Basic Athletic Training
Course Pack A

Clinical Examination and Diagnosis

SECTION II

CHAPTER

5

Evidence-Based Health Care

STUDENT OUTCOMES

1. Explain how to practice evidence-based health care.
2. Describe methods to search for health care–related evidence.
3. Recognize different methods to evaluate health care evidence.
4. Identify values used to assess the accuracy of diagnostic tests.
5. Identify values used to assess the effectiveness of clinical treatments.

INTRODUCTION

Evidence-based health care (EBHC) is part of a cultural shift in current health care practice that aims to provide patients with the best quality care by integrating the best evidence with clinical expertise and the individual patient's values and circumstances.[1] One reason for the need for EBHC is the volume of new information available to clinicians each day. Traditional sources of information such as textbooks and expert opinion are quickly outdated given the volume of new information available on a daily basis (>760,000 new biomedical articles were added to PubMed, a biomedical database, in 2014 alone).[2]

TYPES OF EVIDENCE

An athletic trainer has a patient who recently tore her anterior cruciate ligament (ACL). The patient is asking what type of treatment would be best. What type of outcome evidence should the clinician use to provide treatment recommendations?

As health care providers try to determine the best quality care for patients, it is important to consider patient-oriented evidence as much, if not more, than disease-oriented evidence. Patient-oriented evidence or **patient-oriented evidence that matters (POEM)** provides information on areas about which patients would be most concerned (e.g., mobility, mortality, symptom improvement, health care cost, and quality of life). Typically, patient-oriented evidence provides a more holistic view of a patient's health status. **Disease-oriented evidence** is physiological information such as blood pressure and joint range of motion measures, or symptoms such as headache and nausea. Because disease-oriented evidence has traditionally been gathered by clinicians, it may also be referred to as **clinician-oriented evidence**. Changes in disease-oriented evidence may or may not have a significant impact on a patient's health status; therefore, its use when determining the best quality care

for patients may be limited. Patient-oriented outcome measures are common forms of patient-oriented evidence. **Patient-oriented outcome measures** usually are self-reported questionnaires that patients complete throughout treatment to assess their quality of life. Generic patient-based outcome measures assess quality of life from a broad prospective, whereas specific disease or anatomical patient-based outcome measures assess quality of life from a narrower point of view. Examples of outcome measures are given in Box 5.1.

The athletic trainer should seek out POEM to provide the best treatment advice for the patient with an ACL tear. Patient-oriented outcome measures would provide a broad quality of life assessment. Examples of patient-oriented evidence that could be used in this case would include Short Form 12 (SF-12) health survey questionnaire or the International Knee Document Committee (IKDC) Subjective Knee Form.

BOX 5.1 Examples of Patient-Based Outcome Measures

Generic Outcome Measures

INSTRUMENT	AVAILABLE AT
Quality of Well-Being (QWB)	https://hoap.ucsd.edu/qwb-info/
Short Form 36 Health Survey Questionnaire (SF-36)	https://www.optum.com/optum-outcomes/what-we-do.html
Short Form 12 Health Survey Questionnaire (SF-12)	https://www.optum.com/optum-outcomes/what-we-do.html

Disease-Specific Outcome Measures

INSTRUMENT	AVAILABLE AT
Hip Disability and Osteoarthritis Outcome Score (HOOS)	http://www.koos.nu/
Rheumatoid and Arthritis Outcome Score	http://www.koos.nu/

Anatomy-Specific Outcome Measures

INSTRUMENT	AVAILABLE AT
Foot and Ankle Outcome Score (FAOS)	http://www.koos.nu/
Disabilities of the Arm, Shoulder and Hand (DASH) Questionnaire	http://www.orthopaedicscore.com/scorepages/disabilities_of_arm_shoulder_hand_score_dash.html
International Knee Document Committee (IKDC) Subjective Knee Form	http://www.sportsmed.org/Research/IKDC_Forms/
Michigan Hand Outcome Questionnaire	http://www.orthopaedicscore.com/scorepages/Michigan_Hand_Outcome_Questionnaire.html
Oswestry Low Back Pain Score	http://www.orthopaedicscore.com/scorepages/oswestry_low_back_pain.html
Oxford Hip Score	http://www.orthopaedicscore.com/scorepages/oxford_hip_score.html

HOW TO PRACTICE EVIDENCE-BASED HEALTH CARE

? An athletic trainer has a patient who injured her knee. Which diagnostic tests would provide the athletic trainer with the most accurate diagnosis? What is the best treatment for this patient's condition? How can the athletic trainer quickly answer these questions?

There are five steps athletic trainers need to follow to practice EBHC: (1) Develop a clinical question; (2) search for the best evidence; (3) evaluate the evidence for validity, impact, and applicability; (4) integrate the evidence into the clinical decision; and (5) evaluate the efficiency and effectiveness of steps 1 to 4.

Developing a Clinical Question

Clinicians commonly use the **PICO** or **PICO/T** formats to develop their clinical questions for more effective evidence searches. PICO/T is an acronym:

- **P = Patient/Problem**
- **I = Intervention/Variable of Interest**

- **C = Comparison**
- **O = Outcome**
- **T = Time**

Comparison and Time are not always included in the clinical question; therefore, those terms are optional. Templates for and examples of different types of clinical questions are shown in **Box 5.2**.

BOX 5.2 PICO or PICO/T Clinical Question Templates

For an Intervention or Therapy

- In ___(P)___, what is the effect of ___(I) on ___(O) compared with ___(C) within ___(T)___?

  Example: In patients with ankle sprains, what is the effect of cryotherapy on pain compared with electrical stimulation treatment?

For Etiology

- Are ___(P)___ who have ___(I)___ at [increased/decreased] risk for/of ___(O)___ compared with ___(P)___ with/without ___(C)___ over___(T)___?

  Example: Are football players who have a family history of cardiovascular conditions at an increased risk for a cardiovascular condition compared with football players without a family history of cardiovascular conditions over a football season?

