

Rating the Levels of Evidence in Sports–Medicine Research

Jennifer M. Medina, PhD, ATC, CSCS; Patrick O. McKeon, MS, ATC, CSCS; and Jay Hertel, PhD, ATC • University of Virginia

AS DISCUSSED in this column in the last issue of *Athletic Therapy Today*,¹ it is important for clinicians to recognize the generalization of evidence-based research from journals, how to incorporate it into clinical practice, properly identify the type of study, and understand where the type of study fits into the hierarchy of study design. The next step in evidence-based sports medicine is to determine the quality of evidence each study design provides. In order to do this, clinicians must have an appreciation for the levels of evidence in clinical research.

What Is Evidence and Why Should It Be Graded?

Previously¹ we reviewed a variety of sources that provide clinical evidence. These sources can range from unpublished clinical observations to quantitative reviews of well-designed randomized controlled trials (RCTs). Clinicians must know that not all evidence is equal. There are various methods introduced to determine the quality of evidence found in research. We present two of the most commonly used methods: the Oxford Centre of Evidence-Based Medicine's levels of evidence and grades of recommendation² and the American Academy of Family Physicians' strength of recommendation taxonomy (SORT).³ Each system classifies evidence based on quality of study design, quantity of evidence, and the consistency of results across studies. For example, a systematic review of multiple double-blinded RCTs with consistent results across studies would be rated as a much higher level

of evidence than a clinical case report or a case-control study.

Levels of Evidence and Grades for Recommendation

The Oxford Centre for Evidence-Based Medicine (CEBM) has separated the types of studies into categories designed to direct clinicians in answering clinically relevant questions (see Table 1). Knowledge of the level of evidence for each type of study helps clinicians determine the value of the results reported. The levels of evidence range from 1 to 5; Level 1 is the highest quality and Level 5 is the lowest (see Figure 1).

Level 1 evidence represents the best and most unbiased information available regarding the five types of clinical studies. It is composed of well-designed RCTs or high-quality prospective, diagnostic, or well-designed cost-analysis studies. Level 2 evidence arises from nonrandomized cohort studies. Cohort studies, because they lack randomization, represent a lower quality

TABLE 1. TYPES OF STUDIES USED IN EVIDENCE-BASED MEDICINE

Type of Study	Used to . . .
Therapeutic	Investigate the result of a treatment
Prognostic	Report the effect of a patient characteristic on the outcome of disease
Diagnostic	Assess diagnostic tests
Economic	Develop an economic or decision model

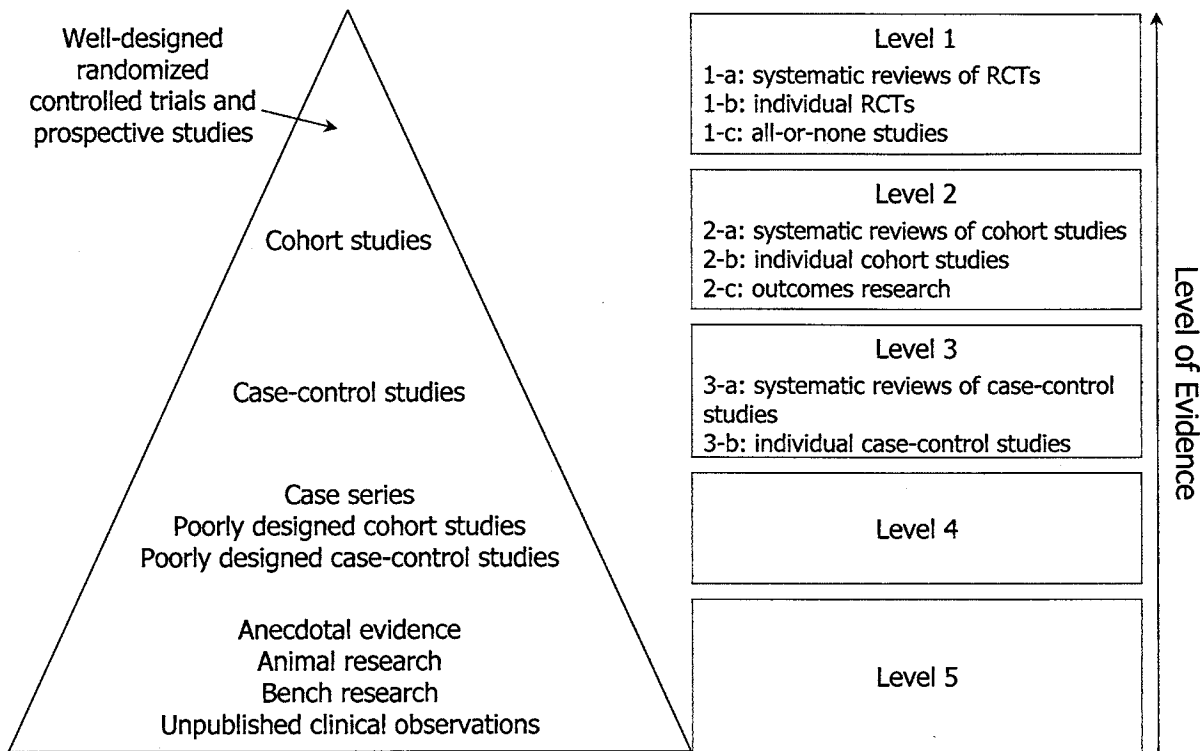


Figure 1 Levels of clinical evidence and study design. Adapted from the Oxford Centre of Evidence-Based Medicine.

