Delfini Evidence Tool Kit
Study Validity & Usability: Primer and Evaluation Tool for Primary Sources

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
General	1.	☐ **Who is sponsoring and funding the study? What are the affiliations of the authors?** **Considerations:** This can be helpful information and is worth noting since it may have implications relating to potential biases. However, it is important to recognize that any entity or person involved in research, even if strictly academic, may have a bias.	
Study Design Assessment	2.	☐ **Is the design appropriate to the research question?** ☐ **Is the research question useful?** **Considerations:** ☐ For questions of **efficacy,** is this an **experimental study design** (meaning there was **no choice** made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, emotional functioning, physical functioning or health-related quality of life) ☐ Reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ Are **intermediate outcome markers** used? If yes, is there valid and useful evidence that proves that the intermediate marker is valid for clinical significance?	
Internal Validity Assessment 🡻	3.	☐ **Can bias, confounding or chance explain the study results?** Ensure **prespecified** (*a priori*) and **appropriate** — ☐ research questions, ☐ populations to analyze, ☐ outcomes (see below for **composite outcomes**), **Threats for Composite Outcomes include** — ☐ Composite outcomes are not appropriate ☐ Outcomes are or may be double-counted ☐ Outcomes are not correlated ☐ Less important outcome – or outcome subject to external control – is driving the results ☐ It is not possible to analyze data in the composite endpoints to tell what is driving the measures of outcomes ☐ Combination is of severe outcomes with mild ones or with process measures ☐ Outcomes are not independent of other variables or are subjective or under external control, choice or influence ***Aside**: If several of the outcomes are biologically related, this could lend support that these are true outcomes and not a result of chance findings*	
Confounding	4.	☐ Absence of known or unknown **confounders** Might there be another factor associated with the study variable that may account for the conclusion **instead or in addition to** of the variable under study? (This would be a **confounder**.) If yes, have the researchers identified potential confounders and adjusted for them? **Caution:** Adjustment may not eliminate this problem. This is best addressed through successful **randomization**. ***Aside:** Confounding is difficult to assess. Most confounders are unknown.*	
Selection Bias	5.	**Study Groups** are — ☐ appropriate for study ☐ of appropriate size ☐ concurrent	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		☐ similar baseline characteristics (e.g., demographics and prognostic factors) ☐ **Observations**: similar as possible at the study's inception, except for the study topic ☐ If no, how robust are the findings — are the results similar when the populations are adjusted and unadjusted? **Review** ☐ How subjects were identified and selected for study, ☐ Demographics, inclusions & exclusions, baseline characteristics **Considerations** ☐ Age, race, gender, prognostic factors and disease spectrum issues such as range of symptoms, severity, signs, lab results and other diagnostic tests, rate of disease progression, response to therapy, etc. ☐ Is the sample size large enough to be representative? (Rule of thumb: less than 100 in each group can be considered small – medium.) ☐ Was the sample chosen for study representative of the defined population at risk or typical patients? ☐ Are all patients accounted for in the report of baseline characteristics? ☐ Are there statistically significant differences in baseline characteristics between groups? If there are many statistically significant differences, this might suggest that there were problems with randomization and/or concealment of allocation. ☐ Is there a difference in the numbers of patients in important prognostic areas or in combinations of prognostic areas that might affect the study results? ☐ Non-responders of placebo being excluded can bias study results in favor of the intervention ☐ Non-responders of the study drug — but not the comparator — being excluded can bias study results in favor of the intervention ☐ Study is a comparison of a monotherapy versus a combination therapy (which includes the monotherapy agent), and patients not responsive to the monotherapy were excluded, which can bias the study results in favor of the combination therapy **Considerations for Observational Studies** ☐ Is the control group a logical one for study and selected in such a way to avoid bias? ***Asides*:** *Selection bias is often a significant issue, especially in observational studies. Any difference between groups could actually be the true cause of the outcomes.* *Baseline characteristics are important to assess who was actually studied*	
Randomization	6.	☐ Randomization methods for sequence generation are truly **random** and sequencing avoids potential for anyone **affecting assignment** to a study arm ☐ Was there any variation in randomization that might have resulted in patients not being truly randomized or in being "unrandomized" such as anything that disrupts the effects of randomization. This could involve eliminating subjects after randomization or not accounting for subjects lost to follow-up in the analysis. **Adequate** ☐ Computer generated random numbers ☐ Urn randomization ☐ Random number tables (not open) ☐ Drawing lots ☐ If block assignment utilized, random sequence generation method plus use of permuted blocks (i.e., randomly chosen block sizes that vary) **Threats include —** ☐ Authors state study was randomized, however no details of randomization reported	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		☐ Deterministic, open or predictable methods, such as a sequential method, were used which are not random (e.g., alternation, date of hospitalization, date of birth, open tables of random numbers, related to prognosis, etc.) ☐ Block size may have been known to investigators prior to assignment – problematic if concealment of allocation is not adequate or if stratified by site ☐ Block size equal to the number of arms. ☐ Threat is increased if concealment of allocation is not adequate	
Randomization Concealment of the Allocation Sequence	7.	☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm ▪ Were adequate steps taken to ensure that no one could affect the group to which a subject would be assigned? Were sufficient measures taken to ensure no one could decipher the code, obtain the code, manipulate an assignment, withhold an assignment, channel a patient, affect the timing for assignment, etc. **Adequate** Combinations of training and methods employed to ensure that the upcoming group assignment is not learned or deciphered in advance of patient enrollment or in such a way that assignments can be manipulated or withheld or that patients can be directed into groups such as — ☐ Training provided to staff on concealment and its import ☐ Central randomization such a call-in center with enrollment performed prior to group assignment and just-in-time per enrolled patient ☐ Irreversible assignment mechanism such as numbered vehicle or truly random method with no opportunity to break the blind, such as sealed, sequentially numbered identical containers ☐ Other convincing method **Possibly Adequate** ☐ Some details reported, but without sufficient detail as to fully ascertain the likelihood of undermining the process. Example, stating central randomization was performed, but without providing further details to know that staff were trained so that they would not provide several upcoming assignments at once. **Threats include —** ☐ No details of concealment reported ☐ Inadequate methods used ☐ Same person administers and without other methods of protection against concealment ☐ Envelopes used without some kind of assurance of procedures that strictly prevent anyone from learning of the upcoming group assignment such as opening the envelope or being able to read through the envelope ☐ Deterministic, predictable or open methods of randomization (see Randomization) ***Aside:*** *Studies that do not perform effective concealment of allocation or report it may have relative differences in benefit size, typically favoring the intervention of up to 20-40 percent*	
Observation Bias ⬇	8.	☐ **Might bias have been introduced during study performance, follow-up or assessment?**	
Performance Bias ⬇	9.	☐ **No bias or difference, except for what is under study, between groups during course of study** **Such as —** ☐ intervention design and execution (example – if this were a surgery study, were the surgeons skillful?) ☐ co-interventions ☐ concomitant medication use (review allowed and disallowed medications and compare usage, dosing and duration between groups; if there is no report, assess likelihood there is non-	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat

reported use given the population)
- ☐ adherence (review method used and compare outcomes between groups or consider if adherence differs because of difference in interventions or other factor — for example, surgical adherence would be 100% and medical treatment would be significantly less in all likelihood)
- ☐ inappropriate exposure or migration (see below in this section)
- ☐ cross-over threats (see specific questions in **cross-over design** section)
- ☐ protocol deviations
- ☐ measurement methods (see below in this section)
- ☐ study duration (consider length of time needed for occurrence of primary effect, side effects, symptoms, regression, remission, recurrence, survival, duration of treatment, duration of follow-up, etc.)
- ☐ other?

Timing

- ☐ In the course of the study, could passing of time have had a resulting impact on any outcomes? Examples —
  - Might the population under study have significantly changed during the course of the trial? (If overall, this could be an external validity issue. If different between groups, this would be an internal validity issue.)
  - Might the time period of the study affect study results (eg, studying an allergy medication, but not during allergy season)?
  - Could awareness of adverse events affect future reporting of adverse events?
  - Could registry data be affected by clinicians changing their selection of patients for certain drugs (e.g., excluding high-risk patients) with the result being fewer reported adverse events over time.
  - Could test timing or a gap in testing result in misleading outcomes (e.g., in studies comparing one test to another, might discrepancies have arisen in test results if patients' status changed in between applying the two tests)?
  - Could transmittal of records result in a gap in timing that could result in misleading outcomes (eg, in a study utilizing a method to compare mortality counts between investigators and independent reviewers, might a gap in time result in discrepancies in counting mortality)?