Diagnosis or Diagnostic Test

- Are/Is ___(I)___ more accurate in diagnosing ___(P)___ compared with ___(C)___ for ___(O)___?

  Example: Is the Lachman's test more accurate in diagnosing patients with knee injuries compared with the anterior drawer test for ACL tears?

Prevention

- For ___(P)___, does the use of ___(I)___ reduce the future risk of ___(O)___ compared with ___(C)___?

  Example: For soccer players, does the use of ankle braces reduce the future risk of ankle sprains compared with ankle taping?

Prognosis/Predictions

- Does ___(I)___ influence ___(O)___ in patients who have ___(P)___ over ___(T)___?

  Example: Does the type of ACL graft used influence knee stability in patients who have ACL replacement surgery over 20 years?

Adapted from Melnyk B, Fineout-Overholt E. *Evidence-Based Practice in Nursing & Healthcare*. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.

Searching the Literature

To efficiently search the literature, it is important to use appropriate evidence databases, which are collections of organized information. **Box 5.3** provides a list of the most common health care–related databases used by athletic trainers. Some databases contain only **filtered information**. Filtered information means that other clinicians and researchers have already searched the existing evidence, evaluated it, and synthesized a clinical recommendation. Examples of filtered information include clinical practice guidelines, critically appraised topics (CATs), Cochrane reviews, evidence-based synopses, meta-analysis, and systematic reviews (SRs). Other databases contain both filtered information and **unfiltered information**. Unfiltered information includes individual research studies and expert opinions. Busy clinicians will find that filtered information is a more efficient and effective means of quickly assessing information to inform their clinical practice.

BOX 5.3 Relevant Evidence Databases for Athletic Trainers

Databases That Only Contain Filtered Information

Cochrane (http://www.cochrane.org)

Centre for Evidence-Based Physiotherapy: PEDro (http://www.pedro.org.au)

National Guideline Clearing House (http://www.guideline.gov)

TRIP (http://www.tripdatabase.com/)

Databases That Contain Both Filtered and Unfiltered Information

CINAHL (https://www.ebscohost.com/nursing/products/cinahl-databases)

EMBASE (https://www.elsevier.com/solutions/embase-biomedical-research)

PubMed/MEDLINE (http://www.ncbi.nlm.nih.gov/pubmed)

Clinical Queries (http://www.ncbi.nlm.nih.gov/pubmed/clinical)

SPORTDiscus (https://www.ebscohost.com/academic/sportdiscus-with-full-text)

Athletic trainers can improve the effectiveness of their evidence search by using appropriate **search terms**, which are words with precise meaning. If an athletic trainer has used the PICO/T format to develop the clinical question, then the athletic trainer can use the PICO/T terms as the search terms. Some databases, such as PubMed/MEDLINE, Cumulative Index of Nursing and Allied Health Literature (CINAHL), and SPORTDiscus, support the use of **Boolean terms** (e.g., "and," "or," and "not"). In these databases, health care providers can focus their literature search by using the Boolean terms in addition to their search terms (see **Box 5.4**). Using **Medical Subject Heading (MeSH)** terms can also improve the effectiveness of one's search within the PubMed/MEDLINE databases. MeSH terms are the controlled medical

vocabulary used by the U.S. National Library of Medicine in the PubMed/MEDLINE databases (see Box 5.5 for examples).

BOX 5.4 Using Search and Boolean Terms to Streamline Evidence Searches

Example 1

To answer the following clinical question:

> Is the Lachman's test more accurate in diagnosing patients with knee injuries compared with the anterior drawer test for ACL tears?

The following searches were done in the PubMed database (www.ncbi.nlm.gov/pubmed):

Term Used During a Search	Number of Articles Returned in a PubMed Search
Knee Injuries	26,265
ACL	17,419
Lachman	1,725
Anterior Drawer	705
Anterior Drawer **NOT** Ankle	547
Lachman **AND** Knee Injuries	480
Lachman **AND** Anterior Drawer	210
Lachman **AND** Anterior Drawer **AND** ACL	202
Lachman **AND** Anterior Drawer **AND** ACL **AND** Diagnosis	156

Notice how the PICO terms became the search terms and the addition of the Boolean terms "AND" and "NOT" decreased the number of results one would need to evaluate.

Example 2

To answer the following clinical question:

> In patients with ankle sprains, what is the effect of cryotherapy on pain compared with electrical stimulation treatment?

The following searches were done in the PubMed database

(www.ncbi.nlm.gov/pubmed):

Term Used During a Search	Number of Articles Returned in a PubMed Search
Ankle Sprain	13,956
Cryotherapy	24,527
Cryotherapy **OR** Cold Therapy	43,559
Electrical Stimulation	163,553
Pain	597,854
Cryotherapy **AND** Ankle Sprain	67
Cryotherapy **AND** Ankle Sprains **AND** Electrical Stimulation	4
Cryotherapy **AND** Ankle Sprains **AND** Pain	25

Notice how the PICO terms became the search terms and the addition of Boolean term "OR" increased the number of results, whereas the use of "AND" decreased the number of results one would need to evaluate.

BOX 5.5 Examples of MeSH Terms

Search Terms

Ankle sprain

Chronic ankle sprain

Syndesmotic ankle sprain

MRI

Shoulder dislocation

Glenohumeral dislocation

Shoulder labral tear

Shoulder

Labral

Tear

MeSH Term Used by MEDLINE/PubMed

Ankle injuries

Ankle injuries

Ankle injuries

Magnetic resonance imaging

Shoulder dislocation

Shoulder dislocation

Shoulder

Shoulder

No related MeSH term

Tear

Lacerations

Evaluate the Evidence

Once evidence related to the clinical question is obtained, it must be evaluated for impact (level of importance), reliability (how reproducible are the results), validity (do the results really represent what we think they represent), and applicability (how well does the evidence apply to the current clinical question). Larger quantities of strong evidence should have a greater impact on clinical decisions than smaller quantities of weaker evidence. The **hierarchy** or **level of evidence** (Fig. 5.1) can help clinicians decide which pieces of evidence are stronger or weaker than others. Filtered information tends to be stronger than unfiltered information; therefore, it appears higher on the evidence hierarchy. See Box 5.6 for a description of each type of evidence. In addition to the type of evidence, studies that evaluate patient-oriented evidence are more meaningful than those that only report disease-oriented evidence. Finally, treatments that demonstrate a larger effect size are more meaningful. **Effect size** measures the difference in outcomes between the treatment and nontreatment groups. The larger the difference or effect size, the greater the impact the treatment will have.

