

Systems for grading the strength of recommendations in clinical practice guidelines in oncology

Anna Rychert, Dominik Roman Dziurda, Magdalena Koperny, Magdalena Maria Krasztel,
Katarzyna Joanna Kędzior, Wojciech Wysoczański, Roman Topór-Mądry

Agency for Health Technology Assessment and Tariff System, Warsaw, Poland

Introduction. In order to improve the applicability of clinical practice guidelines, their authors assign recommendations with grades denoting the degree of conviction regarding their practical application. Nevertheless even within one branch of medicine, significant differences between the grading systems arise.

Material and methods. To identify these systems, websites of societies and institutions publishing oncology guidelines were searched. Only high-quality, regularly updated guidelines were included.

Results. Five systems were analysed – all incorporate quality of evidence and strength of recommendation, but vary in the methods of their assessment and structure of the scales.

Discussion. The described systems depend on the review of data, the quality of which supports the ascribed strength. Systems differ with regard to the methods of assessing the quality, quantity and consistency of evidence, potentially leading to assigning different grades of strength to recommendations based on the same studies.

Conclusions. The introduction of unified grading systems across each branch of medicine could aid the development of unambiguous recommendations that are easy to introduce within the healthcare system.

Key words: grading system, guidelines, recommendation, quality of evidence, strength of recommendation

Introduction

Increasingly often, decisions concerning diagnostics and treatment in contemporary medicine are made in accordance with the paradigm of evidence-based medicine (EBM). However, due to the increasing number of clinical studies conducted and published, as well as the need to adapt the quality and effectiveness of health care to changing conditions, it has become necessary to systematize the gathered knowledge and make it more accessible. Various documents such as clinical practice guidelines (sets of recommendations) or healthcare standards are being developed in response to these needs.

In this work, the significance of the “strength of recommendation” parameter in the practical implementation of guidelines is thoroughly discussed. As the clinical practice guidelines

employ varied methods of characterizing the strength of recommendation, the aim of this work is to present the grading systems most frequently used in the area of oncology. This study does not exhaust the broad subject of methodology of clinical practice guideline development, but it does constitute a review of the most popular grading systems used in oncology guidelines.

Clinical practice guidelines

According to the definition by the American Institute of Medicine, clinical practice guidelines (CPGs) constitute a summary of scientific evidence “that are informed by a systematic review and an assessment of the benefits and harms of alternative care options” [1]. The CPGs are not only aimed at facilitating

Jak cytować / How to cite:

Rychert A, Dziurda DR, Koperny M, Krasztel MM, Kędzior KJ, Wysoczański W, Topór-Mądry R. *Systems for grading the strength of recommendations in clinical practice guidelines in oncology*. NOWOTWORY J Oncol 2021; 71: 407–416.

the decisions made in specific clinical situations, but they also influence the effectiveness and quality of diagnosis and therapy. The keynote of the guideline development process lies in the strict relationship between the recommendations and the gathered evidence. The process itself consists of many stages, is systematized, consistent with specific quality criteria, and based on a systematic review of literature. In addition to that it involves an assessment of quality and selection of the scientific evidence that will serve as a basis for development of recommendations [2].

From the perspective of the user of such recommendations, proper interpretation of the degree of trust in their content and confidence in the rationale for their application are of key importance. Therefore, methodologically correct guidelines for clinical practice should transparently present methods for the development and assessment of recommendations, as well as the logical connection between alternative care options and health results, plus an appraisal of the quality of evidence and strength of recommendations [1].

Quality of evidence and strength of recommendation

While determining the extent of authors' acceptance for the content of the recommendation (the strength of recommendation), four components are taken into consideration: the quality of scientific evidence (for a single study), the quality of overall evidence gathered, the strength of intervention and the benefit-risk balance. It is a complex process, which makes it necessary to distinguish between the following concepts:

- **The quality of evidence for a single study** referring to the impact of the methodological structure of a clinical trial upon the uncertainty of estimation of intervention results for a specific endpoint in a specific population in a single study.
- **The quality of evidence** describing the quality of the overall evidence gathered on the clinical profile of the intervention in relation to the defined endpoint. It defines the degree of certainty that the available scientific evidence reflects the true dimensions and direction of effects in the context of the target healthcare system's conditions. It is also referred to as strength, certainty or level of evidence.
- **The strength of intervention** refers to the effectiveness of the intervention; it illustrates the magnitude of the achievable effect of the new intervention in comparison to other available options in the population subject to the recommendation.
- **The balance of benefits and harms** – a description of magnitude of benefits in relation to damages/side effects/threats associated with a given intervention.

In the context provided above, **the strength of recommendation** defines the degree of authors' conviction that the content of the recommendation should be applied in clinical practice taken the conditions of the target healthcare system.

The process for determining the strength of recommendations is based on quality of evidence, absolute and relative strength of intervention and the degree of consensus with regard to implementation in clinical practice.

