



**Delfini Group**™, LLC



*Evidence- & Value-based Solutions For Health Care*

*Clinical Improvement Consults, Content, Seminars, Training & Tools*

## Evidence Tool Set

# Study Validity & Evidence Usability: Primer & Tool for Primary Studies

**Delfini Group**™, LLC

**Michael Stuart, MD**  
President

**Sheri Strite, Principal &  
Managing Partner**

[www.delfini.org](http://www.delfini.org)

### **Our Mission –**

**To assist medical leaders, clinicians and other health care professionals by ~**

- Bringing science into medical practice in an **easy-to-understand** way.
- Using **simplified methods** to help navigate the complexities of such areas as evidence-based medicine and other topics.
- Building **competencies** and **confidence** in improving medical care through our well received consultations, educational programs and tools.
- Providing inspiration to others to **improve** medical care and help bring about needed change.

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

### Usable Evidence" = Valid + Useful + Usable

You are trying to determine:

1. Is the information true?
2. If yes, is it useful?
3. If yes, is it usable – and how?

This tool, along with the **Delfini Grading, Conclusion & Results Table tool**, can be used to help you evaluate how well a research study meets the criteria for a well-done study in which the evidence is strong and circumstances are relevant to your patients. It can help you pose questions about a study from several perspectives: epidemiologic, clinical and patient perspective.

See the **Delfini Grading, Conclusion & Results Table tool** for help with —

- Grading the evidence including tips for identifying lethal flaws quickly
- Cautions regarding research findings generally and specific to efficacy and safety
- Reporting and assessing results
- Crafting conclusions

This tool **cannot and is not meant** to provide all the information you need to evaluate a study; however, it can help you identify many studies that have significant threats to validity or problems with usability. For more in-depth analysis, it is recommended that you consult an epidemiologist or biostatistician – especially for evaluation of appropriate statistical methods, which this tool does not address. Ultimately you will need to apply **your own judgment** in determining whether a study can be useful to you or not. This tool can give you key ingredients to help you make that decision.

Apply the validity questions to your article of interest. The Appendix provides helpful critical appraisal tips.

### Match Your Clinical Question to Study Type

#### For Primary Studies of Therapy or Prevention

Experiments, preferably Randomized Controlled Trials (RCTs)

#### For Primary Studies of Natural History or Prognosis

Observations, preferably Cohort designs

#### For Primary Studies of a Diagnostic Test

Experiments or cross-sectional studies, preferably RCTs

#### For Primary Studies of Screening or Case Finding

Experiments, preferably Randomized Controlled Trials (RCTs)

You must assess if early diagnosis and treatment will improve outcomes more than later diagnosis and treatment. You need to review studies of treatment **and** diagnosis.

#### Important Note

- **GRADE UA: Author's Conclusions are Uncertain**

If the author's conclusions are uncertain, it may not be necessary to do a critical appraisal of the study because an uncertain outcome automatically renders an "uncertain" grade designation.