Considerations for Migration or Exposure

RCTs: Migration to Other Study Group

- ☐ Review run-in period and wash-out periods
- ☐ Choice of treatment made renders those outcomes observational

Observations: Exposure

- ☐ Were measures taken to ensure that those in the comparison group weren't "exposed?"
- ☐ Was exposure determined by reasonable means and did it precede the outcome of interest?

Considerations for Measurement Methods

- Consider need to critically appraise study reporting on effectiveness of tool. (Validated does not necessarily mean that the measurement tool is valid.)
- Use of tools that are not standard should be scrutinized.
- Evaluate likelihood of detecting meaningful clinical significance.

Threats include —

- ☐ Measurement instruments or methods do not appear appropriate due to problems with —
  - ☐ Methods of measurement
  - ☐ Sources for data collection
  - ☐ Methods of data collection
  - ☐ Not sufficiently comprehensive

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		***Measurement Tip**: Watch for "reporting bias" such as self-reporting when behaviors are hard to track (e.g., diet) or often hidden (e.g., smoking). Watch for "recall bias" due to differences in groups, such as mothers with babies with birth defects compared to those with normal babies*	
Blinding	10.	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved **Threats include —** ☐ Subjects were not blinded ☐ All those assigning treatment, providing care, performing the intervention or otherwise working with subjects blinded ☐ Intervention was not effectively disguised from the comparator ☐ Blind likely to have been broken ☐ Patients might have discerned active agent due to familiarity with the agent (e.g., taste, sensations or side effects, especially if patients were not naïve to the active agent) ☐ Exposure issues — Example: Titration methods with a placebo arm or patients starting with a placebo inhaler would have discerned which was the active medication. ☐ Outcomes could be affected by knowledge of treatment through subjective measurements used or influence or control of patient or provider ***Asides**: Threat, especially increased if concealment is not adequate. Studies have reported that lack of blinding is likely to overestimate benefit by a relative 17 to 44%. Even objective outcomes can be affected by non-blinding.*	
Comparators	11.	☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) **Considerations** ▪ Ideal are arms for intervention, usual care and placebo. Also make sure that the details of the intervention and the comparison are clear. Combination modalities may increase the potential for bias. ▪ A comparison to "usual" care can be a problem because it is not clear what the intervention is being compared to. Are follow-up periods the same, etc? **Additional threats include —** ☐ No placebo comparison information and reviewers did not seek out this information in other studies (which, if available, may or may not be of value or supply useful results) ☐ Inappropriate comparator ☐ Dosing was not done appropriately ☐ Dosing between arms is not equivalent ***Aside**: Lack of placebo or baseline information can lead to misleading conclusions since it may mask the direction of benefit or risk of active agents under comparison.*	
Cross-over Designs	12.	☐ Where subject is serving as own control, might bias be introduced from what happens in one sequence compared to the subsequent sequence — either due to the intervention or due to other factors such as performance or time or nature of a disease? ☐ If the cross-over design includes the patient or physician choosing to crossing over, the outcomes then become "observational." ☐ Where subject is serving as own control, might bias be introduced from loss? Note: Loss is magnified in cross-over studies in that you are losing both the study subject and the control in a single person. **Threats to validity include —** Randomization ☐ No randomization of sequencing Blinding	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		☐ No blinding ☐ Unconcealed concealed cross-over points ☐ Risk of unblinding due to familiarity with intervention or comparator Timing ☐ Lack of pre-specification of reasonable cross-over points ☐ Potential for carry-over effects of intervention or non-intervention elements, disease issues (e.g., considering issues relating to curative potential, disease fluctuations, rebound, seasonal effect, recall bias resulting with familiarity with measurement instruments for prior sequence, etc.) ☐ Potential for treatment being curative, such that cross-over from drug can't be tested Other ☐ Results calculation issues (cross-over studies generally have to be computed by hand if too complex) ☐ Loss (magnified since patient serves as subject and control) ☐ Choice versus assignment to crossover — choice to cross-over renders outcomes "observational"	
Diagnostic Testing	13.	**Key Considerations (Usefulness)** ☐ Does testing lead to improved outcomes or better value? Example: Extrasystoles following an MI indicate higher risk for cardiac mortality, but does treating them reduce mortality risk? ☐ Do the measures of test function appear to be clinically useful? **Measures of Test Function** See **Delfini Diagnostic Testing Calculator** for more information. **Sensitivity (True Positives)/Specificity (True Negatives)** ▪ Typical rates for sensitivity are 80% with a specificity of 90%. Rarely is lower than 50% used. ▪ May be less useful than predictive values because it provides information people already known to have the disease **Predictive Value** ▪ Predictive values (Positive Predictive Value or PPV and Negative Predictive Value or NPV) may be more clinically useful because they are based on predicting disease from test result. **Likelihood Ratios** ▪ LR+ Represents the change from pre-test odds to post-test odds. Increase is considered small if 2-5, modest if 5-10, and large if >10 ▪ LR- Represents the change from pre-test odds to post-test odds. Increase is considered small if .02-.05, modest if .05-.1, and large if > .1 **Threats include —** ☐ New test (index test) is not compared with "gold standard" or reasonable comparator ☐ The test and the gold standard were not applied to all persons studied or to a random sample of all persons studied ☐ Test was not performed in study subjects known to have the condition and those known to be free of the condition ☐ Assessment was not blinded ☐ New test does not find same abnormality as the old test (Overdiagnosis Bias Example: MRI may find earlier and less threatening breast cancers than mammography) ☐ Indeterminate results could have created a bias—if they are not reported, this is a threat to validity ☐ Time in between application of the reference test and index test creates a risk that the diagnosis may not be the same	