Appraisal of evidence, or description of the degree of certainty that conclusions based on the collected evidence are reliable, constitutes one of the most significant factors, though not the only one, necessary to determine the ascribed strength of recommendation. It is based on the type of clinical trials included in the overall evidence. In accordance with the hierarchy of scientific evidence applied by the World Health Organization, systematic reviews and meta-analyses of RCTs are considered to be of the highest quality, while descriptive studies and expert opinions – of the lowest [3]. In this regard, it is key to appraise the internal and external reliability of the studies by considering other factors that may influence the quality of evidence, such as the risk of bias or inaccurate estimation of effects [4]. These factors – assessed using approved tools (i.a. AMSTAR2, RoB 2.0) – may result in lowering or raising the preliminary quality score determined by the study type. Hence the established quality of gathered scientific evidence, serving as a basis for recommendations, is known as the quality of evidence.

While clinical effectiveness of a given intervention, considering the uncertainty of study results, should be the main factor determining the strength of recommendation [5], the authors, when assigning the strength to each of the recommendations, consider various interrelated factors, such as:

- quality of evidence justifying the recommendation,
- strength of intervention,
- applicability of the evidence to the target clinical conditions,
- certainty in relation to basic risk (the occurrence of a given outcome (event) when a standard procedure is applied).

However, this process also accounts for various other aspects, such as uncertainty in relation to the values and preferences of patients and the significance of the effects of a given intervention – patients' expectations and objectives in terms of quality of life, or experience with the illness – as well as social equality and justice, costs, available resources, acceptability of recommendations and possibilities of utilising alternative treatments. In order to determine the strength of recommendation, it is necessary to provide a full and transparent summary of all the indicated components and to identify the potential effects of the CPGs' implementation. Among the desirable effects, health improvement significant to the patient or cost reduction are of the greatest importance. Whereas, the unfavourable effects include, i.e. adverse effects, a significant increase in organisational or cost burden resulting from application of the procedure compliant with the guideline [6].

The systems for grading strength of recommendation in diagnostics and therapy therefore combine two aspects: the quality of evidence, based on the objective and precise process

of appraisal of the research methodology, and the authors' certainty regarding the presence of reasonable grounds to apply the guideline in practice.

In general, evidence of high quality should result in strong recommendations. However, considering the balance of favourable and unfavourable effects of the intervention, or possible differences between the settings and conditions in the trials and those under consideration, as well as other factors, a recommendation might be assigned a much lower strength. As a result, recommendations may be weak despite a reliable estimation of the clinical effect, or strong despite the poor quality of the estimations, the reason being the need (or even the necessity) to consider prerequisites other than analytical data when developing the guidelines [6].

Due to the multitude of factors which should be considered in the process of recommendation assessment, a risk of discrepancies in the systems employed in various guidelines arises. Such situations may lead to a substantial weakening of the CPGs' implementation potential. Therefore, most organisations utilise commonly known appraisal systems or develop their own methodologies for grading the strength of recommendation to ensure its reliability. The common feature of all these documents lies in the transparency of factors considered during their development, relying upon the EBM principles and utilising systematic reviews of evidence as a basis.

Material and methods

The review of the systems for grading recommendations has been prepared on the basis of selected CPGs developed worldwide by oncology societies and governmental organisations. The thematic scope has been limited due to the legal conditions in Poland, which state that the Minister of Health announces, in the form of a notification, the guidelines for the diagnostic and therapeutic procedure regarding cancer treatment. Additionally, most oncology guidelines treat the process holistically – from prevention and screening to rehabilitation and follow-up – and include recommendations addressed to service providers, patients and their caregivers.

In order to identify the applied grading systems, a review was conducted involving 11 websites of science societies and organisations publishing oncology recommendations. Selection was conducted considering the societies recognized by clinical experts in Poland, which systematically publish new guidelines and update the older ones based on the most recent scientific evidence and global trends, as well as publish the methodology of guideline development. The documents analysed had to be of high quality as characterised by the AGREE II instrument (Appraisal of Guidelines for Research and Evaluation II) [7]. Only systems published in English were included. Systems based solely on appraisal of research quality were excluded.

In the case of each of the systems included in this study, the methodology was described as presented in handbooks for authors and "Guidelines for guidelines" documents. After-

wards, the factors considered when defining and determining the strength of recommendations in the individual systems were compared.

Results

The analysis took into consideration the grading systems applied by the following groups of methodologists, scientific societies and organisations:

- Grading of Recommendations Assessment, Development and Evaluation (GRADE),
- National Comprehensive Cancer Network (NCCN),
- National Institute for Health and Clinical Excellence (NICE),
- European Society for Medical Oncology (ESMO),
- Scottish Intercollegiate Guidelines Network (SIGN).

Grading of Recommendations Assessment, Development and Evaluation

Grading of Recommendations Assessment, Development and Evaluation (GRADE) offers a transparent and organised process for developing and presenting summaries of scientific evidence for the purpose of:

- preparation of systematic reviews,
- determination of healthcare standards,
- development of CPGs [8].