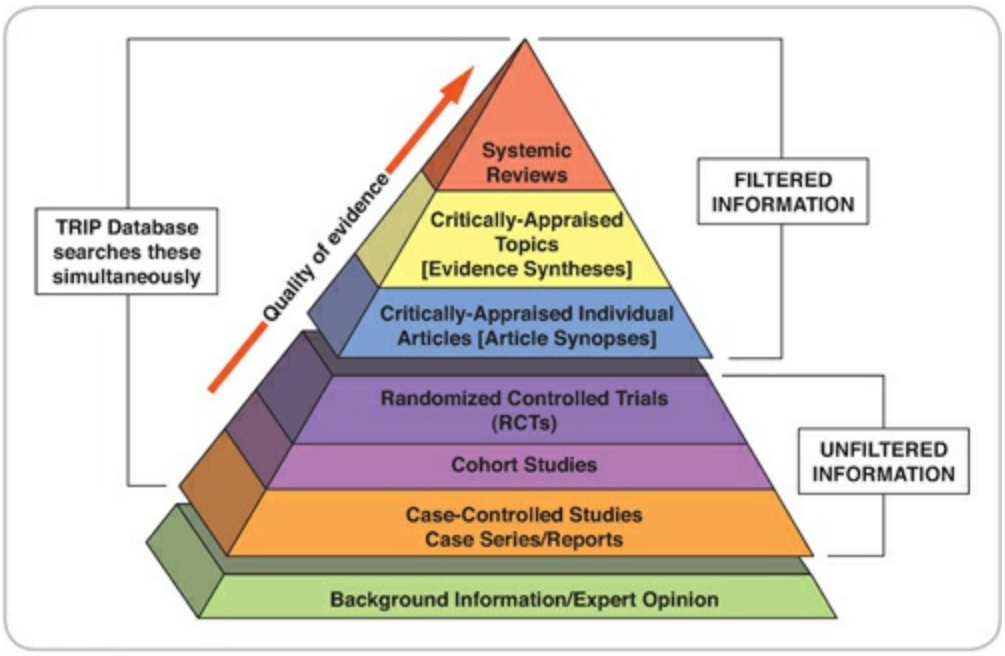

Figure 5.1. Evidence hierarchy.

BOX 5.6 Types of Evidence

Filtered

Meta-analysis	A study that pools results of two or more studies to obtain an overall answer to a question or interest[a]
Systematic review	A methodical review of the existing literature on a clearly described specific question. A description of how the evidence on the topic was found, including the databases, search terms, and inclusion and exclusion criteria, is included in the article.[a]
Critically appraised topic	A short summary of the best available evidence on an individual topic. Unlike in a systematic review, an exhaustive search of the literature was not completed.
Cochrane review	A systematic review created by a global independent network of health practitioners from over 120 countries[b]

Unfiltered

Randomized controlled (clinical) trial	A research study in which a group of patients is randomized into an experimental group and a control group. These groups are followed up for the outcomes of interest.[a]
All-or-none	Case series in which all of the patients experience the same outcome; for example, everyone one who wears an ankle brace does not sustain an ankle sprain.[c]
Cohort study	A research study in which two groups of patients, one that did receive the treatment/injury of interest and one that did not, are followed and an outcome of interested is measured[a]
Case-control	A research study in which two groups of patients, one that did receive the treatment/injury of interest and one that did not, are examined by looking back at existing records for an outcome of interest[a]
Case series	Describes characteristics of between two and five patients with an uncommon disease/injury or who have undergone a similar procedure[a]
Case report	Describes characteristics of a single patient with an uncommon disease/injury or who has undergone a unique treatment[a]
Controlled laboratory study	An in vitro or in vivo investigation in which one group receiving an experimental treatment is compared with one or more groups receiving no treatment or an alternate treatment. Laboratory studies that only include healthy people (no patients) are considered controlled laboratory studies.[a]
Descriptive laboratory study	An in vivo or in vitro study that describes characteristics such as anatomy, physiology, or kinesiology of a broad range of subjects or a specific group of interest[a]
Expert or consensus opinion	An idea that cannot be substantiated by direct evidence. It is a hypothesis based on related information.

[a]Hertel J. Keep it simple: study design nomenclature in research article abstracts [editorial]. *J Athl Train*. 2010;45(3):213–214.

[b]The Cochrane Collaboration. About us. http://www.cochrane.org/about-us. Accessed January 8, 2015.

[c]Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence.' *BMC Med Res Methodol*. 2009;9:34.

