a model decision making process, the group should include leaders with both task-oriented and socio-emotional skills, but large multidisciplinary panels suggest that guidance developers consider separating these roles. For example, in a UK guidance development project, where guidance groups were led by an experienced leader (the so-called small group leader) who was responsible for protecting the guidance processes (making sure that the group functions satisfactorily) and a facilitator (mediator) who was responsible for making sure that guidance development was completed [31].

The impact of most of these factors can be modified to some extent by formalising the decision-making procedures at each stage of the guidance process. Decision-making schemes are open or implicit rules according to which panel members reach consensus.

■ Formal methods of consensus building

The aim of the consensus is to formalise the degree of agreement between the experts by identifying and selecting elements/points with which they agree, disagree or are undecided. This is a rigorous and unambiguous method based on the involvement of many people: clinical practice professionals, patient representatives or other stakeholder groups. This process involves the knowledge, experience and perspective of individual experts, but the decisions and individual recommendations do not reflect the opinion of one person, but of the whole group of experts drafting the document [32]. The consensus combines knowledge from current scientific evidence with the experience of experts (clinicians, other health care professionals) and values. Consensus mainly results in suggestions, recommendations and guidelines for action.

The most common formal methods of achieving consensus are the Delphi method, the nominal group technique, the RAND/UCLA method, the consensus conference, and combinations of different approaches.

The Delphi method

The Delphi method is a structured process of collecting opinions on a given topic from a group of experts by means of a questionnaire developed by the researcher. The experts exchange their opinions anonymously by answering the questionnaire and on

their basis the moderator (researcher) develops the answers received. A round of questions is then repeated and the opinions (answers) of the group are again collated in quantitative form. The aim is to reduce the range of different responses and to reach an expert consensus to contribute to the formulation of suggestions/recommendations [33, 34]. This method not only reflects the individual knowledge and experience of each expert but also maximises their knowledge [35].

The classic Delphi method was developed by the American RAND Corporation, in particular by researchers Norman Dalkey, Olaf Helmer and Nicholas Rescher. The non-profit RAND Corporation was founded after World War II, in 1948, to conduct research and analysis in the context of defence planning and the military sector. To this end, the Delphi method was developed in the 1950s with the original aim of forecasting the impact of technology on warfare. All the same, it has been very popular for nearly 70 years and is used in many fields of science such as business, management, education, including health care [36]. There have also been many modifications of this method [37], which at the beginning was used only for forecasting purposes [35]. Computerisation and wider access to the Internet have made the Delphi method widely used by scientists and researchers as it allows the consensus of a panel of experts to be reached 'long distance' without the need for meetings. Therefore, the method is time and cost effective [34].

The essence of the Delphi method is to form a panel of experts, often forming an interdisciplinary team, which allows to gather information and opinions from many perspectives. Ideally, the panel should consist of six to eleven experts [38]. The whole process is managed by a moderator (manager, principal investigator), who is responsible for contacting the experts, moderating the discussion, elaborating the questions in the questionnaire, collecting the answers/opinions and preparing their analysis. A key aspect of consensus building using the classic Delphi method is to ensure anonymity of the experts. Anonymity gives more freedom of expression, thus avoiding the influence of other people, e.g. authorities, which is an additional value of this method [36].

The Delphi consensus process consists of several stages:

1. The process begins with the definition of the research problem. Most often the research problem is defined by a group of experts who see a need or are interested in creating solutions or recommendations in a given area. On this basis the team of researchers develops a set of questions – a questionnaire. The questions are based on a literature review and identification of areas that should be evaluated by experts in a given field. A moderator should be appointed within the research team to manage the whole process. The questions to the experts should be clear, precise and unambiguous.
2. Round One starts by sending a questionnaire to the panel of experts by e-mail. Each of the experts

shall reply to the questions together with a justification and a statement of their position. The questions may be either closed (answers may take the form of 'yes' or 'no', the form of a scale, e.g. from 1 to 5, together with an explanation of the scale) or open. The expert may refuse to answer a question if they do not feel competent in the area. In modified forms of the method, the first round can also be used to collect research questions and ideas from the panel of experts, which will be considered in subsequent stages. Such a solution may be useful in situations where there is not enough scientific information on a given topic.