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
General	1.	<input type="checkbox"/> <b>Who is sponsoring and funding the study? What are the affiliations of the authors?</b>  <b>Considerations:</b> 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.	<input type="checkbox"/> <b>Is the design appropriate to the research question?</b>  <input type="checkbox"/> <b>Is the research question useful?</b>  <b>Considerations:</b> <ul style="list-style-type: none"> <li><input type="checkbox"/> For questions of <b>efficacy</b>, is this an <b>experimental study design</b> (meaning there was <b>no choice</b> made to determine intervention)</li> <li><input type="checkbox"/> <b>Clinically significant area</b> for study (morbidity, mortality, symptom relief, emotional functioning, physical functioning or health-related quality of life)</li> <li><input type="checkbox"/> Reasonable <b>definitions for clinical outcome such as response, treatment success or failure</b></li> <li><input type="checkbox"/> Are <b>intermediate outcome markers</b> used? If yes, is there valid and useful evidence that proves that the intermediate marker is valid for clinical significance?</li> </ul>	
Internal Validity Assessment ↓	3.	<input type="checkbox"/> <b>Can bias, confounding or chance explain the study results?</b>  Ensure <b>prespecified (a priori) and appropriate</b> — <ul style="list-style-type: none"> <li><input type="checkbox"/> research questions,</li> <li><input type="checkbox"/> populations to analyze,</li> <li><input type="checkbox"/> outcomes (see below for <b>composite outcomes</b>),</li> </ul> <b>Threats for Composite Outcomes include</b> — <ul style="list-style-type: none"> <li><input type="checkbox"/> Composite outcomes are not appropriate</li> <li><input type="checkbox"/> Outcomes are or may be double-counted</li> <li><input type="checkbox"/> Outcomes are not correlated</li> <li><input type="checkbox"/> Less important outcome – or outcome subject to external control – is driving the results</li> <li><input type="checkbox"/> It is not possible to analyze data in the composite endpoints to tell what is driving the measures of outcomes</li> <li><input type="checkbox"/> Combination is of severe outcomes with mild ones or with process measures</li> <li><input type="checkbox"/> Outcomes are not independent of other variables or are subjective or under external control, choice or influence</li> </ul> <p><i><b>Aside:</b> If several of the outcomes are biologically related, this could lend support that these are true outcomes and not a result of chance findings</i></p>	
Confounding	4.	<input type="checkbox"/> <b>Absence of known or unknown confounders</b>  Might there be another factor associated with the study variable that may account for the	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

**Date:**    **Study Reference:**  
**Reviewer:**  
**Study Type:**  
**Study Aim:**

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<p>conclusion <b>instead or in addition to</b> of the variable under study? (This would be a <b>confounder</b>.)</p> <p>If yes, have the researchers identified potential confounders and adjusted for them? <b>Caution:</b> Adjustment may not eliminate this problem. This is best addressed through successful <b>randomization</b>.</p> <p><i>Aside: Confounding is difficult to assess. Most confounders are unknown.</i></p>	
<b>Selection Bias</b>	<p>5. <b>Study Groups</b> are —</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> appropriate for study</li> <li><input type="checkbox"/> of appropriate size</li> <li><input type="checkbox"/> concurrent</li> <li><input type="checkbox"/> similar baseline characteristics (e.g., demographics and prognostic factors) <ul style="list-style-type: none"> <li><input type="checkbox"/> <b>Observations:</b> similar as possible at the study's inception, except for the study topic</li> <li><input type="checkbox"/> If no, how robust are the findings — are the results similar when the populations are adjusted and unadjusted?</li> </ul> </li> </ul> <p><b>Review</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> How subjects were identified and selected for study,</li> <li><input type="checkbox"/> Demographics, inclusions &amp; exclusions, baseline characteristics</li> </ul> <p><b>Considerations</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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.</li> <li><input type="checkbox"/> Is the sample size large enough to be representative? (Rule of thumb: less than 100 in each group can be considered small – medium.)</li> <li><input type="checkbox"/> Was the sample chosen for study representative of the defined population at risk or typical patients?</li> <li><input type="checkbox"/> Are all patients accounted for in the report of baseline characteristics?</li> <li><input type="checkbox"/> 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.</li> <li><input type="checkbox"/> 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?</li> <li><input type="checkbox"/> Non-responders of placebo being excluded can bias study results in favor of the intervention</li> <li><input type="checkbox"/> Non-responders of the study drug — but not the comparator — being excluded can bias study results in favor of the intervention</li> <li><input type="checkbox"/> 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</li> </ul> <p><b>Considerations for Observational Studies</b></p>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