Clinicians commonly use one of three scales to efficiently communicate the impact or strength of individual pieces of evidence and overall clinical recommendation. Those scales include the following:

- Strength of Recommendation Taxonomy (SORT) scale, which includes a 3-point, alphabetic scale (e.g., A, B, C) to score clinical recommendations, and a 3-point, numeric scale (e.g., 1, 2, 3) to score individual pieces of evidence (**Box 5.7**).[3]
- Oxford Centre for Evidence-Based Medicine (CEBM) 2011 Levels of Evidence, which uses a 5-point numeric scale to rate evidence, where level 1 is the strongest and level 5 is the weakest.[4]

- Grading of Recommendations Assessment, Development, and Evaluation (GRADE), which can be used to rate SRs and clinical guidelines. Clinicians use the GRADE scale to classify evidence as either "high," "moderate," "low," or "very low." Classifications are based on the quality of evidence, the likelihood of both desirable and undesirable effects, common patient values, and a judicious use of resources.[5]

BOX 5.7 SORT Scale

Clinical Recommendation Subscale

Strength of Recommendation	Definition
A	Recommendation based on consistent and good-quality patient-oriented evidence[a]
B	Recommendation based on inconsistent or limited-quality patient-oriented evidence[a]
C	Recommendation based on consensus, usual practice, opinion, disease-oriented evidence,[a] or case series for studies of diagnosis, treatment, prevention, or screening

Individual Evidence Subscale

Strength of Study/ Evidence	Diagnosis	Treatment/Prevention/ Screening	Prognosis
Level 1— good quality	■ Validated clinical decision rule ■ SR/meta-analysis of high-quality studies ■ High-quality diagnostic cohort study[b]	■ SR/meta-analysis of RCTs with consistent findings ■ High-quality individual RCT[c] ■ All-or-none study[d]	■ SR/meta-analysis of good-quality cohort studies ■ Prospective cohort study with good follow-up
Level 2— limited-quality patient-oriented evidence	■ Unvalidated clinical decision rule ■ SR/meta-analysis of lower quality studies or studies with inconsistent findings ■ Lower quality diagnostic cohort study or diagnostic case-control study[d]	■ SR/meta-analysis of lower quality clinical trials or of studies with inconsistent findings ■ Lower quality clinical trial[c] ■ Cohort study ■ Case-control study	■ SR/meta-analysis of lower quality cohort studies or with inconsistent results ■ Retrospective cohort study or prospective cohort study with poor follow-up ■ Case-control study ■ Case series
Level 3— other evidence	Consensus guidelines, extrapolations from bench research, usual practice, opinion, disease-oriented, evidence (intermediate or physiological outcomes only), or case series for studies of diagnosis, treatment, prevention, or screening		

Consistency Across Studies

Consistent	Most studies found similar or at least coherent conclusions (coherence means that differences are explainable). *or* If high-quality and up-to-date systematic reviews or meta-analyses exist, they support the recommendation.
Inconsistent	Considerable variation among study findings and lack of coherence *or* If high-quality and up-to-date systematic reviews or meta-analyses exist, they do not find consistent evidence in favor of the recommendation.

[a]Patient-oriented evidence measures outcomes that matter to patients: morbidity, mortality, symptom improvement, cost reduction, and quality of life. Disease-oriented evidence measures intermediate, physiological, or surrogate end points that may or may not reflect improvements in patient outcomes (e.g., blood pressure, blood chemistry, physiological function, pathologic findings).

[b]High-quality diagnostic cohort study: cohort design, adequate size, adequate spectrum of patients, blinding, and a consistent, well-defined reference standard.

[c]High-quality RCT: allocation concealed, blinding if possible, intention-to-treat analysis, adequate statistical power, adequate follow-up (greater than 80%).

[d]In an all-or-none study, the treatment causes a dramatic change in outcomes, such as antibiotics for meningitis or surgery for appendicitis, which precludes study in a controlled trial.

RCT, randomized controlled trial; *SR*, systematic review.

From 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(3):548–556.

Athletic trainers can also use **appraisal scales** to assess the design of individual research studies. Appraisal scales are intended for specific types of research. Clinicians and researchers should only use preferred reporting items for systematic reviews and meta-analyses (PRIMAS) to assess meta-analyses and SRs, the Jadad and physiotherapy evidence database (PEDro) scales to evaluate randomized controlled trial (RCT), either QUADAS-2 or the Standards for Reporting Diagnostic Accuracy (STARD) statement to assess diagnostics studies, and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) to evaluate epidemiologic research.

Reliability

The reliability of research results affects the impact evidence should have on clinical decisions. The **reliability** indicates how reproducible the results are when the measurement should be the same (e.g., measuring a patient's ankle dorsiflexion range of motion multiple times without treatment or a change in

injury status should result in the same value). **Intrarater reliability** values determine the consistency of the measurements made by a single researcher or instrument (e.g., the same clinician measures ankle dorsiflexion multiple times), whereas the reliability of measurements made between several researchers or instruments is **interrater reliability** (e.g., clinician A and clinician B both measure a patient's ankle dorsiflexion). Interclass correlation coefficients (ICCs) are common statistical values used to report both intrarater and interrater reliability for continuous data (e.g., range of motion, force, and temperature). Kappa coefficients should be reported for categorical data (e.g., positive and negative diagnostic test results) and weighted kappa coefficients for ordinal data (e.g., manual muscle test or grades of edema).[6] ICC and kappa coefficient values fall on a scale of 0 to 1.0, with values closer to 1.0 indicating greater agreement or reliability within or between researchers and/or instruments. Typically, reliability values greater than 0.7 are acceptable.[7]

Validity

Besides study design and reliability of research results, the validity of results should be considered. **Validity** is the assurance that measurements represent what we think they represent (e.g., differences in weight scale results represent true changes in an individual's body weight, or changes in a patient-oriented outcome scale represent true changes to the patient's health status).

Diagnostic Accuracy

There are several statistics measures that may be used to assess **diagnostic accuracy**, that is, the ability of a diagnostic test/technique to discriminate between disease/injury and health. When assessing a diagnostic test, the results of the test are compared to those of a reference standard. **Reference standards** reflect the patient's true status, that is, injured or healthy. The results of the diagnostic test are compared to the reference standard in a 2 × 2 contingency table, and there are four possibilities: true positive, false positive, false negative, and true negative (see **Box 5.8**). **True positives** indicate individuals who have a positive diagnosis according to the diagnostic test and

really have the injury according to the reference standard. **False positives** indicate individuals who have a positive diagnosis according to the diagnostic test but really do not have the injury according to the reference standard. **False negatives** indicate individuals who have a negative diagnosis according to the diagnostic test but really do have the injury according to the reference standard. **True negatives** indicate individuals who have a negative diagnosis according to the diagnostic test and really do not have the injury according to the reference standard.