3. The moderator and the research team collect the answers from the experts in Round 1 and prepare a summary and conclusions (without disclosing the authors of the answers). Quantitative data may include the mean, median and/or frequency distribution for each question/problem. Similar responses shall be grouped into categories. Thanks to that participants can usually see the ranking and structure of their responses compared to other participants and from individual rounds.
4. In Round Two, the moderator again sends questions to the panel of experts together with an analysis of other participants' answers, asking them to answer the same questions after hearing the opinions of others. The procedure is analogous to the first round, i.e. the researcher prepares the conclusions of the second round and sends them back to the experts.
5. Through successive rounds and presenting the answers of all participants, the aim is to stimulate experts to exchange views and information, but also to modify their assessments and accept the opinions of others. With each subsequent round the degree of agreement between the experts increases. The process is repeated until the best possible level of consensus is reached or until the number of rounds is completed a priori. The process must involve at least two rounds.
6. The conclusions and analysis of the collected responses and expert opinions shall include the determination of the level of consensus, a summary of the assessments, and the categorisation of responses. The level of consensus can be measured using categorical variables, i.e. the percentage of identical responses, which can take values from 51% to 100%, or quantitative variables, i.e. a specific value of the interquartile range (IQR).

There are two main types of Delphi method. In the first one, a common opinion is sought (a consensus-building Delphi), and in the second one, an overview of possible policy Delphi positions is carried out (the policy Delphi) [35, 39].

In clinical situations and formulation of clinical practice guidelines, recommendations are based on the best available scientific evidence, i.e. randomized control group trials (RCTs), meta-analyses, high quality systematic reviews with low risk of systematic error and clinical records. However, where there is no or

insufficient scientific data, an additional option is to base recommendations on expert advice. Differences in opinions are resolved using scientific evidence. It should be stressed that the Delphi method does not serve to evaluate whether or not an expert consensus decision is right or wrong [40]. The result of the survey constitutes an opinion.

The advantages of the Delphi method include, above all, the possibility of involving experts from various geographically distant places, since there is no need for direct meetings, each participant can complete the task at a time convenient to them [41]. Another important feature of the Delphi method is its anonymity and lack of bias. The repetitiveness of the process and at least two rounds that it involves allow more thoughtful or credible conclusions to be reached than during a single 'face-to-face' meeting organised during conferences, congresses or specially dedicated meetings. However, the use of multiple rounds may burn out participants and increase reluctance to respond in subsequent rounds. In addition, the participants' opinions are influenced by the feedback from subsequent rounds [32].

The method also has its limitations. A large number of survey participants and too many rounds may result in reluctance to re-evaluate and reassess. The coordination of a large group may also have limitations and make consensus difficult to reach. Regarding problems where direct contact is desirable, this method is not applicable [42]. Furthermore, there are opinions in the literature which indicate that the results obtained by the Delphi method are not objective, but are a collection of subjective judgments of a group of participants.

To sum up, the Delphi method, despite its weaknesses, is a commonly used technique for reaching consensus. It should be properly conducted and according to the following rules:

- a selection of experts representing the various communities concerned by the research problem;
- an adequate number of experts and close supervision by the trainer;
- providing conditions for free speech and expression of own opinions;
- not omitting any opinion in the process, with particular reference to extreme positions, and paying attention to their arguments;
- usefulness in solving problems and establishing a recommended course of action on a given issue [43].

Given that it is used in many fields and areas, the Delphi method has undergone many modifications over the years. In medicine, a modified Delphi method may include, in the first round, reviewing the scientific evidence in a given clinical problem, then ranking it in importance/significance/urgency, then expressing expert opinions on selected priority clinical problems. While the first round can take place anonymously, in the modified method, in the next round, experts can meet directly to discuss the results after the first round and establish a final consensus [44].

It also seems important to define the concept of consensus. Some researchers specify cut-off points for

consensus before the process, others use an arbitrary cut-off point to reduce the number of proposals/opinions under consideration (e.g. cut-off threshold of 70%, or much lower, of 51%) [45].

Nominal group technique

The purpose of the nominal group technique is to encourage the group to generate ideas and then work out the best possible solution. This process is similar to the Delphi method except that group discussions (face-to-face) take place between rounds where individual opinions are recorded [45]. Discussions should be moderated and conducted by an experienced facilitator. The technique was developed in the 1960s by Delbecq and Van de Ven.

There are many modifications to this method, but the basic principles are similar:

1. Identification of the problem to be analysed by the research team.
2. The members of the group shall individually and independently record their ideas and considerations on the problem before starting the discussion.
3. Ideas are presented to the entire group using a round-robin format. Each participant presents one idea to the group. Each of the proposals is recorded, and after noting all the ideas, a discussion begins.
4. Similar proposals/solutions are grouped and each proposal is debated, discussed and evaluated by the group.
5. Each member of the group establishes a hierarchy of ideas on his or her own and writes down his or her assessment anonymously on a piece of paper. After summing up the votes, the final position that received the most support is determined.