**Date:**    **Study Reference:**  
**Reviewer:**  
**Study Type:**  
**Study Aim:**

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<input type="checkbox"/> Is the control group a logical one for study and selected in such a way to avoid bias?  <b>Asides:</b> Selection bias is often a significant issue, especially in observational studies. Any difference between groups could actually be the true cause of the outcomes.  <i>Baseline characteristics are important to assess who was actually studied</i>	
<b>Randomization</b>	6. <input type="checkbox"/> Randomization methods for sequence generation are truly <b>random</b> and sequencing avoids potential for anyone <b>affecting assignment</b> to a study arm  <input type="checkbox"/> 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.  <b>Adequate</b> <input type="checkbox"/> Computer generated random numbers <input type="checkbox"/> Urn randomization <input type="checkbox"/> Random number tables (not open) <input type="checkbox"/> Drawing lots <input type="checkbox"/> If block assignment utilized, random sequence generation method plus use of permuted blocks (i.e., randomly chosen block sizes that vary)  <b>Threats include —</b> <input type="checkbox"/> Authors state study was randomized, however no details of randomization reported <input type="checkbox"/> 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.) <input type="checkbox"/> Block size may have been known to investigators prior to assignment – problematic if concealment of allocation is not adequate or if stratified by site <input type="checkbox"/> Block size equal to the number of arms. <input type="checkbox"/> Threat is increased if concealment of allocation is not adequate	
<b>Randomization Concealment of the Allocation Sequence</b>	7. <input type="checkbox"/> <b>Concealment of allocation</b> strategies are employed to prevent anyone affecting assignment to a study arm  <ul style="list-style-type: none"> <li>▪ 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.</li> </ul> <b>Adequate</b> 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 —  <input type="checkbox"/> Training provided to staff on concealment and its import	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<ul style="list-style-type: none"> <li><input type="checkbox"/> Central randomization such a call-in center with enrollment performed prior to group assignment and just-in-time per enrolled patient</li> <li><input type="checkbox"/> 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</li> <li><input type="checkbox"/> Other convincing method</li> </ul> <p><b>Possibly Adequate</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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.</li> </ul> <p><b>Threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> No details of concealment reported</li> <li><input type="checkbox"/> Inadequate methods used</li> <li><input type="checkbox"/> Same person administers and without other methods of protection against concealment</li> <li><input type="checkbox"/> 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</li> <li><input type="checkbox"/> Deterministic, predictable or open methods of randomization (see Randomization)</li> </ul> <p><i>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</i></p>	
<b>Observation Bias</b> ↓	8. <input type="checkbox"/> <b>Might bias have been introduced during study performance, follow-up or assessment?</b>	
<b>Performance Bias</b> ↓	9. <input type="checkbox"/> <b>No bias or difference, except for what is under study, between groups during course of study</b>	
	<p><b>Such as —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> intervention design and execution (example – if this were a surgery study, were the surgeons skillful?)</li> <li><input type="checkbox"/> co-interventions</li> <li><input type="checkbox"/> 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-reported use given the population)</li> <li><input type="checkbox"/> 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)</li> <li><input type="checkbox"/> inappropriate exposure or migration (see below in this section)</li> <li><input type="checkbox"/> cross-over threats (see specific questions in <b>cross-over design</b> section)</li> <li><input type="checkbox"/> protocol deviations</li> <li><input type="checkbox"/> measurement methods (see below in this section)</li> <li><input type="checkbox"/> study duration (consider length of time needed for occurrence of primary effect, side effects,</li> </ul>	

Use of this tool implies agreement to the legal terms and conditions on the Delfini website.