BOX 5.8 Determining the Accuracy of Diagnostic Tests

Diagnostic Data Contingency Table[a]

		Reference Standard		
		Dx+	Dx−	Row Total
Thessaly Test (Diagnostic Test)	Dx+	True positives 31 (A)	False positives 24 (B)	55 (A + B)
	Dx−	False negatives 8 (C)	True negatives 17 (D)	25 (C + D)
	Column total	39 (A + C)	41 (B + D)	Grand total 80 (A + B + C + D)

[a]Adapted from Mirzatolooei F, Yekta Z, Bayazidchi M, et al. Validation of the Thessaly test for detecting meniscal tears in anterior cruciate deficient knees. *Knee*. 2010;17(3):221–223.

Formulas for Diagnostic Validity Measures

Term	Formula	Example
Diagnostic accuracy	(A + D) / N	(31 + 17) / 80 = 60%
Sensitivity	A / (A + C)	31 / (31 + 8) = 79%
Specificity	D / (B + D)	17 / (24 + 17) = 41%
False positive rate	B / (B + D) (1 − Specificity)	24 / (24 + 17) = 59% 1 − 0.41 = 0.59 = 59%
False negative rate	C / (A + C) (1 − Sensitivity)	8 / (31 + 8) = 21% 1 − 0.79 = 0.21 = 21%
Positive predictive value (PV+)	A / (A + B)	31 / (31 + 24) = 56%
Negative predictive value (PV−)	D / (C + D)	17 / (8 + 17) = 68%
Prevalence	(A + C) / N	(31 + 8) / 80 = 49%
Positive likelihood ratio (LR+)	Sensitivity / (1 − Specificity)	0.79 / (1 − 0.41) = 1.3
Negative likelihood ratio (LR−)	(1 − Sensitivity) / Specificity	(1 − 0.79) / 0.41 = 0.5

The sensitivity and specificity of a diagnostic test can help clinicians determine which tests to use when there are several options. Sensitivity and specificity can be calculated from the contingency table values and are scaled from 0% to 100%, where 100% is perfect sensitivity or specificity (see **Box 5.8** for data regarding the Thessaly test for detecting meniscal tears in an ACL-deficient knee[8]). **Sensitivity** is the ability of the diagnostic test to detect an injury. When a clinician uses a diagnostic test with a high sensitivity, a negative test helps to rule out the injury. A mnemonic to help remember this relationship is with high sensitivity (*Sn*), a negative (*N*) test rules out (*out*) the injury, or *SnNout*.[1] Using the example for **Box 5.6**, the Thessaly test had a sensitivity of 79%, making it a useful test to rule in a meniscal tear in those patients with an ACL-deficient knee. **Specificity** is the ability of a diagnostic test to detect health. A mnemonic to help remember this relationship is with high specificity (*Sp*), a positive (*P*) test rules in (*in*) the injury, or *SpPin*.[1] Using the example for **Box 5.6**, the Thessaly test had a specificity of 41%, making it less useful at ruling out a meniscal tear in those patients with an ACL-deficient knee. Clinicians will find diagnostic tests with higher sensitivity and/or specificity more accurate and therefore more helpful when making clinical decisions.

The reciprocal of sensitivity and specificity are the false negative and false

positive rates. The **false negative rate** is the *inability* of a diagnostic test to detect injury. The **false positive rate** is the *inability* of a diagnostic test to detect health. Sensitivity and specificity values are reported and used more often than the false negative and false positive rates.

Predictive values can help clinicians determine whether a diagnostic test would be effective as a screening tool and can also be calculated from contingency table results (see Box 5.8). The **positive predictive value (PV+)** estimates how many people who have a positive test actually have the injury. The **negative predictive value (PV−)** estimates how many people who have a negative test are actually healthy. In the example in Box 5.8, the Thessaly test has a PV+ of 56% and PV− of 68%, meaning it would be a better tool for screening for a healthy meniscus than for an injured meniscus during a preseason physical exam.

The sensitivity, specificity, and predictive values for a diagnostic test are all influenced by the prevalence of the injury. **Prevalence** describes how common the injury is, that is, the number of injury cases in a given population. In the example in Box 5.8, 49% of the studied population had meniscal tears. If the prevalence of an injury is high, then the possibility of positive diagnostic test result by chance increases and results in higher sensitivity, specificity, and positive predictive values. If the prevalence of an injury is low, then the possibility of a positive diagnostic test by chance decreases, resulting in lower sensitivity and specificity values and high negative predictive values.

Likelihood Ratios

Likelihood ratios can help clinicians determine how likely it is that a patient does or doesn't have an injury based on the diagnostic test results. A **positive likelihood ratio (LR+)** tells a clinician how much more likely the patient is to have the condition if the diagnostic test is positive. **Negative likelihood ratios (LR−)** tell an athletic trainer how less likely the patient is to have the condition if the diagnostic test is negative. Based on the likelihood ratio of a diagnostic test, providers can view those test results as being unimportant and unhelpful to very important and helpful (Fig. 5.2). Unlike sensitivity, specificity, and prediction values, the prevalence of a condition will not

influence the likelihood ratio results. Based on the scale presented in Figure 5.2 and the Thessaly test, LR+ of 1.3 and LR− of 0.5 (**Box 5.8**) has limited value to a clinician trying to rule in or rule out a meniscal injury in an ACL-deficient knee.