In the nominal group technique, it is not required that the literature review be completed before the expert meeting in the first phase, but the analysis of the decision problem and basic information on the issue to be discussed should be provided to participants. According to Humphrey-Murto (2017), the number of members in a group usually ranges from 5 to 12 people [32]; according to Waggoner (2016), the panel should include between five and nine participants [38]; and according to Horton (1980) between seven and ten participants [46]. No fewer than five members aim to ensure the diversity and quality of the decisions taken (ensuring the generation of different opinions), while too many members can result in a very wide variety of opinions and solutions, and a lengthening of the consensus process [47]. The whole process can take an estimated one and a half to six hours [48]. The criteria for the selection of experts, the size of the group, the procedure and the criteria for reaching consensus should be clearly defined and the whole process documented, including how and when consensus was reached.

The technique of nominal groups is conducive to generating a large number of ideas and solutions, and so has a great potential for discussion and debate between experts. It is used to assess the legitimacy of using a specific intervention – a drug, a medical device in health

care, and ultimately the impact on the clinical decision making process. The course of the stages of establishing consensus using the nominal group technique depends on several factors. Among others, the size of the group, but also experts invited to cooperate. On the one hand, a large variety of experts representing different groups (e.g. representatives of doctors of different specialities, other medical personnel, patients' representatives, decision-makers) generates many ideas and solutions, stimulates creativity, which may be conducive to finding an optimal answer to a given nominal question. This method is often used in combination with the Delphi method.

On the other hand, the nominal group technique involves fewer participants than the Delphi method. The large diversity of the group may result in a more difficult communication, difficulties in setting common elements and priorities, interpersonal differences, and some representatives/experts may dominate over others, which may translate into difficulties in reaching a common consensus, for example due to the lack of anonymity. Furthermore, this method requires a face-to-face meeting, and therefore generates higher costs than the Delphi method.

RAND/UCLA method

The RAND/UCLA Appropriateness Method (RAM) was developed in 1980 by the University of California in Los Angeles (UCLA) and RAND Health, and was used and refined in North America and Europe in the following years. The RAND/ULCA method combines elements of the Delphi and nominal group techniques. This technique, used in the so-called appropriateness study, was aimed at identifying and assessing the benefits or detriment of medical or surgical procedures. Randomized studies with a control group, which are a gold standard, are not always available. There are situations when there is no strong scientific evidence indicating the benefits of using given procedures and doctors have to make clinical decisions in everyday medical practice. Therefore, the aim of the RAND/UCLA method is to systematically combine the best available and up-to-date scientific evidence with expert advice in order to create guidelines for health areas for which precise scientific data are lacking. The method has become an important tool for quality assessment in medicine and aims to reach a formal agreement on how research results should be interpreted and how they can be used under real conditions. This method indicates principles that can help to develop reliable guidelines. It is worth mentioning that although it is commonly called and used as a consensus method, in reality it is not – its aim is to assess the level of agreement between experts and not, in itself, to reach a consensus [40].

The first step in the RAND/UCLA method is to select and define the research problem to be discussed by the expert group. The next step is to collect current scientific evidence and guidelines for the previously defined research problem. At this stage, a literature review and participant questionnaires should be developed with a list of recommendations for assessing validity. These

documents may include e.g. patient characteristics and indications for management or specific clinical scenarios. All information, i.e. the literature review, the necessary definitions to avoid different interpretations, and the lists and statements to be completed shall be sent to the subjects. The process consists of two rounds. In the first round, as in the Delphi method, participants do not meet face-to-face, but fill in and assess the individual elements individually (questionnaire sent by mail or e-mail). At this stage the experts assess the relevance and feasibility of the listed interventions/procedures on a qualitative and/or quantitative scale. Each intervention/procedure is assessed in terms of the benefits and potential harm it could generate on a numerical scale, originally 1 to 9, where 1 means expected harm far outweighs expected benefits, and 9 means expected benefits far outweigh expected harm [40].

The next stage draws on the technique of nominal groups as at this stage direct meetings of experts are held face-to-face. During the meeting, the answers of each member are presented to the group. The participants have the opportunity to compare their assessments and decisions and then discuss the individual positions. Then another round of assessments takes place. The final result can be classified into one of three groups: 'appropriate', 'uncertain', or 'inappropriate'. Different methods are used to summarise the results. To assess the levels of appropriateness, a median rating panel method is used which classifies results as follows: ranges 1–3 as inappropriate, ranges 4–6 as uncertain, and ranges 7–9 as appropriate. Inappropriate scores are also assigned in the absence of consensus [40].