One of the main objectives of the GRADE group was to eliminate misunderstandings caused by different methods of appraising evidence and classifying recommendations used in healthcare. For this purpose a transparent approach to assess the quality of evidence and strength of recommendation was developed, which includes some strictly defined criteria for estimation of the strength of recommendation, as presented in table I.

To facilitate the development of recommendations, the authors of GRADE propose to categorise the quality of evidence gathered for a given endpoint using four grades: high, moderate, low and very low. It should be kept in mind that these are not purely quantitative; they also involve some quality-based decisions (fig. 1), which require experience not only in conducting systematic reviews and analysis of scientific evidence, but also clinical knowledge concerning a given health problem. The preliminary grade of quality of evidence may be either high or low, depending on the design of the studies informing the recommendation.

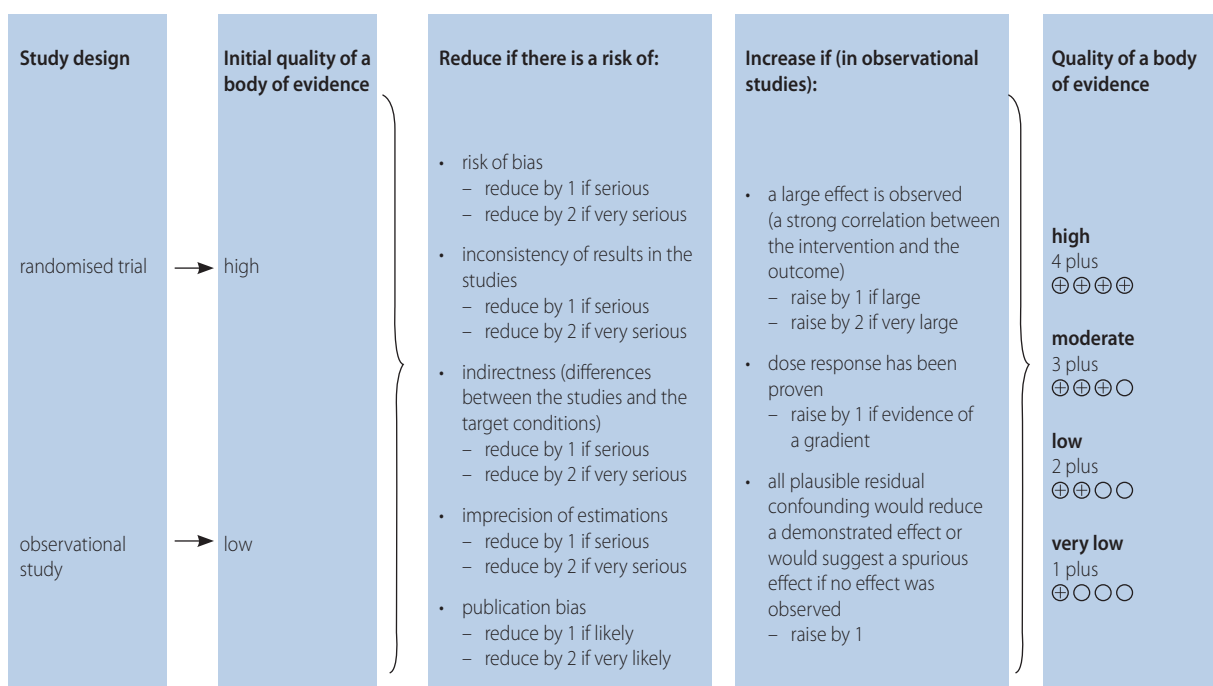
In order to facilitate the interpretation and application of recommendations, the GRADE introduced a descriptive four-step designation of quality of evidence (certainty of effect estimation) and a two-step designation of strength of recommendation, recognising two grades: strong or weak (conditional). Although the division into separate grades requires arbitrary decisions, GRADE assumes that this approach has more merits than detriments; among other things, it provides precise instructions for patients, physicians and healthcare system managers. As a result, recommendations can be classified as presented in table II.

Table I. The criteria that contribute to the strength of a recommendation according to GRADE

Domain	Comment
balance between desirable and undesirable outcomes (trade-offs) taking into account: • best estimates of the magnitude of effects on desirable and undesirable outcomes • importance of outcomes (estimated typical values and preferences)	the larger the differences between the desirable and undesirable consequences, the more likely a strong recommendation is warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely a weak recommendation is warranted
confidence in the magnitude of estimates of the effect of the interventions on important outcomes (overall quality of evidence for outcomes)	the higher the quality of evidence, the more likely a strong recommendation is warranted
confidence in values and preferences and their variability	the greater the variability in values and preferences, or uncertainty about typical values and preferences, the more likely a weak recommendation is warranted
resource use	the higher the costs of an intervention (the more resources consumed), the less likely a strong recommendation is warranted

The term "outcome" is used in accordance with the original source, although it does not fully correspond to the differentiation between terms "outcome" and "endpoint" employed in this work

