ACR Appropriateness Criteria®
Evidence Table Development

The Evidence Table (ET) summarizes the sources cited in the Appropriateness Criteria (AC) topic narrative. Additionally, the ET quantifies a source's quality based on the number of study quality elements described in that source. The creation and revision of the ET is performed by ACR staff in order to apply the methodology consistently. It also alleviates some of the work burden placed on the topics' authors. Once an ET is constructed, the topic's author and panel members review the ET for completeness and validity.

Description of the ET components

The ET includes five components extracted from a source. These are the **reference** citation information, study **type,** number of **patients or events**, study **objective(s)**, and study **result(s)**.

- The study type designates the source's purpose and design. The purpose of **diagnostic studies** is to diagnose or assess patients by utilizing diagnostic tools while **therapeutic studies** assess the use of treatments and interventions in treating patients. Furthermore, diagnostic and therapeutic studies have different study quality elements that help assess the amount of bias that may be introduced, which may affect the results and conclusions of the study.
- There are four study design types: **experimental**, **observational**, **review/other**, and **meta-analysis**. Additional information for classifying these study types is found in the appendix. These broadly defined study designs contribute to the assessment of study quality. Well-designed and well-executed experimental studies typically are better at controlling biases and determining causality where other study types, like observational studies, may determine only when there is a relationship between events and outcomes. Because of the general inconsistency in the medical literature regarding study design names and definitions, the varying degrees of adherence to study design, the hybridization of study designs, and sometimes the lack of complete information in a source's study methodology to correctly assess the study design, a broad categorization of study design may identify important differences in study quality based on study design.

The sixth component of the ET is **study quality**. A source's **study quality** is determined by the elements described below.

Explanation of Study Quality Elements that apply to both Diagnostic and Therapeutic Studies

- **Statistical Measure**. The statistical measure must compare the results of the treatments, interventions, or diagnostic tools. Measures that only relate to describing the study population or number of events do not meet the criteria for this element. If there is no statistical measure to compare results, the source is categorized as *Review/Other* and has a Study Quality of 4 (not useful as primary evidence). The statistical measure does not contribute to the total number of study quality elements that determine the study quality. Examples of diagnostic measures are sensitivity/specificity, PPV/NPV, mean, median, Kappa, Pearson r, regression co-efficient, etc. Examples of therapeutic measures are odds ratios, survival rates/curves, hazard ratios, mean or median, etc.
- **Uncertainty measure (or range) of the statistical measure**. Examples of diagnostic uncertainty measures are standard errors, confidence intervals, p-values, statistical comparison tests such as t-test, Fisher exact probability, Mann-Whitney U, etc. Examples of therapeutic uncertainty measures are standard errors, confidence intervals, percentiles, power calculations for sample size, etc. Some uncertainty measures are incorporated into the statistical measures. When one of these statistical measures is used for the study's results and there is no specific discussion of an uncertainty measure, the criteria for the uncertainty measure is fulfilled. For example, if the study states the PPV or NPV but does not explicitly state the uncertainty measure, it fulfills the criteria because an uncertainty measure can be calculated.
- **Prospective** study, i.e., the study was designed prior to the data collection, such as before performing the intervention or comparing the index test and reference standard.

Explanation of Study Quality Elements that apply to only Diagnostic Studies

- **Systematic recruitment** of patients *or* if the study has recruited a **consecutive series** of patients
- Imaging, pathologic or clinical **standard of reference** *or* a comparison of at least two imaging, pathologic or clinical tests
- **Reference standard applied** to all subjects in the same way *or* each imaging, pathologic or clinical test in the study has been applied to all subjects in the same way
- Two or more **independent readers** of the index test (not consensus reads) *or* two or more independent readers for each test in the study
- **Index test results** were interpreted without knowledge of the results of the reference standard[1]
- **Reference standard results** were interpreted without knowledge of the results of the index[2]

Explanation of Study Quality Elements that apply to only Therapeutic Studies

- **Allocation of subjects into control group(s) and the intervention group(s)**. The term "control group" refers to a group that does not receive the primary intervention, therapy, or treatment being evaluated in the study. The control group may be comparable subjects who receive no intervention *or* who receive another intervention whose outcomes are accepted *or* may have been previously studied.
- **Random allocation** of subjects into control and intervention groups. This implies that each subject being entered into a trial has the same chance of receiving any of the possible interventions. It also implies that the probability that a subject will receive a particular intervention is independent of the probability that any other subject will receive the same intervention.
- **Length of follow-up** for the study. When length of follow up is present, staff will record it in the ET. If there is no information, the author states the reason for the omission or provided analysis to correct for the lack of this measurement. Examples of the length of follow up are length of follow up, survival rates, re-occurrence rates, toxicity rates or similar measures.
- Account for all of the subjects in the study, i.e., the **disposition of all subjects** in the study. If there are no data, the author states the reason(s) why subjects did not complete the study or why they were excluded.

Blinding as an Element in Assessing Therapeutic Study Quality

There can be many qualifications that would affect the blinding element's contribution to study quality. The ACR Methodology subcommittee recognizes blinding subjects to aspects of a study is important when those aspects can influence how participants report subjective measures (such as, pain levels experienced) that are used as outcomes of the study. However, blinding both researchers and study subjects is likely to be an important step in treatment studies that use measures that are more objective, such as those involving administration of medications versus placebos. Because it may not be ethically responsible to blind subjects or researchers in many treatment trials and because of the complexity of the blinding element and its contribution to study quality, the subcommittee felt it could not be assessed for therapeutic studies consistently and objectively.