The second approach to the presentation of results is the measurement of disagreement. The results can be analysed by estimating the Interpercentile Range Adjusted for Symmetry (IPRAS). The value of IPRAS is necessary to determine the lack of consensus and depends on the internal symmetry of the assessments. For this purpose, the inter-percentile range (IPR), which is the difference between the lower and upper limits of the inter-percentile range, is estimated:

$$IPR = \text{lower IPR} - \text{upper IPR}$$

The central point of IPR (central point of IPR, IPRAS) is the arithmetic mean of the lower and upper IPR (arithmetic mean of the 30th and 60th percentiles), and is the basis for estimating the asymmetry index ($AI = 5 - IPRCP$). The IPRAS is the sum of 2.35 (the inter-percentile interval required for disagreement assuming full symmetry) and the quotient of 1.5 (correction factor for asymmetry) and the asymmetry index. If the IPRAS is greater than the IPR there is agreement, otherwise there is disagreement (Disagreement Index; $DI = IPR/IPRAS$).

The RAND/UCLA method requires the collection of the necessary scientific reports on the problem before the experts can give their opinion and assess it; it is not a technique to generate new ideas and solutions. The data collected is used to develop a questionnaire for evaluation, and therefore the literature review stage

is very important. This method also requires a moderator/facilitator to coordinate the entire process. Initially, only clinicians were involved in the process; however, with the popularization of the method and increasing complexity of health care services, the group of experts has started to include representatives of other groups depending on the analyzed area, e.g. nurses, physiotherapists, radiologists, patients' representatives and their carers. A group of panelists may be homogenous and consist only of doctors from a given specialization, e.g. clinical oncologists; however, more and more often it takes the form of a multidisciplinary team. The RAND/UCLA manual indicates that the optimal number of panelists is nine (or between seven and fifteen). The size of the group is large enough to ensure a variety of positions and opinions and at the same time small enough to allow all experts to participate actively in the discussion [40].

When there is a large variety of stakeholders involved in a given process (health problem), it is possible to create several independent panels that assess different parts of the problem. This approach has been used in the evaluation of validation and the possibility to standardize patient-centred geriatric surgery in the USA by The Coalition for Quality in Geriatric Surgery. This organization brings together 58 different organizations and entities (doctors, surgeons, geriatric surgeons, payers, patient organizations, nursing staff, social workers, legislative bodies). Its aim is to set high quality standards of perioperative care for older people. The Coalition for Quality in Geriatric Surgery decided to involve all representatives in the process of setting quality standards by dividing them into three smaller, interdisciplinary panels, which at the same time participated in the process of consensus setting. By doing so, the authors of the study wanted to check the differences in accuracy of results between representatives of different groups (surgeons versus doctors of other specialty), compare them with the results of representatives of groups that are not clinicians, and assess the level of compatibility between three parallel multidisciplinary teams. The results showed that surgeons are more critical of the standards than physicians of other specialties. However, three independent panel discussions concluded that 96–100% of rules/standards are appropriate. The authors concluded that the RAND/UCLA method can be successfully applied to large stakeholder groups [49].

The advantage of the method is the development of a literature review and available, up-to-date scientific evidence before the start of the consensus process. It allows both individual (anonymous) assessments by participants and group discussions. The involvement of many stakeholders allows for the creation of an interdisciplinary panel in order to take different perspectives on a given problem. On the other hand, this form of consensus-building requires large financial and organisational resources, and the use of a 9-stage scale for evaluation can be problematic for evaluators and can cause difficulties in categorisation. When developing clinical guidelines, it is often necessary to vote on many topics/issues and scenarios.

The RAND/UCLA method can be used in the process of developing clinical guidelines, selecting quality indicators for patient care, defining appropriate criteria for medical interventions [50].

Consensus conference

The U.S. National Institutes of Health, which assesses and disseminates information about medical technologies and research, has developed a conference panel consensus model. The aim of the conference panel consensus is to develop a mutually acceptable declaration of consensus on a pre-defined topic. The conference panel consensus conferences bring together representatives of different groups – doctors, researchers, patients, and decision makers. This process includes a literature review, a summary of the current state of knowledge, expert presentations and discussions. Conferences often last for two or more days and aim to develop an agreed written common position. Many European countries have developed their own versions of consensus conferences.

Consensus conferences are based on a prior review of current scientific evidence. The results of the research and the existing international guidelines are presented orally at the conference, but information can also be made available to participants before the conference. During the conference, previously described methods may be used to establish consensus. The number of experts participating in the panel may vary, but usually ranges from 9 to 18 (38.51). Nair (2011) indicates an optimal number of 10 panelists [50].