of study design than RCTs. At the same time, poorly controlled RCTs are also considered to be Level 2 evidence. Level 3 evidence is attained from retrospective case-control studies. Level 4 evidence represents data obtained from case-series studies or from poor-quality nonrandomized cohort studies or case-control studies. The lowest level of clinical evidence is obtained from data based on anecdotal evidence or expert opinion without explicit testing in a clinical research study. It might also be based on the physiology of how things are theorized to work or on evidence from animal studies and other bench research.

Each level of evidence can be further subdivided into categories related to the assessment of the quality, quantity, and consistency of the evidence. For Levels 1 and 2, there are three subcategories of evidence, a, b, and c. Subcategory a evidence is gained from meta-analyses or systematic reviews of multiple studies that demonstrate consistent clinical findings. Because meta-analyses and systematic reviews are reports of all the results collected after an exhaustive search of the literature, they provide the best evidence to answer a clinical research question. Subcategory b evidence is associated with individual studies within each level. The results must be statistically significant or nonsignificant with little variation, noted as small standard deviations and narrow confidence intervals around the

results. This implies that the results are valid. For Level 1, Subcategory c represents the grading for all-or-none studies, which are rarely used in sports medicine. (All-or-none studies demonstrate some degree of efficacy for an intervention. All: Before a particular treatment was accessible all of the patients would die; now some would survive. None: Before a treatment was accessible some patients would die; now with accessible treatment, none die.) Level 2-c represents the grading for outcomes studies, which are commonly used in sports medicine. Outcomes studies were further explained in the July column.¹

Along with the levels of evidence, the CEBM has developed a systematic method for grading the recommendation for evidence to be used in clinical practice, in other words, how well the evidence answers the clinical question of interest. The recommendation can be graded as A, B, C, D, or I based on the strength of the recommendation and the level of evidence. Evidence is given a grade of A when it is obtained from Level 1 evidence with consistent results. A strong recommendation can be made for or against a particular treatment based a grade of A. Grade B recommendations represent a fair level of confidence for making a recommendation. This grade is based on consistent Level 2 or 3 evidence or inconsistent Level 1 evidence that shows promise. Grade C recommendations arise

from evidence that is conflicting or that comes from Level 4 evidence. It is difficult to make a recommendation for or against a particular treatment when the source of the evidence only allows a grade of C. A grade of D or I (for insufficient) indicates that there is insufficient evidence to make a recommendation. The grading system indicates how confident clinicians can be when making a clinical recommendation.

Strength of Recommendation Taxonomy

In 2004, a task force from the American Academy of Family Physicians developed the strength of recommendation taxonomy (SORT) to enhance the classification and grading of medical evidence. Similar to the CEBM, the SORT can be used to evaluate studies that focus on diagnosis, treatment/prevention/screening, or prognosis. The SORT was designed specifically to separate patient-oriented evidence from disease-oriented evidence. Rather than using five levels of evidence, the SORT uses three. Levels 1 and 2 pertain to high- and low-quality patient-oriented evidence, respectively. This is based on the hierarchy of research study design as discussed in the previous article.¹ Level 3 evidence consists of other evidence such as bench research, disease-oriented outcomes, and anecdotal or expert opinion. (see Figure 2).

In the SORT, there are no subcategories for each level based on the quality of the research design. Rather, levels are determined by the consistency or inconsistency of the results from similar studies. The strength of the recommendation in the SORT is graded as A, B, or C. A grade of A represents a strong recommendation, whereas B is a moderate and C is a weak recommendation.

Rating the Level of Evidence in Practice

The most important consideration when implementing evidence into practice is how effective the evidence is in answering a clinical question of interest. The level of evidence and grade of recommendation are crucial in this decision-making process (see Figure 3). For example, does home-based rehabilitation improve outcomes

after anterior cruciate ligament (ACL) reconstruction over standard physical therapy? In 2005, Grant et al.⁴ conducted an RCT in which 145 patients who had undergone ACL reconstruction were randomized into two groups: standard physical therapy and home-based therapy. At the 3-month follow-up, a significantly greater number of patients in the home-based physical therapy group had regained acceptable knee-extension and -flexion range of motion compared with the standard physical therapy-based group.⁴ The authors concluded that the home-based program was more effective than the standard physical therapy-based program in restoring acceptable knee range of motion in the first 3 months after ACL reconstruction.⁴

The methods of this RCT were very well designed and the patient-oriented outcome measures very conclusive. The authors reported that this study was Level 1 evidence based on the CEBM levels of evidence. The study would be more completely rated as a Level 1-b because it is a well-designed, individual RCT. It can be rated as a Level 1 on the SORT because it provides

