**Levels of Evidence for Human Studies of Integrative, Alternative, and Complementary Therapies**

**Introduction**

A classification system has been developed by the National Cancer Institute's [PDQ Adult Treatment Editorial Board](https://www.cancer.gov/publications/pdq/editorial-boards/adult-treatment) to allow the ranking of human cancer treatment studies according to statistical strength of the study design and scientific strength of the treatment outcomes (i.e., endpoints) measured. This classification system has been adapted to allow the ranking of human studies of integrative, alternative, and complementary therapies for cancer. The purpose of classifying studies in this way is to assist readers in evaluating the strength of the evidence associated with particular treatments. However, not all human studies are classified. Only those reporting a therapeutic endpoint(s), such as tumor response, improvement in survival, or measured improvement in quality of life, are considered. In addition, anecdotal reports and individual case reports are not classified because important clinical details are often missing, the evidence from them is generally considered weak, and there is an increased probability that similar results (either positive or negative) will not be obtained with other patients. Furthermore, reports of case series are excluded when the description of clinical findings is so incomplete as to hinder proper assessment and interpretation.

**Strength of Study Design**

In the classification system, a numeric scale from 1 to 4 is used to indicate the statistical strength of the study design, with 1 assigned to studies having the strongest design and 4 assigned to studies having the weakest design. Further subdivision of some design categories yields finer measures of strength. The various types of study design are described below in descending order of strength:

1. **Randomized controlled clinical trials:** Studies in which participants are assigned by chance to separate groups for the comparison of different treatments. It is the patient's choice to be in a randomized trial, but neither the researcher(s) nor the patient can choose the group in which he or she will be placed. Using chance to assign people helps to ensure that the groups will be similar and that the treatments they receive can be compared objectively. At the time of a trial, there is uncertainty about which of the treatments is best. These trials can be "double-blinded" or "nonblinded." Double-blinded trials have a stronger study design. A designation “s” is given to acupuncture studies where the patient is blinded when using a sham control.
   1. **Double-blinded:** Neither the patients nor the researcher(s) know which patients are receiving the therapy under study or the comparison (i.e., control) treatment.
   2. **Nonblinded:** The researcher(s) and the patients know what treatment is being given.
2. **Nonrandomized controlled clinical trials:** Studies in which participants are assigned to a treatment group based on criteria that may be known to the researcher(s), such as the patient's birth date, chart number, or day of clinic appointment. With this type of study design, there is less confidence that the group receiving the treatment under study and the control group are comparable.
3. **Case series:** Studies that describe results from a group or series of patients who all received the treatment that is being investigated. These studies have a weak design, due, in part, to the absence of a control group. Different types of case series, in descending order of strength, are as follows:
   1. **Population-based, consecutive case series:** The study population is well-defined and is either the entire population of interest or a representative random sample of the larger population from which it is drawn. The study subjects receive treatment in the order in which they are identified by the researcher(s).
   2. **Consecutive case series:** Studies describing a series of patients who were not limited to a specific population and who received treatment in same order in which they were identified by the researcher(s).
   3. **Nonconsecutive case series:** Studies describing a series of patients who were not limited to a specific population and who do not represent a consecutive series of patients identified and treated by the researcher(s).
4. **Best Case Series:** From a larger series of patients, only the cases that appear to have benefited from the treatment under study are reported. These studies have the weakest design.

**Strength of Endpoints Measured**

The scientific strength of a study's findings is determined by the endpoint(s) measured. In the classification system, a progressive alphabetic scale is used to indicate the scientific strength of endpoints, with the letter A assigned to the strongest endpoint that can be measured and the letter D assigned to the weakest endpoint. Commonly measured endpoints in human cancer treatment studies are listed below in descending order of strength:

1. **Total mortality:** The proportion of the study population that died. Frequently called the death rate. Measured from a defined point in time, such as the time of diagnosis or the time since treatment was initiated. This is the most easily defined and objective endpoint. The inverse of total mortality (i.e., overall survival), may be the reported value.
2. **Cause-specific mortality:** Death from a specified cause in the population under study, for example, death from cancer versus death from side effects of therapy versus death from other causes. This endpoint is more subjective than total mortality. When death from disease (e.g., cancer, heart disease, etc.) is the measured endpoint, the inverse value (i.e., disease-specific survival), may be reported instead.
3. **Carefully assessed quality of life:** Although a very subjective endpoint, quality of life is an extremely important endpoint to patients. The strength of a quality of life assessment depends on the validity of the instruments (i.e., questionnaires, psychologic tests, etc.) used.
4. **Indirect surrogates:** These are measures that substitute for actual health outcomes, and they are subject to investigator interpretation. In descending order of strength, indirect surrogates include the following:
   1. **Disease-free survival:** Length of time no cancer was detected after treatment.
   2. **Progression-free survival:** Length of time disease was stable or did not get worse after treatment.
   3. **Tumor response rate:** The proportion of patients whose tumors responded to treatment and the degree or extent to which the tumors responded.

**Combined Level of Evidence Score**

A combined level of evidence score is calculated for each qualifying study—except for Best Case Series—by joining the score for statistical strength of study design with the score for strength of the endpoint(s) measured. Because of their extremely weak study design, Best Case Series are given a score for strength of study design only (i.e., a level of evidence score of 4). The level of evidence scores for the remaining study types range, in decreasing order of strength, from 1iA to 3iiiDiii.