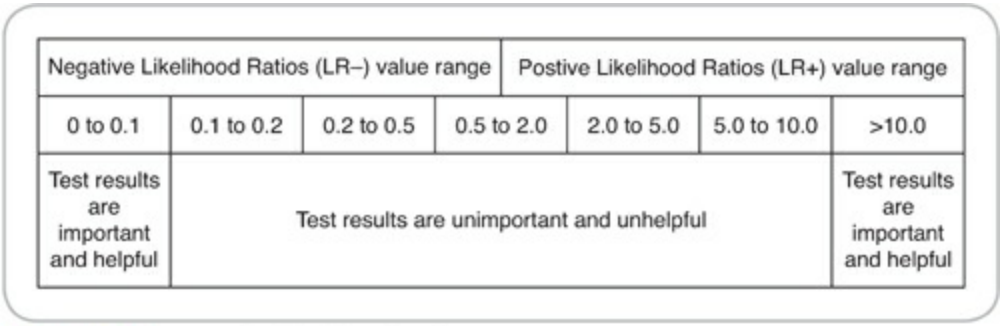

Figure 5.2. Interpreting likelihood ratios.

Probability of a Diagnosis

Clinicians start with a **pretest probability**, which is their hypothesis that a patient has an injury. Clinicians can use their past clinical experience, the patient's medical history, and signs and symptoms, as well as the prevalence of the injury, to estimate the pretest probability that a patient is injured. The clinician can then choose which diagnostic test(s) to use, based on each test's sensitivity and specificity values. The best tests would have demonstrated both high sensitivity and high specificity. Based on the diagnostic test result and the diagnostic test's likelihood ratios, the clinician can determine the **posttest probability** of the patient having the injury on a nomogram (**Fig. 5.3**). Using the Thessaly test example in **Box 5.8**, the pretest probability of a meniscal tear base on prevalence would be 49%. If a patient's Thessaly test was positive and based on the Thessaly test LR+ value of 1.3, the posttest probability of a meniscal tear would be approximately 56%. If a patient's Thessaly test was negative and based on the Thessaly test LR− value of 0.5, the posttest probability of a meniscal tear would be approximately 32%. If several diagnostic tests are used, then the posttest probability becomes the pretest probability for the second test, and so on.

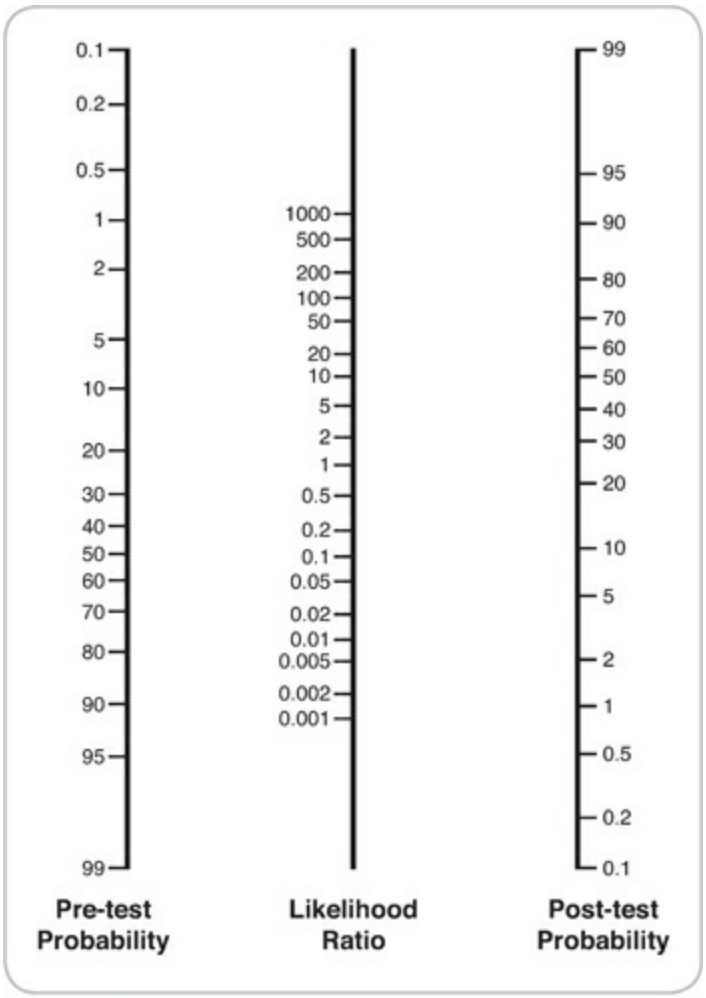

Figure 5.3. Nomogram. (Reprinted with permission from Fagan TJ. Letter to the editor: a nomogram for applying likelihood ratios. *N Engl J Med.* 1975;293:257. Copyright © [1975] Massachusetts Medical Society.)

Risk of Injury

Data from observational studies are used to determine the distribution of injury/disease in a population. The **incidence** of injury is the number of new injuries occurring during a set amount of time, whereas the **incidence rate (IR)** is the number of new cases during a set observation period (IR = number of new cases during / total person-time at risk). The total person-time at risk for sport injury IRs could be the number of practices and/or games in which athletes participated. The **prevalence (P)** is the total number of existing patients with the injury or disease at a given point of time (P = number of existing cases / total population at risk for injury or disease).

Treatment Effectiveness

To assess the effectiveness of a treatment or intervention, it is best to use an RCT, in which patients are randomly assigned to either a treatment or control (i.e., no treatment, or some other comparison) group. The researchers then need to define what a successful treatment outcome would be (e.g., patient returns to participation or patient reports pain reduction on a 5-point Likert scale of global effect). A 2 × 2 contingency table is again used to organize the treatment data (see **Box 5.9** for data regarding a comparison between foot orthoses and inserts on patellofemoral pain[9]). To assess the effectiveness, the athletic trainer looks at the number of unsuccessful or failed treatments. The treatment or **experimental event rate (EER)** is the number of adverse outcomes or unsuccessful treatments. In the example in **Box 5.9**, we can see that foot orthoses failed to reduce pain in 15% of the patients treated with orthoses. The **control event rate (CER)** is the number of adverse or unsuccessful outcomes in the control group. In the example in **Box 5.9**, we can see that 43% of the patients treated with inserts failed to have their pain reduced.