Source: Andrews J.C. et al. (2013). *GRADE guidelines: 15. Going from evidence to recommendation – determinants of a recommendation's direction and strength*

**Figure 1.** The GRADE approach to rating the quality of a body of evidence – factors influencing the appraisal of quality of evidence

Source: Table compiled based on Leśniak W. (2015). *Od danych naukowych do praktycznych zaleceń – tworzenie wytycznych według metodologii GRADE*, and Balshem H. (2011). *GRADE guidelines: 3. Rating the quality of evidence*

Basically, a **strong** recommendation means that the authors are convinced that beneficial (adverse) effects of the intervention substantially outweigh the adverse (beneficial) effects; evidence in this regard (of appropriately high quality) is available. A **weak (conditional, optional)** recommendation means that while the evidence weighs in favour/against the intervention, the authors are not convinced of the significant advantage (disadvantage) of the intervention, either due to insignificant differences in effects, lack of data or low quality of data. In general, high-quality evidence should provide strong recommendations; however, other factors, such as cost effectiveness or opinions of patients may lead to a weak recommen-

dation. Also, in the case of low quality of evidence, additional factors may justify increasing the strength [8].

It should be underlined that the strength of a recommendation is not equivalent to high priority of such recommendation [8], which is particularly true when more than one recommendation is developed in order to consider co-morbidities or ethnic origin.

The main advantages of the GRADE system include:

- a clear grading of quality of evidence and strength of recommendation,
- a straightforward evaluation of significance of effects of alternative interventions,

Table II. Description of quality of evidence and strength of recommendation according to GRADE

Quality of evidence	
high	we are very confident that the true effect lies close to that of the estimate of the effect
moderate	we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
low	our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
very low	we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect
Strength of recommendation	
strong recommendation	in favour of the intervention
	against the intervention
weak recommendation	in favour of the intervention
	against the intervention

Source: Guyatt G. et al. (2013). *GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes*, and Andrews J.C. et al. (2013). *GRADE guidelines: 15. Going from evidence to recommendation – determinants of a recommendation's direction and strength*

- comprehensive criteria for lowering and raising the grades of evidence quality,
- a transparent process of moving from evidence to recommendations,
- recognition of values and preferences of stakeholders,
- an explicit, pragmatic interpretation of strong and weak recommendations for both clinicians, patients and decision-makers.

Due to its comprehensiveness and clarity, the GRADE methodology is presently recognised as a standard for the guideline development process. It is used by the World Health Organisation, the Guidelines International Network, scientific societies, state agencies responsible for guideline development and HTA, e.g., in the United States, Canada, Belgium, and Germany.

The GRADE methodology is mainly applicable for evaluation of drug technologies, as well as surgery or radiotherapy procedures. It may be used for evaluation of other non-drug technologies; however, certain limitations should be expected during evaluation of quality of evidence. When defining the key questions in the assessment of quality of evidence for diagnostic technologies, it is necessary to differentiate clearly between selection of endpoints referring to accuracy of the diagnostic test and results which are of significance to the patients [11]. Moreover, to warrant reliability of evaluation, clear and accurate definition of the assessment criteria, and analysts' experience in application of GRADE methodology are of key significance, since they might cause discrepancies in the interpretation of data and lead to diversified grades with little coherence between individual analysts [12].

National Comprehensive Cancer Network

National Comprehensive Cancer Network (NCCN) is a non-profit organisation dedicated to improving the quality, effectiveness and efficiency of healthcare. Its CPGs are aimed at facilitating the decision-making processes in cancer care [13].

According to NCCN, a significant diversity of clinical study types in oncology – from big RCT trials to small retrospective studies – has made it necessary to place greater attention on the experience and opinion of specialists and clinical experts, and take them into account in the course of evaluating scientific evidence. To do so, the NCCN develops its recommendations based on the critical assessment of evidence, combined with the expertise and consensus of a multidisciplinary expert panel, especially in situations where high-quality evidence is lacking. Additionally, since most interventions in cancer treatment have adverse effects, the panel is obliged to evaluate them with attention to efficacy, utility, safety and toxicity [14].

The NCCN categories for recommendations (tab. III), which serve to denominate the strength of recommendation, are determined on the basis of quality of scientific evidence (single studies and overall evidence) and the stance (consensus) of the panel with regard to validity of the intervention. The panel consensus is determined based on voting on incorporating the recommendation. A uniform consensus, allowing the recommendation to be categorised as category 1 or 2A, requires the support of at least 85% of panel members. Consensus leading to the recognition of a recommendation as category 2B requires at least 50% of votes to support the recommendation. On the other hand, recommendations which are associated with substantial differences in opinions with regard to their validity must obtain at least 25% of votes to be included in the guidelines as category 3 [14]. All NCCN recommendations are considered appropriate and the guidelines do not indicate the interventions which in authors' opinions should not be used in the clinical practice, e.g. due to poor balance of benefits and harms.