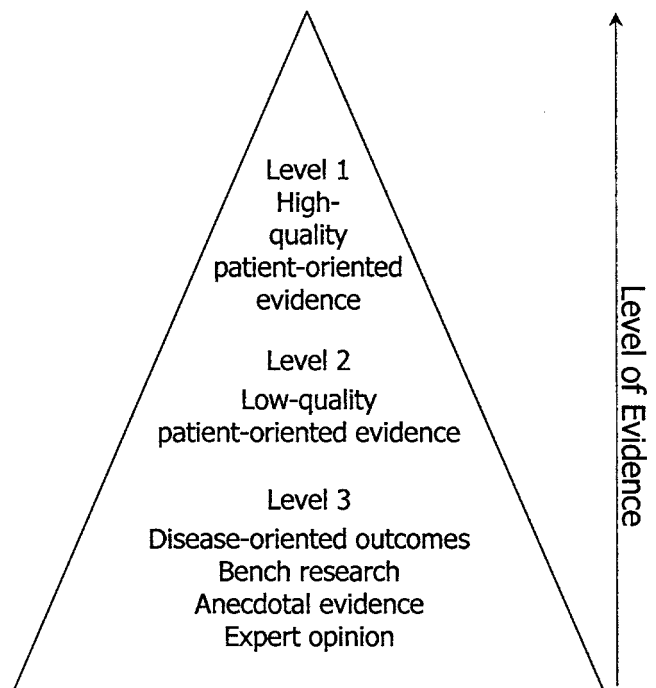


Figure 2 Strength of recommendation taxonomy (SORT) levels of evidence. Adapted from the American Academy of Family Physicians.

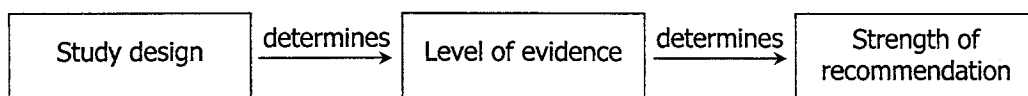


Figure 3 Determinants for grading clinical evidence.

patient-oriented evidence from a well-designed RCT. In answering the clinical question, this evidence would receive a Grade B recommendation on both the CEBM and SORT scales because of the novel study methods. In order to increase the grade of recommendation to A, further independent research is needed to confirm these results.

Another question would be, are there deficits in function in individuals who report having functional ankle instability? Docherty et al.⁵ designed a case-control study to compare individuals with self-reported functional ankle instability with healthy subjects on the performance of four functional tasks. The results revealed that individuals with functional ankle instability displayed significant deficits on some of the functional tasks compared with controls. Because of the study design, the evidence was rated as Level 3-b on the CEBM and a Level 2 on the SORT. In answering the clinical question, this study would receive a Grade B recommendation because it has demonstrated patient-oriented evidence consistent with results from other Level 2 or 3 studies. To upgrade the level of evidence and grade of recommendation, this research question would need to be studied from a prospective perspective and to be confirmed by multiple independent studies.

Conclusion

It is important for clinicians to recognize that all evidence is not created equal. There are several systems for rating the levels of evidence and grading clinical recommendations. We have presented the structure of the two most commonly used rating systems in

sports-medicine research. Increasing numbers of peer-reviewed journals are reporting the level of evidence and grade of recommendation for clinical research to enhance the process of incorporating evidence-based medicine into clinical practice. Most important, clinicians should be able to assess the evidence based on its quality, quantity, and consistency in order to make effective evidence-based clinical decisions. ■

References

1. McKeon PO, Medina JM, Hertel J. Hierarchy of research design in evidence-based sports medicine. *Athl Ther Today*. 2006;11(4):42-45.
2. Centre for Evidence-Based Medicine, Oxford University. Levels of evidence and grades of recommendation. Available at: http://www.cebm.net/levels_of_evidence.asp. Accessed May 20, 2006.
3. Ebell MH, Siwek J, Weiss BD, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *Am Fam Physician*. 2004;69:549-557.
4. Grant JA, Mohtadi NG, Maitland ME, Zernicke RF. Comparison of home versus physical therapy-supervised rehabilitation programs after anterior cruciate ligament reconstruction: a randomized clinical trial. *Am J Sports Med*. 2005;33(9):1288-1297.
5. Docherty CL, Arnold BL, Gansneder BM, Hurwitz S, Gieck J. Functional-performance deficits in volunteers with functional ankle instability. *J Athl Train*. 2005;40(1):30-34.

Jennifer Medina is a doctoral student at Pennsylvania State University and adjunct faculty member at the University of Virginia. Her research interests include lower extremity alignment and the neuromechanics of gait, evidence-based medicine, and management strategies for posttraumatic osteoarthritis.

Patrick McKeon is currently a doctoral student in the sports-medicine program at the University of Virginia. His research interests include evidence-based practice, the dynamical-systems model of motor control, and sensorimotor deficits associated with ankle instability.

Jay Hertel is an assistant professor in the graduate athletic training and sports-medicine programs at the University of Virginia. His research interests include ankle instability, sensorimotor deficits associated with lower extremity injuries, and evidence-based practice.