BOX 5.9 Determining the Effectiveness of Treatment

	Outcomes		
	Unsuccessful/Adverse	Success	Row Total
Foot orthoses (treatment group)	6 (A)	35 (B)	41 (A + B)
Flat inserts (control group)	17 (C)	23 (D)	40 (C + D)
Column total	23 (A + C)	58 (B + D)	Grand total 81 (A + B + C + D)

Term	Formula	Example
Experimental event rate (EER)	A / (A + B)	6 / (6 + 35) = 15%
Control event rate (CER)	C / (C + D)	17 / (17 + 23) = 43%
Relative risk (RR)	EER / CER	0.15 / 0.43 = 35%
Relative risk reduction (RRR)	(CER − EER) / CER	(0.43 − 0.15) / 0.43 = 65%
Absolute risk reduction (ARR)	CER − EER	0.43 − 0.15 = 28%
Number needed to treat (NNT)	1 / ARR	1 / 0.28 = 3.6
Absolute risk increase (ARI)	EER − CER	0.15 − 0.43 = −28%
Number needed to harm (NNH)	1 / ARI	1 / −0.28 = −3.6

The EER and CER are used to calculate the relative risk of the treatment. The **relative risk (RR)** tells us how likely an adverse event or unsuccessful treatment will occur in the treatment group relative to the control group. In the example in Box 5.9, patients who received the foot orthoses were only 35% as likely to continue to experience knee pain compared to those treated with flat inserts.

Clinicians may find the relative risk reduction more helpful than the relative risk. The **relative risk reduction (RRR)** indicated the reduction in unsuccessful treatments relative to the control group. For the example presented in Box 5.9, patients who received orthoses were 65% less likely to have pain compared to those who received flat inserts. Because the RRR is a relative value, it does not provide any idea how large the treatment effect is; therefore, clinicians might find the absolute risk reduction more helpful. The **absolute risk reduction (ARR)** provides us with the actual difference in risk between the treatment and control group. For the example in Box 5.9, the ARR is 28%, indicating that orthoses reduced pain 28% more often than the flat inserts.

The **number needed to treat (NNT)** is helpful when determining whether a treatment is a good allocation of time and/or resources. The NNT indicates how many patients would need to receive the therapy to prevent one adverse or unsuccessful outcome. Positive values indicate that a treatment is helpful, and the closer the value is to 1, the more effective the treatment. In the treatment example in Box 5.9, the NNT for foot orthoses is 3.6, which is rounded to 4. This means that for every four patients with patellofemoral pain treated with

orthoses, we should expect one patient to experience pain relief.

To know how good a treatment is at preventing injury or assessing how harmful a treatment is (i.e., serious side effects or complications), clinicians should use the **absolute risk increase (ARI)** and **number needed to harm (NNH)**. The ARI provides clinicians with the actual increase in risk between the treatment and control groups, whereas the NNH tells a clinician how many patients need to receive the treatment for just one patient to experience a harmful treatment. For the example in **Box 5.9**, there is no increase in risk as indicated by a positive ARR and NNT and if erroneously calculated, both a negative ARI of −28% and negative NNH value of −3.6. This means the use of foot orthoses in patients with patellofemoral pain will not increase the number of patients who experience pain compared to those patients who receive flat inserts, and therefore, they will be harmed by the orthoses.

Integrating Evidence into Clinical Decisions

After a clinician has gathered all the information relative to the clinical question, the clinician must decide how to use that evidence. Remember that the purpose of EBHC is to provide the best possible care by integrating evidence with clinical experience and the individual patient's values and circumstances. Evidence should be assessed on the strength (evidence hierarchy), the types of outcomes measured (patient- versus disease-oriented), consistency of the results (all for, all against, or mixed results), and the applicability (evidence is from a similar population or the exact same treatment settings/dosage used). Clinical experience should influence how a clinician uses evidence but should not be an excuse to ignore the best available evidence. Finally, patients should provide input regarding the care they receive.

Disability Models

Athletic trainers will find the models of disablement helpful when trying to understand how disease/injury and treatment affects a patient. The Nagi disablement model is a classic model that represents how disease/injury at a

tissue level affects body systems, the whole person, as well as the person's role in society (Fig. 5.4).[10] Today, most health care professions are moving toward the World Health Organization's (WHO) International Classification of Functioning, Disability and Health (ICF) (Fig. 5.5).[11,12] Unlike the Nagi model, the ICF model takes into consideration how personal and environmental factors may also influence an individual's function, disability, and health. Changes in disease-oriented outcomes typically represent changes in Nagi's impairments or the WHO's body function or structure, whereas changes in patient-oriented outcomes represent changes in Nagi's functional limitations and disabilities or the WHO's activities and participation.

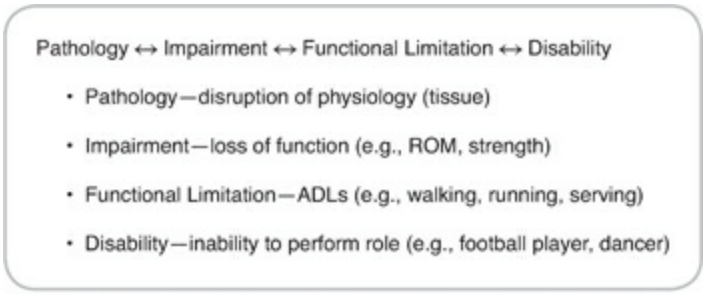

Figure 5.4. Nagi disablement model. *ADLs*, activities of daily living; *ROM*, range of motion.