Screening	14.	**Threats include —**	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		☐ There is no evidence that early diagnosis and treatment will improve outcomes compared to later diagnosis and treatment (if true consider not applying these results) ☐ There is no assessment of treatments and diagnostic testing ☐ Lead time bias might explain the potential beneficial outcomes **Explanation:** Appearance of increased "survival" time due to early detection— meaning, date of death is not different from what it would be if detection was later, prompted by onset of symptoms. Lead time bias is not an issue when randomization is used to determine who is screened and who is not – but it is an issue in observational studies comparing screened to unscreened. ☐ Length bias might explain the potential beneficial outcomes **Explanation:** Increased selection of slower growing tumors (i.e., missing fast growing tumors) resulting in overestimation of survival time ☐ Volunteer bias might explain potential beneficial outcomes ☐ Overdiagnosis bias **Explanation**: A finding of a disease at an asymptomatic stage in a patient who would not have become symptomatic or harmed by the disease	
Prognosis	15.	☐ Was validation conducted with an independent set of patients? ☐ **Explanation:** Similar results suggest initial sample was representative of the larger population.	
Attrition Bias	16.	☐ Might attrition, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes? Many researchers, biostatisticians and others struggle with this area—there appears to be no clear agreement in the clinical research community about how to best address these issues. There also is inconsistent evidence on the effects of attrition on study results. We, therefore, believe that studies should be evaluated on a case-by-case basis. The key question is, "Given that attrition has occurred, are the study results likely to be true?" It is important to look at the contextual elements of the study and reasons for discontinuation and loss-to-follow up and to look at what data is missing and why to assess likely impact on results. Attrition may or may not impact study outcomes depending, in part, upon the reasons for withdrawals, censoring rules and the resulting effects of applying those rules, for example. In general, we think it is important to attempt to answer the following questions: Examining the contextual elements of a given study— ▪ What could explain the results if it is not the case that the reported findings are true? ▪ What conditions would have to be present for an opposing set of results (equivalence or inferiority) to be true instead of the study findings? ▪ Were those conditions met? ▪ If these conditions were not met, is there any reason to believe that the estimate of effect (size of the difference) between groups is not likely to be true. Differential attrition issues should be looked at especially closely. Unintended differences between groups are more likely to happen when— ▪ Patients have been allocated to their groups in non-blinded fashion (e.g., predictable allocation methods, lack of effective concealment of allocation to study groups), groups are not balanced at the onset of the study, and/or the study is not effectively blinded. Knowing the treatment a patient is receiving can result in such problems as treating patients differently as clinicians have greater beliefs in the new treatment, patients potentially more likely to discontinue because of disappointment that they are	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		not receiving the newest agent, etc. ▪ An effect of the treatment is causing the attrition. This may be informative and may, at times, be informative about the effectiveness of the treatment (especially in an otherwise valid study) or signal a safety issue. **Example of a Contextual Review In a Study with Over 50 Percent Discontinuations** "In this placebo-controlled trial, effective randomization, balance in resulting study groups, concealed allocation to study assignment and blinding in this study mitigate against patients being treated differently due to knowledge of a study assignment. This is borne out by balance in use of co-interventions—patients discontinuing were treated similarly regardless of study group. Groups were otherwise treated identically except for the investigational agents. Adherence was balanced and very high (over 80 percent). Protocol deviations were balanced and very low at less than 90 percent. Censoring rules were assessed at low risk of bias and were designed to prevent double counting. There is no evidence of selective reporting. Balance in discontinuation reasons and numbers is supportive that patients were not discontinuing because of discovery of being on one agent as compared to another. The superior outcomes in the active agent group would have to be explained by superiority of a co-intervention that was disproportionately applied to the active agent group. However, co-interventions were balanced between groups, therefore, it is reasonable to conclude that the results are due to the efficacy of the active agent.	
Assessment Bias	17.	☐ Assessors are **blinded** ☐ **Selective reporting** occurred or is suspected (e.g., key outcomes or prespecified outcomes not reported, reporting is incomplete or based on non-prespecified assessments ☐ Likelihood of findings due to **chance, false positive and false negative outcomes** (judgment call on statistical significance, including **confidence intervals**) ○ For valid studies, consider what you judge to be a reasonable range for clinical significance – this need not be hard and fast ○ Look at the borders of the confidence intervals and interpret their clinical meaning based on your judgment – are the boundaries outside what you deem clinically significant? If yes, then the findings are inconclusive. ○ If the study was **terminated early**, this increases the potential for chance effects especially in studies with fewer than 500 events and which are not confirmed in other studies. ***Example: If a monitoring committee examines the data every X months in a Y year study then the p-value would be (X months / X*Y months) x the study-wide p-value** ☐ **Intention-to-Treat Analysis** performed (all people are analyzed as randomized + reasonable method for imputing missing values which puts the intervention through a challenging trial or other reasonable sensitivity analysis) or missing values are very small. **Caution**: Use of the mean is not appropriate. Last Observation Carried Forward (LOCF) is generally considered a biased method, but, in the case of progressive conditions may at least provide useful data on the direction of the results under certain circumstances: http://www.delfini.org/delfiniClick_PrimaryStudies.htm#LOCFhelp ☐ Use of **modeling** only with use of reasonable assumptions ☐ If **time-to-event analysis** used (see also Kaplan Meier curves, survival curves), was it appropriate to do, is there transparency especially with regard to censoring and is it unbiased. (See http://www.delfini.org/Delfini_Pearls_Analyzing_Results_Time_To_Event_Analysis_KM.pdf.) **Additional threats include —** **Statistical** ☐ No report of statistics used	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

