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

Citation: _____

Design 1	Prospective cohort study; Cross-sectional study
Design 2	Retrospective cohort study; Case-control study
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/U: Not reported, not applicable, or unclear.

			Comments
Inclusion criteria defined & appropriate	Y	N	NR / NA / U
Appropriate sampling	Y	N	NR / NA / U
Spectrum of cases appropriate to likely test usage	Y	N	NR / NA / U
All patients received criterion standard	Y	N	NR / NA / U
Standard administered independent of study test	Y	N	NR / NA / U
Standard interpreted independent from study test	Y	N	NR / NA / U
Appropriate level of attrition	Y	N	NR / NA / U
Accounting for drop-outs	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**, all patients in the study must have their diagnosis determined by an objective criterion standard, or a surrogate measure that is reliable and valid. Studies not describing concealed assessment 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.
   b. The **spectrum of cases** evaluated should be appropriate to the likely use of the study. Although this relates to generalizability, this also may introduce bias in measures of diagnostic accuracy.
   c. The **outcome measure** specified should be the criterion (gold) standard. Occasionally, studies will use surrogate outcomes; if the surrogate is valid and reliable, such studies need not be downgraded.
   d. The **criterion standard** should also be administered and interpreted independent of the diagnostic test being evaluated. A reasonable surrogate may be used in place of the criterion standard. Bias can be introduced by an investigator's expectation of the result of the criterion standard. Masked outcome assessment is important when the outcome being measured is subject to expectation bias. When a study does not employ objective or masked criterion standard, 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 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., use of conventional diagnostic testing statistics – sensitivity, specificity, predictive values, likelihood ratios, and receiver operating characteristics curves, 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