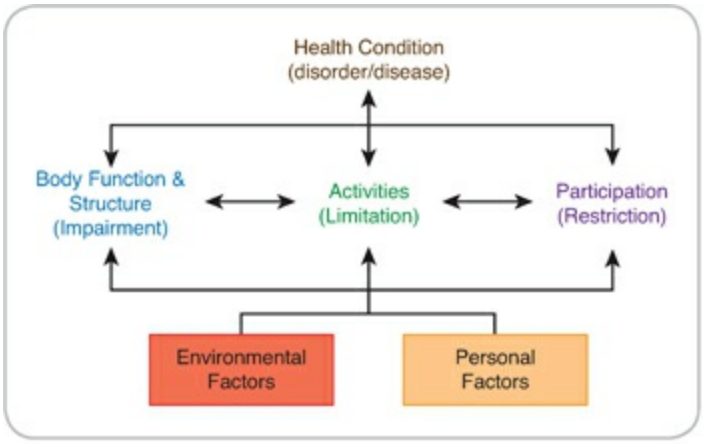

Figure 5.5. WHO's International Classification of Functioning, Disability, and Health Model. (From World Health Organization. *Towards a Common Language for Functioning, Disability and Health: ICF.* Geneva, Switzerland: World Health Organization; 2002.)

Evaluate the Efficiency and Effectiveness of Steps 1 to 4

The last step of EBHC is to reflect on and evaluate the EBHC process and identify ways the process can be more efficient and effective. Clinicians could ask themselves: (1) Was the clinical question well formulated? (2) Did we use the best databases? (3) Did we use the best search terms? (4) Could the search be streamlined by searching multiple databases or using multiple search terms in a single search? (5) Could we appraise the evidence? or (6) Could the evidence be integrated into the clinical decision? Clinicians may find filtered information more helpful when speed is most important. Learning how to use Boolean terms effectively can reduce the reviewing of unrelated material after an evidence search. Becoming more familiar with EBHC-related terms, the hierarchy of evidence, and diagnostic and treatment statistics may improve one's ability to quickly appraise the evidence and determine its impact, reliability, and validity, as well as its applicability to the current clinical decision.

The athletic trainer should develop two separate PICO/T questions: one to determine the accuracy of diagnostic tests to be considered and the second to assess the treatment being considered. Using one of the filtered databases should expedite the evidence search for both PICO/T questions. The athletic trainer will find that diagnostic tests with LR+ of greater than 10 and LR− of less than 0.1 will have the greatest impact on improving diagnostic accuracy. The use of multiple diagnostic tests can also improve the accuracy of the posttest probability of the diagnosis. The clinician can determine the effectiveness of treatment with the relative and the absolute risk rates, whereas NNT would help decide whether resources (e.g., cost, time, equipment) required for the treatment are worth the risk of an unsuccessful treatment.

SUMMARY

1. The purpose of EBHC is to integrate the best available evidence into

clinical decisions while including patient values.

2. Patient-oriented evidence provides a broader view of health than disease- or clinician-oriented evidence.
3. There are five steps to practicing EBHC: (1) Ask a PICO/T question; (2) search health care–related databases; (3) evaluate the evidence for impact, reliability, validity, and applicability; (4) integrate the evidence into the clinical practice; and (5) evaluate the process and look for ways to become more efficient and effective.
4. The SORT, CEBM 2011 Levels of Evidence, and GRADE scales are used to rate evidence.
5. Diagnostic accuracy is dependent on the sensitivity, specificity, and prevalence of diagnostic tests. Positive and negative likelihood ratios of diagnostic tests help clinicians determine the probability of a diagnosis based on the results of the diagnostic tests.
6. The effectiveness of treatment can be determined by the NNT, RRR ARR, ARI, and NNH.
7. The disablement models help to explain the effects injury/disease can have on a person's role in society. Disease-oriented outcomes represent impairments, whereas patient-oriented outcomes represent a patient's activity and participation level.

APPLICATION QUESTIONS

1. You have been an athletic trainer for 20 years. Why is it important for you to practice EBHC?
2. You suspect a patient has lateral epicondylitis of the elbow. How would you determine which diagnostic tests would be most helpful in confirming the correct diagnosis?
3. You diagnose a patient with lateral epicondylitis of the elbow. How

would you identify the most effective treatment for this patient?

REFERENCES

1. Straus SE, Richardsons WS, Glasziou P, et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. 3rd ed. Edinburgh, Scotland: Elsevier Churchill Livingstone; 2005.
2. U.S. National Library of Medicine. Key MEDLINE® indicators. Published February 7, 2014. Updated December 14, 2014. http://www.nlm.nih.gov/bsd/bsd_key.html. Accessed January 14, 2015.
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(3):548–556.
4. Howick J, Chalmers I, Glasziou P, et al. The Oxford 2011 levels of evidence. http://www.cebm.net/index.aspx?o=5653. Accessed January 9, 2015.
5. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926.
6. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257–268.
7. Fitzpatrick R, Davey C, Buxton M, et al. Evaluating patient-based outcome measures for use in clinical trials: a review. *Health Technol Assess*. 1998;2(14):1–74.
8. Mirzatolooei F, Yekta Z, Bayazidchi M, et al. Validation of the Thessaly test for detecting meniscal tears in anterior cruciate deficient knees. *Knee*. 2010;17(3):221–223.
9. Collins N, Crossley K, Beller E, et al. Foot orthoses and physiotherapy in the treatment of patellofemoral pain syndrome: randomised clinical trial. *BMJ*. 2008;337:a1735.
10. Nagi S. Some conceptual issues in disability and rehabilitation. In: Sussman M, ed. *Sociology and Rehabilitation*. Washington, DC: American Sociological Association; 1965:100–113.
11. World Health Organization. *Towards a Common Language for Functioning, Disability and Health: ICF*. Geneva, Switzerland: World Health Organization; 2002.
12. Snyder AR, Parsons JT, Valovich McLeod TC, et al. Using disablement models and clinical outcomes assessment to enable evidence-based athletic training practice, part I: disablement models. *J Athl Train*. 2008;43(4):428–436.