[1] In the absence of a reference standard and when more than one type of imaging, pathologic or clinical tests are being compared, this element is fulfilled when the results of at least one of the test(s) has been interpreted without the knowledge of the results of the other test(s) in the study.
[2] In the absence of a reference standard and when more than one type of imaging, pathologic or clinical tests are being compared, this element is fulfilled when the results of all of the test(s) has been interpreted without the knowledge of the results of the other test(s) in the study.

Definitions of Study Quality Categories

Study Quality Category Name	Study Quality Category Definition	Criteria for Diagnostic Studies	Criteria for Therapeutic Studies
Category 1	The study is well designed and accounts for common biases.	The source has all 8 diagnostic study quality elements present.	The source has 5 or 6 therapeutic study quality elements present.
Category 2	The study is moderately well designed and accounts for most common biases.	The source has 6 or 7 diagnostic study quality elements present.	The source has 3 or 4 therapeutic study quality elements present.
Category 3	The study has important study design limitations.	The source has 3, 4, or 5 diagnostic study quality elements present.	The source has 1 or 2 therapeutic study quality elements present.
Category 4	The study or source is not useful as primary evidence. The article may not be a clinical study, the study design is invalid, or conclusions are based on expert consensus. The study does not meet the criteria for or is not a hypothesis-based clinical study (eg, a book chapter or case report or case series description); *or* The study may synthesize and draw conclusions about several studies such as a literature review article or book chapter but is not primary evidence; *or* The study is an expert opinion or consensus document.	The source has 0, 1, or 2 diagnostic study quality elements present.	The source has zero (0) therapeutic study quality elements.
Category M	Meta-analysis studies are not rated for study quality using the study element method because the method is designed to evaluate individual studies only. An "M" for the study quality will indicate that the study quality has not been evaluated for the meta-analysis study.	n/a	n/a

Review of the ET

It is the role of the panel to assess the quality of the ET and to question its assessment of study quality for the sources listed in the document. An initial review of the ET takes place before the first rating round. Panel members are expected review the ET after any edits to the document and prior to the second rating round. Ratings are based on the evidence, which can be best interpreted when the rater understands the quality of the evidence supporting the recommendations.

In the event a panel member (the inquirer) disagrees with the study quality assessment for any article in the ET, the inquirer will contact the ACR staff, report the disputed study quality element, and provide an explanation. Staff will re-evaluate the study quality assignment for any error and report the results to the inquirer. If a requested change to the study quality is warranted based on the accepted methodology, then the change to the study quality is made in the ET.

If the inquirer is not satisfied with the result, a second level review involving the topic author, vice-chair and possibly the chair may be initiated. The second level reviewers will recommend either to use the study quality recommended by staff or propose an alternative and provide an explanation.

The proposed changes will be presented to the all members of the relevant panel for approval. If two thirds of the panel members approve, the study quality change will be made to the ET. If agreement cannot be achieved by the panel (i.e., less than two thirds of the panel agrees), the staff assignment made using the methodology will be the final study quality assessment.

Appendix

Study Type Categories

- *Experimental*

  Experimental studies create differences in the explanatory (independent) variable under controlled conditions and examine any resulting changes in the response (dependent) variable. These studies include methodologies that reduce the potential for bias, for example, randomization, blinding. An example is the randomized controlled trial.

  Characteristics of Experimental Design

  - True experiments have control and manipulation
  - Specifies an experimental group and control group
  - Test cause and measure effect

- *Observational*

  Investigators observe subjects and measure variables of interest (independent variables) without assigning treatments, interventions or outcomes to the subjects. The treatment, intervention or outcome that each subject receives is determined beyond the control of the investigator

  Characteristics of Observational Design

  - Investigator observes variables
  - Specifies cohorts (groups with similar characteristics of interest) or a case (groups with the variable of interest) and control (groups without the variable of interest) group.
  - Test association between variables but not causality.

- *Review/Other*

  Reviews or other studies are case reports, systematic literature reviews, clinical practice guidelines, consensus statements, book chapters, etc. These sources may not have a statistical measure that compares the results but include published literature that examines or reviews other studies, data, surveys, opinions, etc. and summarizes results or concludes outcomes.

  Other sources in this category may be studies that have descriptive statistics only that do not provide a result or outcome to the study, such as incidence or prevalence studies that only describe population or disease trends or patterns.

- *Meta-analysis*

  Meta-analysis studies aggregate information in order to achieve a higher statistical power for the measure of interest, as opposed to a less precise measure derived from a single study. Other methods that do not create pooled samples using statistical methods such as systematic literature reviews and clinical practice guidelines are not included in the definition of meta-analysis. Meta-analysis studies are not multisite studies even when one of the studies in the meta-analysis is a multisite study because the same protocol may not be implemented in each study.

  Characteristics of Meta-analysis Design

  - Systematic review of literature
  - Pooled results
  - Provides a precise estimate of treatment effect or diagnostic performance