Some of the documents developed by the NCCN, apart from the categories of recommendation, also classify interventions in terms of the categories of preference (tab. III). The scale has been developed, firstly – to describe institutional preferences and those of the panel, thus providing users with information on which recommendations are considered to be

Table III. NCCN categories for recommendations and categories of preference

NCCN categories for recommendations	
category 1	based on high-level evidence, there is uniform NCCN consensus that the intervention is appropriate
category 2A	based upon lower-level evidence, there is uniform NCCN consensus that the intervention is appropriate
category 2B	based upon lower-level evidence, there is NCCN consensus that the intervention is appropriate
category 3	based upon any level of evidence, there is major NCCN disagreement that the intervention is appropriate
NCCN categories of preference	
preferred intervention	interventions that are based on superior efficacy, safety, and evidence; and, when appropriate, affordability
other recommended intervention	other interventions that may be somewhat less efficacious, more toxic, or based on less mature data; or significantly less affordable for similar outcomes
useful in certain circumstances	other interventions that may be used for select patient populations (defined with recommendation)

Source: The National Comprehensive Cancer Network. *Development and Update of the NCCN Guidelines**

of choice; secondly – to describe the scope of the recommended interventions, which address various clinical situations and preferences of patients [14]. However, the NCCN methodology fails to transparently specify how the values and preferences of patients are taken into account during recommendation development.

Undoubtedly, the NCCN guideline process allows for a quick development of detailed recommendations on complex health problems such as neoplasms. However, the factors that allow the guidelines to be kept up-to-date are also their main limitation, with the documents lacking formal and transparent review and assessment of available trials. Thus, Wayant [15] suggests that, while panel members determine the quality of evidence in some fields, in order to enhance the objectivism, applicability and comparability of the NCCN guidelines, the GRADE approach should be adopted. At the same time, the recommendations include only a limited description of the assessed efficacy, safety, quality and consistency of evidence, and financial impact, which are additionally ascribed solely to the systemic therapy.

National Institute for Health and Clinical Excellence

National Institute for Health and Clinical Excellence (NICE) is a British agency responsible, among other things, for publishing guidance in four areas:

- use of both new and existing health technologies by the National Health Service (NHS),
- clinical practice (diagnosis and treatment),
- promotion of health and prophylaxis,
- social care.

The process of development of CPGs published by NICE is mainly informed by the appraisal of intervention's efficacy and cost effectiveness, considering existing circumstances, clinical conditions and patient preferences. Depending on the assessment of these factors, the recommendations can be ascribed a different level of authors' conviction regarding their strength (validity) – some of them can be established on

the basis of evidence of higher quality and greater certainty as to the positive effect of the treatment. As a result, NICE guidelines can be ascribed one of three grades of strength of recommendation [16]:

- interventions which must be applied (or must not be applied),
- interventions which should be applied (or should not be applied),
- interventions which can be applied.

According to the NICE methodology, the guidelines are considered to be "strong" if most experts and patients would choose this specific intervention – mostly due to the positive effects of therapy outweighing the adverse ones in relation to the cost effectiveness of a given intervention. However, if the balance of benefits and harms is not that clear, and many patients would not choose a given intervention, although some could decide to do so, the recommendation will be a weak one [16] (tab. IV).

The grading system applied by the NICE, based on the development of recommendations using the appropriate verbs (particularly modal verbs) and grammatical forms, is distinguished by its simplicity and ease of use; still, it only defines the degree of authors' certainty with regard to the application of the intervention. The NICE recommendations do not refer directly to the quality of evidence gathered, although, all of the scientific evidence identified is assessed using the GRADE methodology, and the results of this process are published in an annex to the guidelines [16]. At the same time, despite being very straightforward in its form, one needs to remember that every language is characterized by ambiguity, so the recommendations may not be considered that unequivocal to their user.

European Society for Medical Oncology

European Society for Medical Oncology (ESMO) publishes its guidelines with the aim of enhancing the quality and effectiveness of cancer care. The system for grading recommendations adopted in the ESMO guidelines is based on the Infectious Di-

Table IV. Examples of strength of recommendation designation used by the NICE

Examples of recommendations with different strength of recommendation
recommendations on interventions which must or must not be applied: provide treatment without undue delay for people who have lung cancer that is suitable for radical treatment or chemotherapy, or who need radiotherapy or ablative treatment for relief of symptoms
recommendations on interventions which should or should not be applied: offer surgery to people with rectal cancer (cT1–T2, cN1–N2, M0, or cT3–T4, any cN, M0) who have a resectable tumour
recommendations on interventions which could be applied: consider daily aspirin, to be taken for more than 2 years, to prevent colorectal cancer in people with Lynch syndrome

seases Society of America–United States Public Health Service Grading System [17]. Determination of the quality of evidence and strength of recommendation is mandatory for every statement, but the methodology does not require a systematic review; instead, it allows for the recommendations to be based on both the evidence gathered in a non-systematic review and expert opinions.

