

ACR APPROPRIATENESS CRITERIA® Evidence Table Development

I. Evidence Table Development

- A. **Purpose:** To categorize each article using an existing typology for research studies, as well as rating of the strength of evidence in each study.
- B. **General Information:** The research associate (RA) develops an evidence table (ET) for each Appropriateness Criteria topic. The ET consists of the citation, study type, number of patients, study objective, study results and strength of evidence (SOE) for all articles included in the narrative reference list.
- C. **Defining the Elements of the Evidence Table**
 1. *Citation* – Author(s), title, journal, volume and date should be included.
 2. *Study Type* – Studies are classified based on the study types shown below. The first nine study types are used for studies on treatments while the remaining eight are used for studies on diagnostic interventions. (see appendix A)
 3. *Patients/Events* – The number of patients included in the study and/or the number of images evaluated is listed in this column. The number of observers or interpreters of images, if mentioned in abstract, should also be included. If patients in the study are placed in different groups, the number within each group should be listed separately. If the study is a quantitative review, the number of studies included in that review should be listed. The column should be filled with N/A to indicate “not available” if the number of patients is unavailable or if the document is not concerned with individual patients (as in the case of guidelines).
 4. *Study Objective* – The central question(s) addressed by the study is listed in this column. Some indication of how the study was structured may be included. For example, “A retrospective review of patient records was undertaken to determine whether those who had an MRI had a lower rate of surgery than patients who had ultrasound”.
 5. *Study Results* – The principal findings of the research should document the overall conclusions of the study authors (e.g., CT should not be performed for this patient group). Additionally, specific numerical results should be included when practical.
 6. *Strength of Evidence (SOE)* – A four category rating scale is used to describe the evidence. “Category 1” denotes the strongest level of evidence and “Category 4” the weakest. For each study, a rating from the scale below is assigned, with the exception of “book chapters” which are assigned “N/A”.

CODE	CATEGORY NAME	CATEGORY DEFINITION
1	Category 1	The conclusions of the study are valid and strongly supported by study design, analysis and results.
2	Category 2	The conclusions of the study are likely valid, but study design does not permit certainty.
3	Category 3	The conclusions of the study may be valid but the evidence supporting the conclusions is inconclusive or equivocal.
4	Category 4	The conclusions of the study may not be valid because the evidence may not be reliable given the study design or analysis.

II. Rating Strength of Evidence

There are distinct criteria for rating SOE for studies of diagnostic accuracy and studies of treatment efficacy. Each is discussed below.

A. Controlled Trials

The characteristics that characterize high quality controlled trials include:

1. Random assignment to treatment and control groups
2. Number and representativeness of patients included in study
3. Prospective statement of study hypothesis
4. Uniform application of treatment as it would be performed by most clinicians
5. Blinding (of both patients and investigators)
6. Appropriate statistical analyses

ACR APPROPRIATENESS CRITERIA[®] Evidence Table Development

A study that includes all of these elements should be assigned to category 1 (e.g., randomized control trials) for the SOE category and a study that has none of them should be assigned to category 4 (e.g., qualitative reviews, case reports, and opinions). For example, a blinded, prospective study with large number of patients, adequate description and standardized application of the diagnostic test and an appropriate reference standard should be coded as a “1”.

The lack of any of the critical elements weakens the study design. Studies with one or more of these design elements missing should be coded as a “2” or lower. There is some judgment involved with whether a study is rated a 2 or 3. For example, blinded, prospective studies with small number of patients or retrospective studies with large number of patients may be coded as “2” or lower depending on the strength of the other elements. Retrospective studies with small number of patients may be rated only “3” or lower depending on which of the other critical design elements are present.

When assigning the category, the RA first determines if the study is clearly a category 1 or category 4 by reviewing the abstract for the critical design elements (i.e., all of the design elements are present or none of the design elements are present, respectively). If the study does not fit these two categories, the RA determines the number and quality of the critical design elements that are in the abstract. From this information, the RA is able to categorize the study’s strength of evidence as category 2 or category 3. In general, if there is only one element missing, the study would likely be categorized “2”. If none but two elements are present, the study would likely be categorized “3”. However, the actual category depends on the number and quality of the elements in the study. Sometimes, the full article will be reviewed and provided additional information that increases the confidence in the categorization.