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<p>symptoms, regression, remission, recurrence, survival, duration of treatment, duration of follow-up, etc.)</p> <p><input type="checkbox"/> other?</p> <p><b>Timing</b></p> <p><input type="checkbox"/> In the course of the study, could passing of time have had a resulting impact on any outcomes? Examples —</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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.)</li> <li><input type="checkbox"/> Might the time period of the study affect study results (eg, studying an allergy medication, but not during allergy season)?</li> <li><input type="checkbox"/> Could awareness of adverse events affect future reporting of adverse events?</li> <li><input type="checkbox"/> 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.</li> <li><input type="checkbox"/> 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)?</li> <li><input type="checkbox"/> 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)?</li> </ul> <p><b>Considerations for Migration or Exposure</b></p> <p><b>RCTs: Migration to Other Study Group</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Review run-in period and wash-out periods</li> <li><input type="checkbox"/> Choice of treatment made renders those outcomes observational</li> </ul> <p><b>Observations: Exposure</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Were measures taken to ensure that those in the comparison group weren't "exposed?"</li> <li><input type="checkbox"/> Was exposure determined by reasonable means and did it precede the outcome of interest?</li> </ul> <p><b>Considerations for Measurement Methods</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Consider need to critically appraise study reporting on effectiveness of tool. (Validated does not necessarily mean that the measurement tool is valid.)</li> <li><input type="checkbox"/> Use of tools that are not standard should be scrutinized.</li> <li><input type="checkbox"/> Evaluate likelihood of detecting meaningful clinical significance.</li> </ul> <p><b>Threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Measurement instruments or methods do not appear appropriate due to problems with — <ul style="list-style-type: none"> <li><input type="checkbox"/> Methods of measurement</li> <li><input type="checkbox"/> Sources for data collection</li> <li><input type="checkbox"/> Methods of data collection</li> <li><input type="checkbox"/> Not sufficiently comprehensive</li> </ul> </li> </ul>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<p><b>Measurement Tip:</b> 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</p>	
<b>Blinding</b>	<p>10. <input type="checkbox"/> <b>Double-blinding</b> methods employed (i.e., subject and all working with the subject or subject’s data) and achieved</p> <p><b>Threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Subjects were not blinded</li> <li><input type="checkbox"/> All those assigning treatment, providing care, performing the intervention or otherwise working with subjects blinded</li> <li><input type="checkbox"/> Intervention was not effectively disguised from the comparator</li> <li><input type="checkbox"/> Blind likely to have been broken</li> <li><input type="checkbox"/> 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)</li> <li><input type="checkbox"/> Exposure issues — Example: Titration methods with a placebo arm or patients starting with a placebo inhaler would have discerned which was the active medication.</li> <li><input type="checkbox"/> Outcomes could be affected by knowledge of treatment through subjective measurements used or influence or control of patient or provider</li> </ul> <p><b>Asides:</b> 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.</p>	
<b>Comparators</b>	<p>11. <input type="checkbox"/> Reasonable <b>intervention</b> and reasonable <b>comparator</b> used (e.g., placebo)</p> <p><b>Considerations</b></p> <ul style="list-style-type: none"> <li>▪ 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.</li> <li>▪ 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?</li> </ul> <p><b>Additional threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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)</li> <li><input type="checkbox"/> Inappropriate comparator</li> <li><input type="checkbox"/> Dosing was not done appropriately</li> <li><input type="checkbox"/> Dosing between arms is not equivalent</li> </ul> <p><b>Aside:</b> 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.</p>	
<b>Cross-over Designs</b>	<p>12. <input type="checkbox"/> 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?</p>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<ul style="list-style-type: none"> <li><input type="checkbox"/> If the cross-over design includes the patient or physician choosing to crossing over, the outcomes then become "observational."</li> <li><input type="checkbox"/> 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.</li> </ul> <p><b>Threats to validity include —</b></p> <p>Randomization</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> No randomization of sequencing</li> </ul> <p>Blinding</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> No blinding</li> <li><input type="checkbox"/> Unconcealed concealed cross-over points</li> <li><input type="checkbox"/> Risk of unblinding due to familiarity with intervention or comparator</li> </ul> <p>Timing</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Lack of pre-specification of reasonable cross-over points</li> <li><input type="checkbox"/> 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.)</li> <li><input type="checkbox"/> Potential for treatment being curative, such that cross-over from drug can't be tested</li> </ul> <p>Other</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Results calculation issues (cross-over studies generally have to be computed by hand if too complex)</li> <li><input type="checkbox"/> Loss (magnified since patient serves as subject and control)</li> <li><input type="checkbox"/> Choice versus assignment to crossover — choice to cross-over renders outcomes "observational"</li> </ul>	