One of the most common methodologies of the NIH consensus development conference was developed by the National Institutes of Health in 1997. The Office for Medical Applications of Research (OMAR) NIH is responsible for organizing the conference and assisting in the direct use of biomedical research results in medical practice. Each conference is jointly funded and organized by one or more NIH and OMAR Institutes or Centres (ICs). Issues for which consensus positions are prepared concern important public health issues. Prior to the conference, experts define research problems and define the scope of the conference. Another group of experts (about 10) is invited to participate in a panel or decision-making group. The scientific evidence is then presented to the conference by various experts appointed by the IC and OMAR, who are not members of the decision-making group. Typically, the Agency for Research and Quality in Health Care (AHRQ) prepares a systematic literature review. During the presentation the experts of the panel or decision-making group and conference participants can participate in an open discussion. In the next stage, the panel of experts meets in closed discussion to talk about the evidence presented and to establish consensus. The Chair is the moderator responsible for leading and conducting both the open part of the conference and the closed discussion in the panelists' group, and helps to reach a consensus on a topic/question on which two or more experts have differing opinions. The draft consensus position is then presented at a plenary session. After a possible discussion by

participants in the session, the panel of experts may modify the position if it considers it appropriate. The final consensus position is published and distributed in order to achieve maximum impact on clinical practice and medical research [50].

■ Examples of the application of consensus methods in the guidance process

European Society for Medical Oncology

The European Society for Medical Oncology (ESMO) is a leading European organisation in the field of medical oncology developing – among other things – clinical practice guidelines, expert positions or recommendations established at consensus conferences. The ESMO has formal operating procedures for conducting consensus conferences. Members participating in consensus conferences are experts in the field coming both from and outside Europe. The group usually consists of up to 40 people. Prior to a consensus conference, expert working groups are set up (5–10 members of the group) and these are assigned specific issues and prepare questions on the basis of which recommendations will be developed (work mainly by e-mail or teleconference). The modified Delphi method is used at this stage. The members of the group are also required to review the literature and find scientific evidence on the formulated questions, which are then presented at the consensus conference. If the Chair considers it appropriate to conduct a systematic review, it should be developed by the working group. The conference usually takes three days (but may be longer), during which the objectives, scope and process are presented. Thereafter, parallel sessions of the working groups take place. Each working group discusses topics/questions with a view to making recommendations, and assesses the level of scientific evidence. Divergent opinions should be recorded and presented at the plenary session together with a summary report of the session; no voting takes place at this stage. The plenary session is followed by a discussion on each clinical issue/question and a vote (percentage representation of votes following the pattern of 'I agree', 'I disagree', 'I abstain'). The ultimate stage is the final plenary session where the questions and recommendations are summarised together with the percentage of agreement for each recommendation. The President agrees with the members the next steps of developing the final document and the deadline for the publication of the recommendations in the form of an expert position developed at the consensus conference. If a consensus conference fails to achieve the required consensus among the experts for a specific recommendation (lack of agreement or agreement < 75% or disagreement > 20%), the post-conference consensus shall be carried out by taking into account the GRADE methodology for assessing the quality of scientific evidence. The recommendations for which there are inconsistencies are sent with the GRADE form to the experts for completion within 72 hours. Then the results are combined, and if a < 75% consensus is reached,

the developed recommendations are sent again to the experts asking them to vote again [52].

British Association for Surgery of the Knee

Examples of the application of many formal consensus methods at different stages are the guidelines and expert position papers developed by the British Association for Surgery of the Knee (BASK) on the treatment and classification of patients with knee meniscus lesions, including the definition of criteria for the eligibility of patients for surgical treatment. The working group included 21 experts who are surgeons. Delphi, nominal groups and RAM methods were used throughout the process. The first stage was a meeting to determine the scope of the terminology used, the scope of required studies, and the classification of meniscus lesions. During the meeting the technique of nominal groups was used. Then a literature review concerning arthroscopic meniscectomy was carried out. Following the Delphi method, questionnaires with proposed recommendations were sent to group members in order to agree the position. After receipt of all responses, changes were reviewed and made, and the group was again asked to respond to individual points. Based on the findings of the Delphi population definition, a total of 45 different clinical scenarios were prepared and evaluated by the panel of experts. For this purpose the RAM method was used – the first round was conducted anonymously between meetings. A 9-grade scale was used and the degree of compatibility between the participants was assessed using the median and the dispersion between the assessments. After the first round of RAM voting, the group met to discuss the results. Each participant received a summary of their assessments and their co-participants. During the meeting, all the scenarios with regard to which there were inconsistencies in the expert assessments were presented and discussed. Afterwards, participants reassessed all scenarios. The aim of the two rounds of voting was to distinguish the real clinical inconsistency from possible fatigue, misunderstanding or errors ('artefactual' disagreement). After the second round of RAM, 11 scenarios (24%) were assessed as appropriate for arthroscopic meniscus surgery, 24 scenarios (53%) as inappropriate,

and 10 scenarios (22%) were evaluated as 'uncertain'. A recommendation was possible for 42 (96%) scenarios, with a compliance threshold of 50%. The other two scenarios were discussed by the steering group and recommendations were formulated based on the number of votes and other criteria. For 10 uncertain scenarios, patients were re-evaluated and classified as cases to be re-evaluated after further non-surgical treatment.

