CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia

Special article

GRADE System: Classification of Quality of Evidence and Strength of Recommendation☆

José Luis Aguayo-Albasini, Benito Flores-Pastor, Víctor Soria-Aledo*

Servicio de Cirugía General, Hospital General Universitario JM Morales Meseguer, Campus de Excelencia Internacional Mare Nostrum, Universidad de Murcia, Murcia, Spain

ARTICLE INFO

Article history:
Received 16 April 2013
Accepted 19 August 2013
Available online 12 March 2014

Keywords:
Evidence quality
Strength of recommendation
GRADE system
Clinical practice guidelines

ABSTRACT

The acquisition and classification of scientific evidence and subsequent formulation of recommendations constitute the basis for the development of clinical practice guidelines. There are several systems for the classification of evidence and strength of recommendations; the most commonly used nowadays is the Grading of Recommendations, Assessment, Development and Evaluation system (GRADE). The GRADE system initially classifies the evidence into high or low, coming from experimental or observational studies; subsequently and following a series of considerations, the evidence is classified into high, moderate, low or very low. The strength of recommendations is based not only on the quality of the evidence, but also on a series of factors such as the risk/benefit balance, values and preferences of the patients and professionals, and the use of resources or costs.

Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación

Palabras clave:
Calidad de la evidencia
Fuerza de la recomendación
Sistema GRADE
Guías de práctica clínica

RESUMEN

La adquisición y jerarquización de la evidencia, así como la posterior formulación de recomendaciones, constituyen la base del desarrollo de las guías de práctica clínica. Sistemas de graduación de la calidad de la evidencia y de la fuerza de las recomendaciones han existido muchos y actualmente se va imponiendo el modelo *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE). En el sistema GRADE la calidad de la evidencia se clasifica, inicialmente, en alta o baja, según provenga de estudios experimentales u observacionales; posteriormente, según una serie de consideraciones, la evidencia queda en alta, moderada, baja y muy baja. La fuerza de las recomendaciones se apoya no solo en la calidad de la evidencia, sino en una serie de factores como son el balance entre riesgos y beneficios, los valores y preferencias de pacientes y profesionales, y el consumo de recursos o costes.

☆ Please cite this article as: Aguayo-Albasini JL, Flores-Pastor B, Soria-Aledo V. Sistema GRADE: clasificación de la calidad de la evidencia y graduación de la fuerza de la recomendación. Cir Esp. 2014;92:82–88.

* *Corresponding author.*
E-mail addresses: florespastorbenito2@gmail.com, benitom.flores@carm.es (B. Flores-Pastor).

2173-5077/$ – see front matter

Introduction

Evidence-based medicine (EBM) requires medical practitioners to combine their medical knowledge and judgement with the best existing scientific knowledge. Determining the best evidence requires skills of identification, critical analysis and prioritising published evidence. The former stage is essential, as any recommendation or grade of recommendation proposed in terms of preventive or therapeutic surgery or concerning a diagnostic procedure must be directly related to the quality (and other factors) of the existing evidence.

EBM is chiefly of interest to groups of experts who develop clinical practice guidelines (CPG) for research on a disease or health problem and for diagnosis, treatment and prevention. Up to 8 stages are described in the development of a guideline (Table 1), but only stages 3–8 concern us in this article (formulating questions, acquiring evidence, assigning quality and drawing up recommendations). Obtaining useful CPG is not an easy task due to the varied nature of the individuals making up the groups or committees of experts who create these guidelines, their different points of view and methods, and the similar variability of scientific information available on a particular topic.[1–3] Until a few years ago these groups of experts used an informal methodology to reach a consensus, but recently procedures for prioritising evidence and establishing appropriate recommendations have improved. Here the system for the *Grading of Recommendations, Assessment, Development and Evaluation* (GRADE) comes into play.

There are a great many sophisticated systems for categorising scientific evidence, including the English model, the *Oxford Centre for Evidence-Based Medicine* (OCEBM), the *Scottish Intercollegiate Guidelines Network* (SIGN) or the *American College of Chest Physicians* (ACCP) used by the ACCP itself in their guidelines on venous thrombosis up until their seventh revision.[4–6] All of them attributed different quality levels to studies on a particular problem, which then enabled different degrees of recommendation to be made. However, some disadvantages soon emerged, such as the fact that these systems were developed principally as a result of a consensus of expert opinion and were not validated.[7] Therefore, occasionally, different systems were not categorising the same studies in terms of similar levels of evidence. Indeed, sometimes no agreement was reached on the same model. Moreover, some systems were better at estimating the quality of evidence than establishing the grade of recommendation, and vice versa. All the above-mentioned meant that occasionally the CPG were not completely reliable.

The GRADE working group's proposal was communicated in 2004. It was created by an international and multidisciplinary group of methodologists, experts in CPG and clinical doctors, in an attempt to deal with the problems mentioned above.[8,9] The advantage of the system is that it is a thorough and transparent method for classifying quality of evidence and for allocating a grade or strength of recommendation. We shall develop these points as the GRADE system does, but first we shall outline the steps to be followed in the formulation of clinical questions.

Table 1 – Stages in the Preparation of a Clinical Practice Guideline.

1. Definition of scope and objectives
2. Creation of CPG preparation group
3. Formulation of the clinical questions (PICO)
4. Search for evidence
5. Assessment and synthesis of literature
6. Formulation of recommendations
7. External review
8. Edition

CPG: clinical practice guideline; PICO.

Formulation of Clinical Questions in PICO Format and Search for Answers

Once the scope of a CPG has been established, a series of clinical questions need to be defined which are grouped into sections of organisation, prevention, diagnosis, treatment, prognosis, etc. PICO (acronym for Patient–Intervention–Comparison–Outcome) is the preferred method used to move from a generic clinical question to a specifically formulated one to facilitate a bibliographic search and preparation of recommendations for each question. Thus:

a. Patient: or population, disease statuses, age groups, comorbidities, etc.
b. Intervention: treatment, diagnostic test, aetiological agent, etc.
c. Comparison: possible alternative to intervention under research as a regular treatment or placebo, gold reference standard of a diagnostic test, lack of aetiological agent, etc.
d. Outcomes: relevant outcome variables in the case of studies on efficacy, prognosis or aetiology, and validity estimators in the case of diagnostic tests (sensitivity, specificity, probability coefficients, etc.).

When clinical questions are formulated in PICO format they are defined in a specific manner and there is no ambiguity as to what is being probed and moreover, as each type of question corresponds to a type of study with the appropriate design for its answer, the format helps towards conducting a literature search. During the formulation of clinical questions all the possible outcome variables must be defined. This is an even more relevant issue when used in preparing the GRADE system recommendations, where the variables are qualified as to their importance for clinicians and patients and are weighted on a scale from 1 to 9. Only variables with a score from 7 to 9 are considered key in affecting a GRADE system decision and the clinical questions need to be specified to these key variables. The answers to these questions on key outcomes shall be those which are used to grade the recommendations. Variables with a score of 4–6 are classified as important but not crucial for decision making. Those given a score from 1 to 3 shall be considered unimportant and will not be included in the evaluation or influence the recommendations. The strict and accurate selection of key outcome variables means that the studies are selected equally and thus it is possible that the