If the study is a meta-analysis or systematic review of randomized clinical trials, the RA should consider the following questions:

1. Was the search for potentially relevant research conducted so as to ensure that most of the high quality studies informing the topic have been identified?
2. Were the criteria for deciding which studies to include in the meta-analysis specified and, if so, did the bias the results in any direction?
3. Did the researchers use an accepted method for combining the data and did they report on how results vary if alternative methods are employed for dealing with heterogeneity of results?

If the meta-analysis or systematic review meets all these criteria, the individual studies still need to be assessed for SOE. If all the studies have the same level of SOE, then the overall SOE would be that level. If some of the criteria questions are not met, the overall SOE will be lower. If the SOE for all the studies varies and all the criteria questions are met, the overall SOE should be no more than the highest SOE of an individual study. Also the overall SOE should reflect the number and quality of all of the studies included in the analysis. The ACR is currently reviewing its methods of assigning SOE to meta-analysis and systematic review studies.

B. Observational Studies

Because of its inherent design, observational studies are usually not rated higher than “2”. In diagnostic studies, observational studies may be rated higher than “2” based on the GRADE methodology for assessing evidence in diagnostic studies¹. (Even though the quality of the evidence for diagnostic assessment may be high, the quality of the recommendation for diagnostic tests may be low because diagnostic tests may not predict positive patient outcomes.)

Rating of observational studies is dependant on how well the study design overcomes any potential bias. Both very large samples (in the thousands) and carefully selected controls are ways to reduce the likelihood of treatment assignment bias. Observational studies for treatment outcomes with large samples and carefully selected controls may receive a “2”. An observational study for a similar treatment outcome that examines only a handful of patients and compares their outcomes against a clearly non-comparable group (for example comparing outcomes for younger treated patients with those of older untreated patients) should receive a ‘4’ for SOE because the potential bias introduced may influence the study outcomes.

If the study is a systematic review of observational studies, the RA should consider the same questions as in RCTs:

1. Was the search for potentially relevant research conducted so as to ensure that most of the high quality studies informing the topic have been identified?

¹ Grading quality of evidence and strength of recommendations for diagnostic tests and strategies, Schunemann HJ et al, BMJ, May 2008; 336: 1106 - 1110

ACR APPROPRIATENESS CRITERIA[®]

Evidence Table Development

2. Were the criteria for deciding which studies to include in the systematic review specified and, if so, did the bias the results in any direction?
3. Did the researchers use an accepted method for combining the data and did they report on how results vary if alternative methods are employed for dealing with heterogeneity of results?

The same method for providing an overall SOE to clinical / controlled trials should be performed for observational studies. If the systematic review meets all the questions above, the individual studies still need to be assessed for SOE. If all the studies have the same level of SOE, then the overall SOE would be that level. If some of the criteria questions are not met, the overall SOE will be lower. If the SOE for all the studies varies and all the criteria questions are met, the overall SOE should be no more than the highest SOE of an individual study. Also, the overall SOE should reflect the number and quality of evidence for all of the studies included in the systematic reviews. The ACR is currently reviewing its methods of assigning SOE to systematic review studies.

C. Reviews

Review studies generally are rate as a “4” because they generally do not address the issue of bias. In rating the SOE for reviews, the RA should look for any methods the study may employ that systematically reduce bias. If sufficient techniques have been employed to reduce bias, the RA may rate the study a “3”.

If the reviews discuss more than one study, questions similar to systematic reviews for controlled trials should be addressed only when any of the studies in the review rate higher than a “4”. It is unlikely that a systematic review would be performed on review studies.

D. Additional Guidance for Rating the SOE in Diagnostic Imaging Studies

The critical design elements that characterize high-quality, diagnostic imaging studies are:

Critical Design Element	Examples / Clarification
Representative study population <ul style="list-style-type: none"> ○ Unbiased selection of the study population with sufficient variability to be representative of the patient population of interest ○ “Large” number of representative patients included in study. 	Both representativeness and numbers are important to this element. If the study only has men, it may not be truly representative of the population which could introduce bias in the results. The number is not an absolute value but relative to the prevalence of the condition being studied. Another example that may influence the representativeness of the study is drop out rate.
Uniform application of the diagnostic test. The imaging study is well described and has an accepted or well-known standardized procedure.	Results can be biased if the application of the diagnostic test is dependent on the skill level of the operator, the procedure is variable, or the procedure performed during the study does not conform to how it is normally performed.
Prospective evaluation of patients	Randomly assigning subjects to be in a protocol group or a control group prior to the treatment or intervention reduces the possibility of introducing bias.
Confirmation with a reliable procedure of the observed results.	Imaging results may be confirmed with clinical tests such as biopsy.
The conclusions of the investigators who interpreted the test results were not influenced.	Blinded and double-blinded designs help to reduce investigator bias by not providing any clues about how the results should be interpreted.
Verification of results and avoidance of verification bias	Independent verification of the results confirming the findings of a diagnostic study to reduce any bias. Were patients with negative results on initial diagnostic procedures verified through a different mechanism than patients with positive results?