The scenarios were then grouped according to similar clinical characteristics of patients in order to make more general recommendations.

On their basis, treatment regimens were developed and sent to participants for their opinion. Based on the opinion, amendments were made before the next meeting, at which the guidelines were discussed in detail and taking into account the latest scientific evidence. The changes agreed at this meeting were introduced into the guidelines and then presented at the annual BASK conference and sent electronically to BASK members for their opinion and final approval. The final form of the guidelines was approved by the working group [53].

■ Summary

Clinical guidelines have become an important tool to influence clinical practice. Many international and national scientific societies develop recommendations or adapt guidelines developed by other organisations. Over the years, the panels of experts involved in the development of top-quality guidelines have become more and more extensive, involving different groups interested in the implementation of guidelines. The participation of many experts representing the opinions and interests of the different groups leads to a methodological and structured approach. The aim is to ensure that all participants have the opportunity to express their views, as well as to ensure the transparency of the process and to deal with misunderstandings.

In practice, a combination of different formal consensus methods or their modifications are often used. In the first stage of generating research problems or the issues to be discussed one method can be used as a preliminary consensus (e.g. Delphi), and

Criterion	Delphi method	Normal group technique	RAND/UCLA method	Consensus Conference
Scope of the panel	multidisciplinary	multidisciplinary	homogenous or multidisciplinary	multidisciplinary
Number of panelists	6–11	5–12	7–15	9–18
Need for a direct meeting	No	Yes	Yes	Yes
Literature review	Yes – indispensable	If the problem under consideration requires review. Not obligatory	Yes	Yes – results research/data presented by experts
Anonymity of assessments/opinions	Yes, fully anonymous		Yes, partially	No
Moderator/Chair	Yes	Yes	Yes	Yes

Table I. Characteristics of formal consensus methods

Source: Own work.

then another method could be used for final consensus (e.g. RAND/UCLA or consensus conference). If the rules and steps for a method are not strictly adhered to, the solution is called the modified consensus method. The way consensus is reached and recommendations made should be transparent and documented.

Formal consensus methods can provide a wide range of stakeholders – doctors, decision makers, patients – with agreed rules on a given topic that combine scientific evidence with the practice and experience of experts. It should be remembered that consensus should be based on a literature review and not solely on the opinions of participants.

The main features of these methods of formal consensus are presented in **Table I**.

Notes

¹ The latter are also called performance standards or a minimum standard.

² Established in 2002, Guidelines International Network (G-I-N) is a network of guidelines' authors which comprises 93 organisations and 89 individual members representing 46 countries.

³ The Peer Review method (a qualitative method of assessing scientific achievements, carried out on the basis of merit criteria by other scholars having similar or better scientific achievements and at a comparable stage of their scientific career) is considered to be the most reliable.