The quality of evidence (referred to by ESMO as the level of evidence) points to the quality of research reports collected (e.g., clinical studies, case / control studies, expert opinions), answering a clinical question and more specifically the number of included studies, their sample size, methodology, risk of bias and heterogeneity. According to ESMO, the strength of recommendation (known as grade of recommendation) considers both the quality of evidence and the significance / magnitude of the effect of the intervention (tab. V). This grade may be either positive (a recommended procedure) or negative (a non-recommended procedure). In order to avoid any interpretation difficulties, every recommendation must be expressed as a positive statement and assigned a strength of recommendation (indicating whether the procedure is to be applied or not). Negative statements are not to be used when formulating the recommendations.

While simplicity is definitely an advantage of the grading system used by ESMO, the limitations of this methodology

must be underlined – firstly, there is no obligation to hold a systematic review, thus the recommendations are formulated on the basis of subjectively selected studies, as well as expert knowledge and experience. Secondly, appraisal of the quality of scientific evidence is limited to determination of the study type and, possibly, assessment of the risk of bias (although the tools or criteria applied for this purpose have not been defined). Additionally, the ESMO methodology fails to include any formal procedure of patient involvement or collecting information on patient preferences [18].

Scottish Intercollegiate Guidelines Network

Scottish Intercollegiate Guidelines Network (SIGN) main objective of the SIGN is to improve healthcare by limiting the diversity in both clinical practice and the effects of diagnostic and therapeutic procedures. To achieve this, SIGN develops and disseminates the CPGs describing effective interventions based on current scientific evidence. Methods used by SIGN are founded on the GRADE methodology (especially, the appraisal of quality of evidence and strength of recommendation) and follow the quality standards for CPGs described in AGREE II instrument [7]. In order to develop recommendations in accordance with the EBM paradigm, that are both implementable and take into account the patient opinions, SIGN uses the Evidence-to-Decision framework grounded in GRADE [19].

Table V. The ESMO grading of recommendations

Levels of evidence	
I	evidence from at least one large randomised, controlled trial of good methodological quality (low potential for bias) or meta-analyses well-constructed randomised trials without heterogeneity
II	small randomised trials or large randomised trials with a suspicion of bias (lower methodological quality) or meta-analyses of such trials or trials with demonstrated heterogeneity
III	prospective cohort studies
IV	retrospective cohort studies or case-control studies
V	studies without a control group, case reports, expert opinions
Grades of recommendation	
A	strong evidence for efficacy with a substantial clinical benefit, strongly recommended
B	strong or moderate evidence for efficacy but with limited clinical benefit, generally recommended
C	insufficient evidence for efficacy or benefit does not outweigh the risk or the disadvantages (adverse events, costs, etc.), optional
D	moderate evidence against efficacy or for adverse outcome, generally not recommended
E	strong evidence against efficacy or for adverse outcome, never recommended

Source: The ESMO Guidelines Committee. (2020). *Standard Operating Procedures (SOPs) for Authors and templates for ESMO Clinical Practice Guidelines (CPGs) and ESMO-MCBS Scores*

As the guidelines are developed on the basis of consistent evidence, the ultimate wording of the recommendations is usually reached by informal consensus. When this cannot be reached, the evidence is interpreted by an independent supervision team involving external experts, leading to assigning strength to the recommendations, which can be either “strong” or “conditional”.

A strong recommendation is made when:

- the evidence is of high quality,
- estimates of the effects of the intervention are precise (that is, there is certainty that the effects will be achieved in practice),
- the intervention assessed has a limited number of negative factors,
- there is a high level of acceptance of the intervention among patients [19].

A recommendation is conditional if:

- the evidence is of low quality,
- there are doubts with regard to the magnitude of the effect that can be achieved in practice,
- the benefits and harms of the therapy must be balanced out,
- the acceptance of the intervention varies among patients [19].

Nevertheless, the particular quality of evidence does not automatically lead to a particular strength of recommendation. High-quality evidence should be associated with strong recommendations, but a consideration of the applicability of published

evidence to the target population, its relevance to the NHS and patients, and the balance of benefits and harms may lead to a much lower strength being assigned. Similarly, under some circumstances, when evidence is of lower-quality, but the intervention is characterized by low risk of harm and the problem is sufficiently significant, a strong recommendation can be justified [19]. Where equality, justice and co-morbidities should be accounted for, the authors may prepare several recommendations answering that particular question – one for each subpopulation. Regardless of the circumstances, the ultimate recommendation strength must be specified using one of the grades presented in table VI.

The strong correlation between the SIGN guidelines and requirements of the public payer (the guidelines translate directly to the conditions of provision of benefits) is one of the main factors that determine their high applicability and treatment effects. That said, it is also possibly the main limitation to the guidelines; by assessing the impact of the recommendations on the budget, they may marginalise the significance of the intervention for the target population and their impact on the patient’s quality of life [20].