<b>Diagnostic Testing</b>	<p>13. <b>Key Considerations (Usefulness)</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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?</li> <li><input type="checkbox"/> Do the measures of test function appear to be clinically useful?</li> </ul> <p><b>Measures of Test Function</b> See <b>Delfini Diagnostic Testing Calculator</b> for more information.</p> <p><b>Sensitivity (True Positives)/Specificity (True Negatives)</b></p> <ul style="list-style-type: none"> <li>▪ Typical rates for sensitivity are 80% with a specificity of 90%. Rarely is lower than 50% used.</li> <li>▪ May be less useful than predictive values because it provides information people already known to have the disease</li> </ul> <p><b>Predictive Value</b></p> <ul style="list-style-type: none"> <li>▪ 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.</li> </ul> <p><b>Likelihood Ratios</b></p>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<ul style="list-style-type: none"> <li>▪ 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 &gt;10</li> <li>▪ 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 &gt; .1</li> </ul> <p><b>Threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> New test (index test) is not compared with "gold standard" or reasonable comparator</li> <li><input type="checkbox"/> The test and the gold standard were not applied to all persons studied or to a random sample of all persons studied</li> <li><input type="checkbox"/> Test was not performed in study subjects known to have the condition and those known to be free of the condition</li> <li><input type="checkbox"/> Assessment was not blinded</li> <li><input type="checkbox"/> 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)</li> <li><input type="checkbox"/> Indeterminate results could have created a bias—if they are not reported, this is a threat to validity</li> <li><input type="checkbox"/> Time in between application of the reference test and index test creates a risk that the diagnosis may not be the same</li> </ul>	
<b>Screening</b>	<p>14. <b>Threats include —</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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)</li> <li><input type="checkbox"/> There is no assessment of treatments and diagnostic testing</li> <li><input type="checkbox"/> Lead time bias might explain the potential beneficial outcomes <b>Explanation:</b> 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.</li> <li><input type="checkbox"/> Length bias might explain the potential beneficial outcomes <b>Explanation:</b> Increased selection of slower growing tumors (i.e., missing fast growing tumors) resulting in overestimation of survival time</li> <li><input type="checkbox"/> Volunteer bias might explain potential beneficial outcomes</li> <li><input type="checkbox"/> Overdiagnosis bias <b>Explanation:</b> A finding of a disease at an asymptomatic stage in a patient who would not have become symptomatic or harmed by the disease</li> </ul>	
<b>Prognosis</b>	<p>15. <input type="checkbox"/> Was validation conducted with an independent set of patients? <input type="checkbox"/> <b>Explanation:</b> Similar results suggest initial sample was representative of the larger population.</p>	
<b>Attrition Bias</b>	<p>16. <input type="checkbox"/> Is there sufficient patient drop-out or missing data points that the study's validity is threatened? Missing data is one issue. Another key issue is whether an imbalance between groups has resulted or whether there is an imbalance within groups as randomized as compared to completers. If there is no imbalance between and within various study subgroups (as randomized compared to completers, patients lost to follow-up, etc), then</p>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
	<p>this may present no threat, except in instances in which statistical significance is not reached because it may be that not enough people remained to show a statistically significant difference if one exists. Baseline comparisons of various subgroups is advised where possible.</p> <p><b>Consider the following (consider percents to be approximations) —</b></p> <ul style="list-style-type: none"> <li>▪ Likely minimal threat: &lt;5% and no differential loss* (e.g., 1% vs 4%) (based on known instances where loss greater than 5% results in large P-value changes).</li> <li>▪ Possible threat: &gt;=5% but &lt;10% and no differential loss*</li> <li>▪ Threat in most cases: &gt;=5% with differential loss*, or &gt;= 10% without differential loss, and without worst-case sensitivity analysis, or otherwise reasonable sensitivity analysis, conducted by authors or reviewers, in which statistical significance is confirmed</li> </ul>	
<p><b>Assessment Bias</b></p>	<p>17.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Assessors are <b>blinded</b></li> <li><input type="checkbox"/> <b>Selective reporting</b> occurred or is suspected (e.g., key outcomes or prespecified outcomes not reported, reporting is incomplete or based on non-prespecified assessments)</li> <li><input type="checkbox"/> Likelihood of findings due to <b>chance, false positive and false negative outcomes</b> (judgment call on statistical significance, including <b>confidence intervals</b>) <ul style="list-style-type: none"> <li>○ For valid studies, consider what you judge to be a reasonable range for clinical significance – this need not be hard and fast</li> <li>○ 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.</li> <li>○ If the study was <b>terminated early</b>, this increases the potential for chance effects especially in studies with fewer than 200 events and which are not confirmed in other studies. <b>*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</b></li> </ul> </li> <li><input type="checkbox"/> <b>Intention-to-Treat Analysis</b> 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) <b>Caution:</b> use of last observation carried forward or use of the mean is biased</li> <li><input type="checkbox"/> Use of <b>modeling</b> only with use of reasonable assumptions</li> <li><input type="checkbox"/> If survival curves were used, was double-counting avoided and is it expected that the rate would otherwise not be constant between the groups (e.g., death would occur more early in a surgically treated group as compared to a medically treated group, and therefore, bias the curve)?</li> </ul> <p><b>Additional threats include —</b></p> <p><b>Statistical</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> No report of statistics used</li> <li><input type="checkbox"/> 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</li> <li><input type="checkbox"/> Modeling used which requires assumptions and is frequently not done correctly. It is not</li> </ul>	