<table>
<tr><th>Category</th><th></th><th>Questions to Evaluate
(A study reporting no details should be considered a threat.)</th><th>Threat</th></tr>
<tr><td></td><td></td><td>☐ In appropriate statistics used. Example: One-sided tests were performed when two-sided tests would be preferable because of the possibility of results going in either direction — one-sided tests may favor the intervention
☐ Modeling used which requires assumptions and is frequently not done correctly. It is not possible to evaluate the assumptions used and/or the correctness of the method sufficiently to detect potential bias in results reporting
☐ Inappropriate reliance on adjustments</td><td></td></tr>
<tr><td>Equivalence or Non-inferiority</td><td>18.</td><td>☐ Efficacy of referent agent is not established through valid and clinically useful studies
☐ Referent study and new study are not sufficiently similar
☐ Biases or analyses diminish potential true differences between groups
☐ Authors did not choose an equivalence range or non-inferiority cutoff (Delta) that seems reasonable (e.g., too wide for equivalence trials and too far for non-inferiority trials)</td><td></td></tr>
<tr><td>Oncology Studies</td><td>19.</td><td>Threats include —

☐ Small study size
☐ Short study duration
• Primary outcome is tumor response not survival
Explanation: Survival has always been accepted as an appropriate outcome measure in oncology studies. FDA also accepts tumor response outcomes even though they may not be associated with improved survival. Below are the most common outcome measures in oncology studies. In making judgments regarding clinical usefulness, our preference is to see –

☐ Overall mortality as the primary outcome measure with next preferred outcome measure
☐ The combined outcome measure of mortality and tumor response followed by
☐ Tumor response outcomes