Discussion

As presented above, the common characteristics of all grading systems include:

- transparency of determinants influencing the strength of recommendation,

Table VI. The SIGN grading for recommendations

The levels of evidence	
1 ⁺⁺	high-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias
1 ⁺	well-conducted meta-analyses, systematic reviews or RCTs with a low risk of bias
1 ⁻	meta-analyses, systematic reviews, or RCTs with a high risk of bias
2 ⁺⁺	high-quality systematic reviews of case-control studies or cohort studies. high-quality case-control studies or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal
2 ⁺	well-conducted case-control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal
2 ⁻	case-control or cohort studies with high risk of confounding or bias and a significant risk that the relationship is not causal
3	non-analytic studies, e.g. case reports, case series
4	expert opinion

Forms of recommendation	
Judgement	Recommendation
undesirable consequences clearly outweigh desirable consequences	strong recommendation against
undesirable consequences probably outweigh desirable consequences	conditional recommendation against
balance between desirable and undesirable consequences is closely balanced or uncertain	research is recommended and, possibly, a conditional recommendation for use in clinical studies
desired consequences probably outweigh undesirable consequences	conditional recommendation for
desirable consequences clearly outweigh undesirable consequences	strong recommendation for

Good-practice points
recommended best-practice based on the clinical experience of the guideline development group

Source: SIGN (2019). SIGN 50: a guideline developer’s handbook

- following the EBM principles,
- being informed by a thorough review of evidence.

Thus, the basic aspects influencing the strength of recommendation are the quality of body of evidence, consistency of results, the type and magnitude of the effects, as well as the level of authors' certainty towards the effect. The approaches vary in terms of the recognition of the authors' support for recommendations and the conditions of the local healthcare system, which is of particular significance in the case of guidelines developed by state agencies. Table VII presents a summary of the factors that can determine the strength of recommendation, as considered in the analysed grading systems.

To add to the confusion, the same designations used in the grading systems may refer to different aspects – for instance, in some guidelines, letters of the alphabet refer to the quality of evidence, in others – to the strength of the recommendation. Additionally, the grading systems also differ in the number of grades in a scale, e.g. quality of evidence scale may vary – from four grades (according to GRADE) to eight (according to SIGN). Regardless of the above, the highest grade is commonly assigned to meta-analyses and RCTs, which constitute the highest methodological standard for original studies. Discrepancies between the recommendations formulated by different groups and the assigned strength of recommendation are related to additional factors and use of varying scales to grade the quality of evidence. As a result, on the basis of the same study, different authors may develop recommendations which will vary in terms of strength.

Conclusions

The role of CPGs is to verify the quality of the most recent scientific evidence and to assess the benefits and harms of a given

diagnostic and treatment process (the strengths of individual interventions), instead of imposing a universal approach to patient care, like clinical care standards would.

However, the increasing number of guidelines published impedes proper interpretation and comparison of these documents. The number of available grading systems requires the user to have a thorough knowledge of the methods of their development and factors influencing the recommendations to ensure informed decisions on his(her) part.

The diversity of systems used to grade the strength of recommendation and describe the quality of evidence makes it difficult to compare recommendations developed by different authors – even within one branch of medicine. At the same time, guideline users may have doubts as to the meaning of the applied grades of strength of recommendation. They may believe that the assigned grading refers to the significance of the recommendation, and not to the certainty of the evidence informing it.

To tackle these uncertainties and create a transparent and unambiguous set of recommendations within each healthcare system, it would be advisable to introduce a unified grading system for all the guidelines across each branch of medicine. This would account for the key factors in the target healthcare settings.

Acknowledgements

The authors are grateful to Małgorzata Bała and Wiktoria Leśniak (Jagiellonian University Collegium Medicum) as well as Piotr Szymański (Central Clinical Hospital of the Ministry of Interior and Administration in Warsaw) for critically reading the manuscript. We thank Edyta Florka, Katarzyna Just, Magdalena Kutkowska and Katarzyna Sejbuk-Rozbicka (AOTMiT) for their assistance in collecting data.

Table VII. Factors considered in determination of strength of recommendation by various groups developing guidelines

Factors that determine the strength of recommendation		Organisation				
		GRADE	NCCN	NICE	ESMO	SIGN
quality of body of evidence	study design	+	+	+	+	+
	number of studies / sample size	–	+	–	+	–
	quality of evidence	+	risk of bias	+	risk of bias	+
	consistency of the results	+	+	+	+	+
assessment of effectiveness of the intervention		+	+	+	+	+
balance of benefits and harms / positive or negative effect of the intervention		+	+	+	+	+
degree of the group support for the recommendation		–	+	–	–	–
unanimous consensus of the guideline development group		+	–	+	+	+
level of certainty of the authors of a positive impact of the intervention		+	+	+	+	+
cost effectiveness		+	–	+	–	+
conditions of the target healthcare system		+	–	+	–	+
patients' acceptance of the intervention		+	–	+	–	+