## Delfini Group Study Validity & Evidence Usability: Primary Sources

**Date:**    **Study Reference:**  
**Reviewer:**  
**Study Type:**  
**Study Aim:**

Category	Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat										
	possible to evaluate the assumptions used and/or the correctness of the method sufficiently to detect potential bias in results reporting <input type="checkbox"/> Inappropriate reliance on adjustments											
<b>Equivalence or Non-inferiority</b>	18. <input type="checkbox"/> Efficacy of referent agent is not established through valid and clinically useful studies <input type="checkbox"/> Referent study and new study are not sufficiently similar <input type="checkbox"/> Biases or analyses diminish potential true differences between groups <input type="checkbox"/> 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)											
<b>Oncology Studies</b>	19. <b>Threats include —</b> <input type="checkbox"/> Small study size <input type="checkbox"/> Short study duration <ul style="list-style-type: none"><li>• Primary outcome is tumor response not survival <b>Explanation:</b> 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 – <input type="checkbox"/> Overall mortality as the primary outcome measure with next preferred outcome measure <input type="checkbox"/> The combined outcome measure of mortality and tumor response followed by <input type="checkbox"/> Tumor response outcomes</li></ul> <table border="1" data-bbox="397 1276 1398 1879"> <tbody> <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> </tbody> </table>	Overall Survival	Defined as the time from randomization until death from any cause and is measured in the intent-to-treat population	Progression-Free Survival (PFS)	Defined as the time from randomization until objective tumor progression or death	Disease-Free Survival (DFS)	Defined as the time from randomization until recurrence of tumor or death from any cause	Objective Response Rate (ORR)	Defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period	Time-to-Progression (TTP)	Defined as the time from randomization until objective tumor progression	
Overall Survival	Defined as the time from randomization until death from any cause and is measured in the intent-to-treat population											
Progression-Free Survival (PFS)	Defined as the time from randomization until objective tumor progression or death											
Disease-Free Survival (DFS)	Defined as the time from randomization until recurrence of tumor or death from any cause											
Objective Response Rate (ORR)	Defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period											
Time-to-Progression (TTP)	Defined as the time from randomization until objective tumor progression											

## Delfini Group Study Validity & Evidence Usability: Primary Sources

Date: Study Reference:

Reviewer:

Study Type:

Study Aim:

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
		<p>Time-to-Treatment Failure (TTF) Defined as a composite endpoint measuring time from randomization to discontinuation of treatment for any reason, including disease progression, treatment toxicity and death</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> No information on QOL</li> <li><input type="checkbox"/> No information on physical functioning</li> <li><input type="checkbox"/> No information on tumor-related symptoms</li> <li><input type="checkbox"/> Single arm study</li> </ul>	
<b>Usefulness Assessment</b>	20.	<ul style="list-style-type: none"> <li><input type="checkbox"/> Clinically significant area + sufficient benefit size = meaningful clinical benefit (consider efficacy vs effectiveness)</li> <li><input type="checkbox"/> If composite endpoints used, reasonable combination used — and used for safety if used for efficacy</li> </ul>	
<b>Measures of Outcomes</b>	21.	<ul style="list-style-type: none"> <li><input type="checkbox"/> If the study conclusions promise benefit, are the results actually going to be of reasonable benefit? <ul style="list-style-type: none"> <li>▪ <b>Consider</b> 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 <b>only</b> in combination with absolute measures such as ARR and NNT.</li> <li>▪ 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.</li> </ul> </li> </ul> <p><i>Aside: Other measures may be used as well; however use of relative measures alone should be avoided as it <b>always</b> overestimates benefit. Odds ratios deal with odds, not probabilities are harder to apply since you lose the baseline information.</i></p>	
<b>Safety</b>	22.	<ul style="list-style-type: none"> <li><input type="checkbox"/> Assess potential safety issues.</li> </ul> <p><i>Aside: Safety assessment is highly complex and frequently necessarily based on weaker data. For help, it is strongly recommended to consult the <b>Delfini Grading, Conclusion &amp; Results Table tool</b> for cautions and tips. The <b>Delfini Searching Tool</b> can provide additional help for seeking out other sources to help assess safety.</i></p>	
<b>Article Quality</b>	23.	<ul style="list-style-type: none"> <li><input type="checkbox"/> 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?</li> </ul>	
<b>Plausibility</b>	24.	<ul style="list-style-type: none"> <li><input type="checkbox"/> Do the conclusions make sense? Consider dose-response relationship, biologic plausibility, etc.</li> </ul> <p><b>Caution:</b> often conclusions resulting from reasoning by pathophysiology are erroneous plus plausibility is limited by current thinking and knowledge.)</p>	

Use of this tool implies agreement to the legal terms and conditions on the Delfini website.

## Delfini Group Study Validity & Evidence Usability: Primary Sources

**Date:**    **Study Reference:**  
**Reviewer:**  
**Study Type:**  
**Study Aim:**

Category		Questions to Evaluate (A study reporting no details should be considered a threat.)	Threat
<b>Conclusions</b>	25.	<input type="checkbox"/> 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)?  <i>Aside: Author's conclusions are often optimistic. Review data for accuracy</i>	
<b>External Validity</b>	26.	<input type="checkbox"/> <b>How likely are research results to be realized in the real world considering population and circumstances for care?</b> <input type="checkbox"/> Review n, inclusions, exclusions, baseline characteristics and intervention methods — this is a <b>judgment call</b> .  <i>Aside: Baseline characteristics are important to assess who was actually studied. Application to different populations is hugely subject to judgment.</i>	
<b>Patient Perspective</b>	27.	<input type="checkbox"/> Consider benefits, harms, risks, costs, uncertainties, alternatives, applicability to which patients, adherence issues and patient satisfaction  <b>Safety Note:</b> Appropriately defined as all patients receiving intervention; if medication at least one dose of study drug	
<b>Provider Perspective</b>	28.	<input type="checkbox"/> Satisfaction, acceptability, likely appropriate application and actionability (e.g., FDA approval, affordability, external relevance, circumstances of care, able to apply, tools available)	
<b>Other</b>	29.	<input type="checkbox"/> Other concerns?	
<b>Summary of all findings from above appraisal</b>	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.