<table>
<tr><td></td><td></td></tr>
<tr><td>Overall Survival</td><td>Defined as the time from randomization until death from any cause and is measured in the intent-to-treat population</td></tr>
<tr><td>Progression-Free Survival (PFS)</td><td>Defined as the time from randomization until objective tumor progression or death</td></tr>
<tr><td>Disease-Free Survival (DFS)</td><td>Defined as the time from randomization until recurrence of tumor or death from any cause</td></tr>
<tr><td>Objective Response Rate (ORR)</td><td>Defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period</td></tr>
<tr><td>Time-to-Progression (TTP)</td><td>Defined as the time from randomization until objective tumor progression</td></tr>
<tr><td>Time-to-Treatment Failure (TTF)</td><td>Defined as a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including disease progression, treatment toxicity and death</td></tr>
</table>

☐ No information on QOL
☐ No information on physical functioning
☐ No information on tumor-related symptoms</td><td></td></tr>
</table>

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		☐ Single arm study	
Usefulness Assessment ⬇	20.	☐ Clinically significant area + sufficient benefit size = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ If composite endpoints used, reasonable combination used	
Measures of Outcomes	21.	☐ If the study conclusions promise benefit, are the results actually going to be of reasonable benefit? ▪ **Consider** the how big the benefit expected is by looking at the estimate of effect (e.g., Absolute Risk Reduction - ARR, Number Needed to Treat/Harm/Screen/Prevent - NNT (NNH, etc.), Odds Ratio - OR, Relative Risk - RR). Most helpful are NNT and ARR. Less useful are OR and RR. Relative measures can be helpful **only** in combination with absolute measures such as ARR and NNT. ▪ Keep in mind efficacy versus effectiveness – the study circumstances and population used mean that the study results are likely to be bigger than what you will realize in the clinical setting. ***Aside:*** *Other measures may be used as well; however use of relative measures alone should be avoided as it* ***always*** *overestimates benefit. Odds ratios deal with odds, not probabilities are harder to apply since you lose the baseline information.*	
Safety	22.	☐ Assess potential safety issues. ***Aside:*** *Safety assessment is highly complex and frequently necessarily based on weaker data. For help, it is strongly recommended to consult the* **Delfini Grading, Conclusion & Results Table tool** *for cautions and tips. The* **Delfini Searching Tool** *can provide additional help for seeking out other sources to help assess safety.*	
Article Quality	23.	☐ Is this article particularly difficult to read potentially as a result of lack of skills on the part of the author? (If yes, this may point to problems in research skills and may suggest even more bias than what you might be able to discern from the article.) Or is language used that hints of the use of unconventional methods that are not clear in the article (e.g., use of words like "annualized," or "evaluable," or "we defined the ITT population as…") and which might suggest bias or lethal threats?	
Plausibility	24.	☐ Do the conclusions make sense? Consider dose-response relationship, biologic plausibility, etc. **Caution**: often conclusions resulting from reasoning by pathophysiology are erroneous plus plausibility is limited by current thinking and knowledge.)	
Conclusions	25.	☐ Are the author's conclusions actually justified by the results? Is there data integrity? Is data overstated? Are limitations acknowledged? If composite outcomes are reported, then do the outcomes that are chosen for combined reporting reasonable and not misleading (e.g., combinations of subjective and objective outcomes, combinations of severe outcomes with mild ones or process measures)? ***Aside:*** *Author's conclusions are often optimistic. Review data for accuracy*	
External Validity	26.	☐ **How likely are research results to be realized in the real world considering population and circumstances for care**? ☐ Review n, inclusions, exclusions, baseline characteristics and intervention methods — this is a **judgment call**. ***Aside:*** *Baseline characteristics are important to assess who was actually studied. Application to different populations is hugely subject to judgment.*	
Patient Perspective	27.	☐ Consider benefits, harms, risks, costs, uncertainties, alternatives, applicability to which patients, adherence issues and patient satisfaction	

Study Reference:
Study Type: **Study Aim:**
Date: **Evaluator:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		Safety Note: Appropriately defined as all patients receiving intervention; if medication at least one dose of study drug	
Provider Perspective	28.	☐ Satisfaction, acceptability, likely appropriate application and actionability (e.g., FDA approval, affordability, external relevance, circumstances of care, able to apply, tools available)	
Other	29.	☐ Other concerns?	
Summary of all findings from above appraisal	30.		

Next Steps

1. Grade the study or individual conclusions from the study.
2. Record pertinent study results.
3. Prepare a concluding statement about your findings.

Help with each of these steps can be found in the **Delfini Grading, Conclusion & Results Table** tool.

- The last table in this tool can be used for study grading and recording study results which can be copied and pasted here.
- Also included in this tool is a table that can be copied and pasted below the results table to record a concluding statement.