References

1. Woolf S., Schünemann H.J., Eccles M.P., Grimshaw J.M., Shekelle P., *Developing clinical practice guidelines: Types of evidence and outcomes; values and economics, synthesis, grading, and presentation and deriving recommendations*, "Implementation Science" 2012; 7 (1): 61.
2. Institute of Medicine (US). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Graham R., Mancher M., *Clinical Practice Guidelines We Can Trust*, National Academies Press, Washington, DC 2011.
3. Field M.J., Lohr K.N. (eds.), *Clinical practice guidelines: Directions for a new program*, National Academies Press, Washington, 1990.
4. Jacobs C., Graham I.D., Makarski J., Chasse M., Ferguson D., Hutton B., Clemons M., *Clinical practice guidelines and consensus statements in oncology – an assessment of their methodological quality*, "PLoS One" 2014; 9 (10): e110469.
5. Guyatt G., Oxman A.D., Akl E.A., Kunz R., Vist G., Brozek J.,... Rind D., *GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables*, "Journal of Clinical Epidemiology" 2011; 64 (4): 383–394.
6. Leśniak W., Bała M., Jaeschke R., Brożek J.L., *Od danych naukowych do praktycznych zaleceń – tworzenie wytycznych według metodologii GRADE*, "Polish Archives of Internal Medicine" 2015; 125: 26–41.
7. Qaseem A., Forland F., Macbeth F., Ollenschläger G., Phillips S., van der Wees P., *Guidelines International Network: Toward international standards for clinical practice guidelines*, "Annals of Internal Medicine" 2012; 156 (7): 525–531.
8. Hutchings A., Raine R., *A systematic review of factors affecting the judgments produced by formal consensus development methods in health care*, "Journal of Health Services Research & Policy" 2006; 11 (3): 172–179H.
9. Murphy E., Dingwall R., Greatbatch D., Parker S., Watson P., *Qualitative research methods in health technology assessment: a review of the literature*, "Health Technology Assessment" 1998; 2 (16): iii–ix, 1–274.
10. Shekelle P.G., Schriger D.L., *Evaluating the use of the appropriateness method in the Agency for Health Care Policy and Research Clinical Practice Guideline Development process*, "Health Services Research" 1996; 31 (4): 453.
11. Jaeschke R., Guyatt G.H., Dellinger P., Schünemann H., Levy M.M., Kunz R.,... Bion J., *Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive*, "BMJ" 2008; 337: a744.
12. Moyano M.S., Rivera J.L., *Materiality Analysis of Health Plans Based on Stakeholder Engagement and the Issues Included at ISO 26000: 2010*, "Revista Espanola de Salud Publica" 2017; 91.
13. Santric-Milicevic M., Terzic Z., Markovic G., Filipovic J., *Social responsibility in health system: An approach to assess ISO 26000 application in Serbia Milena Santric-Milicevic*, "European Journal of Public Health" 2013; 23 (suppl. 1).
14. Kumakawa T., *Social responsibility in healthcare system: ISO 26000 and socially responsible investment*, "Journal of the National Institute of Public Health" 2009; 58 (3): 283–293.
15. Polski Komitet Normalizacyjny, *Norma I.S.O. 26000: 2010*, Warszawa 2010.
16. Norma Międzynarodowa dotycząca społecznej odpowiedzialności. Polski Komitet Normalizacyjny 2013, <https://www.pkn.pl/informacje/2013/09/iso-26000> (accessed: 20.06.2019).
17. Jastrzębska E., *Angażowanie interesariuszy jako istota społecznej odpowiedzialności według ISO 26000. 2016*, <https://depot.ceon.pl/handle/123456789/13165> (accessed: 6.06.2019).
18. Massoni F., Ricci P., Ricci S., *Guidelines and evidence based medicine. The importance of stakeholder involvement*, "Reviews in Health Care" 2014; 5 (3): 109–115.
19. Shekelle P.G., Woolf S.H., Eccles M., Grimshaw J., *Developing guidelines*, "BMJ" 1999; 318 (7183): 593–596.
20. Schmeer K., *Stakeholder Analysis Guidelines. Policy Toolkit for Strengthening Health Sector Reform*, WHO, Regional Office of the World Health Organisation, Washington, 1999: 1–33.
21. National Institute for Clinical Excellence, *How NICE clinical guidelines are developed: An overview for stakeholders, the public and the NHS. 2009*, July, 20, 2014.
22. Ahmad R., Kyratsis Y., Holmes A., *When the user is not the chooser: Learning from stakeholder involvement in technology adoption decisions in infection control*, "Journal of Hospital Infection" 2012; 81 (3): 163–168.
23. Kahan J.P., Park R.E., Leape L.L., Bernstein S.J., Hilborne L.H., Parker L.,... Brook R.H., *Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures*, "Medical Care" 1996; 34: 512–523.