Conflict of interest: none declared

Anna Rychert

Agency for Health Technology Assessment and Tariff System

ul. Przeskok 2

00-032 Warszawa, Poland

e-mail: a.rychert@aotmit.gov.pl

Received: 8 Oct 2021

Accepted: 2 Nov 2021

References

1. Graham R, Mancher M, Wolman DM, et al. (ed.). Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US) 2011. <https://www.ncbi.nlm.nih.gov/books/NBK209539/> (2.12.2020).
2. Leśniak W, Bała M, Jaeschke R, et al. Od danych naukowych do praktycznych zaleceń – tworzenie wytycznych według metodologii GRADE. Polish Archives of Internal Medicine. 2015; 125(Special Issue): 26–41, doi: 10.20452/pamw.3232.
3. Jaeschke R, Guyatt G, Cook C. POWAP – czyli praktyka medyczna oparta na wiarygodnych i aktualnych publikacjach. Ewaluacja Badań Medycznych – Evidence Based Medicine (EBM). Medycyna Praktyczna 1999.
4. Koperny M, Maciorowska K, Leśniak W, et al. Clinical guidelines development process in Poland. Przegl Epidemiol. 2017; 71(4): 647–659, indexed in Pubmed: 29417792.
5. Gronseth GS, Cox J, Gloss D et al. Clinical Practice Guideline Process Manual. 2017 Edition. https://tools.aan.com/siteassets/home-page/policy-and-guidelines/guidelines/about-guidelines/17guidelineproc-man_pg.pdf (2.12.2020).
6. Szczekliak A, Gajewski P. Interna Szczeklika–Podręcznik chorób wewnętrznych. Medycyna Praktyczna, Kraków 2014: 2575–2600.
7. Brouwers MC, Kho ME, Browman GP, et al. AGREE Next Steps Consortium, AGREE Next Steps Consortium, AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010; 182(18): E839–E842, doi: 10.1503/cmaj.090449, indexed in Pubmed: 20603348.
8. Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction–GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64(4): 383–394, doi: 10.1016/j.jclinepi.2010.04.026, indexed in Pubmed: 21195583.
9. Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2013; 66(2): 151–157, doi: 10.1016/j.jclinepi.2012.01.006, indexed in Pubmed: 22542023.
10. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation–determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013; 66(7): 726–735, doi: 10.1016/j.jclinepi.2013.02.003, indexed in Pubmed: 23570745.
11. Gopalakrishna G, Mustafa RA, Davenport C, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) to diagnostic tests was challenging but doable. J Clin Epidemiol. 2014; 67(7): 760–768, doi: 10.1016/j.jclinepi.2014.01.006, indexed in Pubmed: 24725643.
12. Hartling L, Fernandes RM, Seida J, et al. From the trenches: a cross-sectional study applying the GRADE tool in systematic reviews of healthcare interventions. PLoS One. 2012; 7(4): e34697, doi: 10.1371/journal.pone.0034697, indexed in Pubmed: 22496843.
13. The National Comprehensive Cancer Network. (b.d.). About the NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). <https://www.nccn.org/professionals/default.aspx> (2.12.2020).
14. The National Comprehensive Cancer Network. (b.d.). Development and Update of the NCCN Guidelines®. <https://www.nccn.org/professionals/development.aspx> (2.12.2020).
15. Wayant C, Cooper C, Turner D, et al. Evaluation of the NCCN guidelines using the RIGHT Statement and AGREE-II instrument: a cross-sectional review. BMJ Evid Based Med. 2019; 24(6): 219–226, doi: 10.1136/bmjebm-2018-111153, indexed in Pubmed: 30904829.
16. The National Institute for Health and Clinical Excellence. Developing NICE guidelines: the manual, 2018. <https://www.nice.org.uk/process/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869> (2.12.2020).
17. European Society for Medical Oncology (ESMO). SOPs/Instructions for Authors and templated for standard ESMO Clinical Practice Guidelines (CPGs) and ESMO-MCBS Scores, 2020. <https://www.esmo.org/content/download/77789/1426712/file/ESMO-Clinical-Practice-Guidelines-Standard-Operating-Procedures.pdf> (31.11.2020).
18. Pentheroudakis G, Cardoso F, Arnold D, et al. ESMO Guidelines Committee. The ESMO guideline strategy: an identity statement and reflections on improvement. Ann Oncol. 2015; 26 Suppl 5: v1–v7, doi: 10.1093/annonc/mdv299, indexed in Pubmed: 26314771.
19. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50: a guideline developer's handbook. Edinburgh: SIGN 2019; (SIGN publication no. 50). https://www.sign.ac.uk/assets/sign50_2019.pdf (31.11.2020).
20. McSweeney LA, Wilson JA, Wilkes S, et al. Is Scottish Intercollegiate Guidelines Network guidance for GP management of tonsillitis suitable? A qualitative study. Fam Pract. 2018; 35(5): 633–637, doi: 10.1093/fampra/cmy017, indexed in Pubmed: 29590327.