A diagnostic study that includes all of these elements may be assigned to category “1” for the SOE and a study that has none of them should be assigned to category “4” (e.g., qualitative reviews, case reports, etc.). The lack of any of the critical elements weakens the study design. Studies with one or more of these design elements missing should be coded as a “2” or lower. There is some judgment involved with whether a study is rated a “2” or “3”. A [similar method as what was described in the treatment section](#) should be followed to assign a “2” or a “3”.

Although the ultimate decision on the score for both diagnostic and treatment studies is subjective, the RA should use the guidance above to determine the appropriate score for each study.

ACR APPROPRIATENESS CRITERIA® Evidence Table Development

Appendix A

Code	Study Type	Description
<i>Treatment Study Types</i>		
<i>Controlled Trials</i>		
1	Randomized Controlled Trial (RCT) – Treatment	Studies that compare treatments by randomly assigning prospective patients to the different treatments of interest and then measure the different outcomes for each group in order to get an estimate of treatment effect.
2	Controlled Trial	Studies that compare treatments by assigning prospective patients to the different treatments of interest and then measure the different outcomes for each group in order to get an estimate of treatment effect. These studies differ from RCTs in that patients are placed into treatment groups without randomized assignment.
<i>Observational Studies</i>		
3a	Cohort Study	A study that follows over time two or more groups that shares at least one common characteristic (where they were treated, when they graduated high school, occupation) to determine the likelihood of disease, the disease’s natural history, prevalence, or pattern of healthcare intervention, or outcomes of care.
3b	Cross-sectional	A study that compares two (or more) groups during a specific period in time to determine differences in disease patterns, treatment, or outcomes.
3c	Case-control	A study that compares one group (cases or treatment group) that received the study treatment or intervention to another group (controls) that did not received the intervention. Characteristics that may add or detract from the intervention effect are hypothesized and identified within the treatment group. These same or similar characteristics are matched to individuals in the control group who have not received the treatment. The differences in the two groups after the intervention are observed and evaluated.
<i>Reviews</i>		
4	Clinical Series	Reports of outcomes for patients treated with a specific therapy without comparisons made to outcomes for other patients should be placed in this category.
5	Case Reviews	Any report on a researchers experiences with treatment for a very small number of patients (<10) should be assigned to this category.
6	Anecdotes	Reports of personal experiences of the author(s)
7	Summaries	Non-quantitative summaries of other research on therapeutic efficacy, including “state-of-the-art” essays should be included in this category.
<i>Diagnostic Study Types</i>		
<i>Controlled Trials</i>		
8	Randomized Controlled Trial (RCT) – Diagnostic	These randomized clinical trials determine the benefits of diagnostic modalities (as opposed to treatment effect). These studies are always prospective in nature.
<i>Observational Studies</i>		
9	Comparative Assessment	Studies that compare the specificity, sensitivity, diagnostic yield, or diagnostic accuracy of two or more modalities. The study can be prospective or retrospective.
10	Clinical Assessment	A quantitative study reporting on the specificity, sensitivity, diagnostic yield, or diagnostic accuracy of a single diagnostic modality.
<i>Reviews</i>		
11	Quantitative Review	This study type involves cumulating quantitative estimates from completed research of the specificity, sensitivity, diagnostic yield, or diagnostic accuracy of one or more diagnostic modalities.
12	Qualitative Review	A summary of research to date that does not include summative statistics.
13	Descriptive Study	A qualitative or quantitative description of the benefits of imaging modalities based on other than a small sample.
14	Case Report	A qualitative description of the benefits of an imaging modality based on a small number (<10) selective cases.
15	Other	Includes epidemiological studies, research on patterns and costs of care, guidelines, studies on prognostic factors, book chapters, and any study designs other than those listed above.