GRADING OF EVIDENCE FOR THERAPEUTIC QUESTIONS
American College of Emergency Physicians Clinical Policies Committee

Citation: _____

Design 1	Randomized controlled study (with concealed allocation)
Design 2	Non-randomized interventional study including patients who do not receive the intervention
Design 3	Case series

Applicable to Clinical Question: Direct Indirect Not

Dimensions for Grading (consider both quality of execution and importance to result):
NR/NA: Not reported, not applicable, or unclear.

			Comments
Inclusion criteria defined & appropriate	Y	N	NR / NA / U
Appropriate sampling	Y	N	NR / NA / U
Patient groups comparable at baseline	Y	N	NR / NA / U
Appropriate blinding (single, double)	Y	N	NR / NA / U
Outcomes defined & appropriate	Y	N	NR / NA / U
Outcome measured in a valid and reliable way	Y	N	NR / NA / U
Objective or masked outcome assessment	Y	N	NR / NA / U
Appropriate level of attrition	Y	N	NR / NA / U
Accounting for drop-outs and crossovers	Y	N	NR / NA / U
Appropriate sample size	Y	N	NR / NA / U
Generalizability	Y	N	NR / NA / U
Data managed appropriately	Y	N	NR / NA / U
Analyses appropriate	Y	N	NR / NA / U
Conclusions supported by the results	Y	N	NR / NA / U
Industry sponsored	Y	N	NR / NA / U

Downgrading:
- ☐ No downgrading (no methodological limitations and directly applicable)
- ☐ Downgrade 1 level (only minor methodological limitations)
- ☐ Downgrade 1 level (indirectly applicable)
- ☐ Downgrade 2 levels (major methodological limitation[s])
- ☐ Fatally flawed or not applicable

Class of Evidence: **I** **II** **III** **X**

Notes:

Reviewer: _____ **Date:** _____

Guidelines for Use:

1. Use the top grid to assign a **Design** (1, 2, or 3) based on the study's design. Some designs may not fit this schema and should be assessed individually.
2. To qualify as **Design 1**, the randomization process must be explicitly described to ensure investigators could not have influenced a patient's treatment assignment (**concealed allocation**). Studies not describing concealed allocation should be considered **Design 2**.
3. **Applicability to the clinical question** relates to whether the study being evaluated is directly, indirectly, or not applicable to the clinical question proposed as part of the clinical policy.
4. Then assess the quality of the execution of the study using the list of important dimensions as reminders. Important dimensions to be considered when assessing the quality of a study include:
   a. A clear description of how patients were included in the study, including explicit and appropriate **inclusion and exclusion criteria** and appropriate **sampling** to generate the study sample from the base population. If conducted poorly, one or both features may introduce selection bias.
   Random allocation is used to avoid confounding. Patients in different study groups should be similar in baseline characteristics that might impact outcome. Occasionally, investigators use statistical techniques (e.g., multivariable analyses) to attempt to adjust for confounders. In general, if patient groups are not comparable at baseline, multivariable analyses should be performed.
   c. To avoid problems resulting from multiple comparisons and post hoc analyses, a **primary outcome measure** should be specified. The outcome measure should also be patient-relevant. Most often patient-relevant outcomes have face validity (i.e., they make sense - e.g., death or disability). Occasionally, studies will use surrogate markers that have been shown to predict patient-relevant outcomes (e.g., hypertension predicts stroke and myocardial infarction). Studies with validated surrogate outcomes need not be downgraded.
   d. Bias can be introduced by an investigator's expectation of the effect of the intervention. **Masked outcome assessment** is important when the outcome being measured is subject to expectation bias. Sometimes, the outcome measure is not subject to expectation bias (e.g., mortality), and under such circumstances the outcome can be considered **objective**. When a study does not employ objective or masked outcomes, expectation bias can be lessened by having an independent observer assess the outcome.
   e. **Attrition** (i.e., patients who do not complete the study) may significantly bias a study. The impact of patients dropping out of a study or crossing over (i.e., receiving an intervention they were not initially intended to receive) can be lessened by various statistical techniques (e.g., intention-to-treat, imputation, sensitivity analyses). In general, attrition should be <20% but in some instances should be significantly less.
   f. **Sample size** should be sufficient to provide adequate precision of estimates and to prevent type II errors (i.e., not finding a difference when one actually exists).
   g. **Generalizability** refers to the ability to generalize the study's results to other patients or settings. Consider the representativeness of the patient population included in the study (e.g., were only patients with severe disease included).
   h. **Data management** refers to whether the data were appropriately handled during collection and analyses; this may include issues of use of a DSMB, whether authors had access to data, and who performed analyses.
   i. **Analyses** should be appropriate and valid for the study design (e.g., appropriate use of multivariable methods, confounders, interactions, collinearity, model fit; use of propensity methods; appropriate handling of missing data; etc.). j. **Conclusions supported by results** refers specifically to whether the conclusions are appropriately aligned with reported results or whether the authors took liberty in over- or under-extending their conclusions.
   k. **Industry sponsored** studies often are influenced, either in their design, performance, or reporting, by the company, which may introduce bias. Who controlled and analyzed the data? Was a DSMB used?
5. At the **Downgrading** section, summarize the quality of execution and applicability to the clinical question into a decision on downgrading. The idea here is that the maximum evidence class that can be assigned is limited by the Design (i.e., Design 1 can support up to Class of Evidence I, but Design 2 can only support Class of Evidence II or lower, and so on). Essentially, the quality of execution is used to "downgrade" studies from the maximum class, as shown in the table below. Additionally, applicability to the clinical question also relates to downgrading. (e.g., not applicable studies receive a Class of Evidence "X"). Evidence Class X studies will not be used to support clinical policies. Use the downgrading results to generate a **Class of Evidence** based on the table below.

	Design		
Downgrading	1	2	3
None	I	II	III
1 level	II	III	X
2 levels	III	X	X
Fatally flawed or NA	X	X	X