24. Coulter I., Adams A., Shekelle P., *Impact of varying panel membership on ratings of appropriateness in consensus panels: A comparison of a multi- and single disciplinary panel*, "Health Services Research" 1995; 30 (4): 577.
25. Leape L.L., Park R.E., Kahan J.P., Brook R.H., *Group judgments of appropriateness: The effect of panel composition*, "International Journal for Quality in Health Care" 1992; 4 (2): 151–159.
26. Pagliari C., Grimshaw J., Eccles M., *The potential influence of small group processes on guideline development*, "Journal of Evaluation in Clinical Practice" 2001; 7 (2): 165–173.
27. Moscovici S., *Social Influence and Social Change*, European Monographs in Social Psychology, Vol. 10, Academic Press, London 1976.
28. Wesółowska E., *Polaryzacja grupowa w warunkach debaty deliberatywnej*, "Civitas et Lex" 2014; 2: 19–30.
29. Janis I.L., *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*, Houghton Mifflin, Oxford, England 1972.
30. Ellemers N., Wilke H., Van Knippenberg, A., *Effects of the legitimacy of low group or individual status on individual and collective status-enhancement strategies*, "Journal of Personality and Social Psychology" 1993; 64 (5): 766.
31. Eccles M., Clapp Z., Grimshaw J., Adams P.C., Higgins B., Purves I., Russell I., *North of England evidence based guidelines development project: Methods of guideline development*, "BMJ" 1996; 312 (7033): 760–762.
32. Humphrey-Murto S., Varpio L., Gonsalves C., Wood T.J., *Using consensus group methods such as Delphi and Nominal Group in medical education research*, "Medical Teacher" 2017; 39 (1): 14–19.
33. Delphi method, <https://www.rand.org/topics/delphi-method.html> (accessed: 20.04.2019).
34. Wood F., Bloor M., *Keywords in Qualitative Methods: A Vocabulary of Research Concepts*, Sage Publications, London 2006.
35. Linstone H.A., Turoff M., *The Delphi Method: Techniques and Applications*, 2002, <https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf> (accessed: 20.04.2019).
36. Murphy M.K., Black N.A., Lamping D.L. et al., *Consensus development methods and their use in clinical guideline development*, "Health Technology Assessment" 1998; 2: 1–88.
37. Cantrill J.A., Sibbald B., Buetow S., *The Delphi and nominal group techniques in health services research*, "International Journal of Pharmacy Practice" 1996; 4 (2): 67–74.
38. Waggoner J., Carline J.D., Durning S.J., *Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research*, "Academic Medicine" 2016; 91 (5): 66–668.
39. Avella J.R., *Delphi panels: Research design, procedures, advantages, and challenges*, "International Journal of Doctoral Studies" 2016; 11 (1): 305–321.
40. Fitch K., Bernstein S.J., Aguilar M.D., Burnand B., La-Calle J.R., *The RAND/UCLA appropriateness method user's manual (No. RAND/MR-1269-DG-XII/RE)*, Rand Corp., Santa Monica Ca 2011.
41. Njuangang S., Liyanage C., Akintoye A., *Application of the Delphi technique in healthcare maintenance*, "International Journal of Health Care Quality Assurance" 2017; 30 (8): 737–754.
42. Fink A., Kosecoff J., Chassin M., Brook R.H., *Consensus methods: Characteristics and guidelines for use*, "American Journal of Public Health" 1984; 74 (9): 979–983.
43. Sudol S., *Delficka metoda badawcza*, "Zarządzanie. Teoria i Praktyka" 2016; 17 (3): 69–74.
44. McKenna H.P., *The Delphi technique: A worthwhile research approach for nursing?*, "Journal of Advanced Nursing" 1994; 19, 1221–1225.
45. Van Teijlingen E., Pitchforth E., Bishop C., Russell E., *Delphi method and nominal group technique in family planning and reproductive health research*, "Journal of Family Planning and Reproductive Health Care" 2006; 32 (4): 249–252.
46. Horton J.N., *Nominal group technique: A method of decision-making by committee*, "Anaesthesia" 1980; 35 (8): 811–814.
47. Delp P., Thesen A., Motiwalla J., Seshardi N., *Nominal group technique. Systems tools for project planning*, "Annals of Global Health" 1977; 14–18.
48. Campbell S.M., Cantrill J.A., *Consensus methods in prescribing research*, "Journal of Clinical Pharmacy and Therapeutics" 2001; 26 (1): 5–14.
49. Berian J.R., Baker T.L., Rosenthal R.A., Coleman J., Finlayson E., Katlic M.R.,... Russell M.M., *Application of the RAND-UCLA Appropriateness Methodology to a Large Multidisciplinary Stakeholder Group Evaluating the Validity and Feasibility of Patient-Centered Standards in Geriatric Surgery*, "Health Services Research" 2018; 53 (5): 3350–3372.
50. Nair R., Aggarwal R., Khanna D., *Methods of formal consensus in classification/diagnostic criteria and guideline development*, "Seminars in Arthritis and Rheumatism" 2011; 41 (2): 95–105.
51. McGlynn E.A., Kosecoff J., Brook R.H., *Format and conduct of consensus development conferences: Multination comparison*, "International Journal of Technology Assessment in Health Care" 1990; 6 (3): 450–469.
52. Consensus Conference Standard Operating Procedures ESMO Guidelines Committee, latest revision April 2019, <https://www.esmo.org/content/download/77792/1426729/file/ESMO-Consensus-Conferences-Standard-Operating-Procedures-Apr2019.pdf> (accessed: 5.06.2019).
53. Abram S.G.F., Beard D.J., Price A.J., *BASK Meniscal Working Group. Arthroscopic meniscal surgery: A national society treatment guideline and consensus statement*, "The Bone & Joint Journal" 2019; 101 (6): 652–659.