Critical Appraisal of the Medical Literature:
A Foundation for
Healthcare Decision-making

Training Packet for
Understanding Medical Literature—
A Simplified Approach

Version 20150607

Critical Appraisal of the Medical Literature Training Program: A Foundation for Healthcare Decision-making

Contents

Page	Items	Description
2	Contents	
3	Bios	
	1. SUMMARIES	
4	Critical Appraisal Matters: Quick Facts	
5	Evaluating Therapies (Superiority)	Core elements for appraising superiority of therapies
6	Evidence Grading	Summarizing evidence assessments
7	Intention-to-Treat Analysis (ITT)	Intention-to-Treat Analysis (ITT)
8	Time-to-Event Analysis	
10	Results	
11	Forest Plots	
12	Safety	
13	Composite Endpoints	
14	Oncology Outcomes	
15	Screening	
16	Diagnostic Testing	
17	Cross-Over Designs	
18	Correlation Analysis	Univariate, multivariable and multivariate analysis
19	Comparative Study Designs	Evaluating non-inferiority and equivalence trials
20	Secondary Studies	Overviews, systematic reviews including meta-analysis
21	Secondary Sources	Issues with sources utilizing primary or secondary studies
22	Secondary Resource Audit Advice	Auditing secondary studies and sources
23	Real World Data	
24	Overview of Evidence Synthesis	
	2. TOOLS	
25	Primary Studies Critical Appraisal Checklist	Checklist: Short critical appraisal checklist
26	Secondary Studies Critical Appraisal Checklist	
31	Directions to Delfini Tools	Download from http://www.delfini.org/delfiniTools.htm
	3. EXERCISES	
32	Myoceptimab and MRP: CV Morbidity	SHOW & TELL Sample Critical Appraisal Report
33	Sample Critical Appraisal Using Template	SHOW & TELL Sample Critical Using Delfini Short Validity 1-pager
36	Yoroceptimab and SG	
38	Myoceptimab: Angina Pectoris ↓	
42	Time-to-Event + Attrition: Myoceptimab	
45	Mini-cases in Critical Appraisal	Abstracts VIGOR 45 · Health Econ 46 · Screening 47
48	Fibromyalgia ↓	
54	Labrie Prostate Cancer Screening	
58	Jupiter Trial	
59	Cryoablation Case	
60	Acupuncture Case	
61	ITT Re-analysis	
63	**Primary Studies Critical Appraisal Checklist**	**EXTRA** (ODD PAGE FOR REMOVING)
65	Critical Appraisal Exercise Worksheets	**EXTRA** (ODD PAGE FOR REMOVING)
69	Book Order Form	(ODD PAGE FOR REMOVING)

About Delfini

Michael Stuart, MD
Co-founder, President & Medical Director
6831 31st Ave NE . Seattle WA 98115
Home Office: 206-527-6146 (call first to fax)
mstuart@delfini.org

Sheri Strite
Co-founder, Principal & Managing Partner
3961 NE 10th Ave . Portland OR 97212-1222
Home Office: 503-288-5154
sstrite@delfini.org

Evidence- and Value-based Solutions for Health Care

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

Delfini Group is a public service entrepreneurship founded to advance applied evidence- and value-based clinical quality improvements and methods through practice, training and facilitation. Much of Delfini's work is dedicated to help solve the little known societal problem of medical misinformation.

Authors of ***Basics for Evaluating Medical Research Studies: A Simplified Approach (And Why Your Patients Need You to Know This)*** **and more—see www.delfinigrouppublishing.com**

Michael Stuart MD & Sheri Ann Strite are medical information scientists, evidologists and clinical improvement experts who combine academic and practical experience to train people how to evaluate medical research studies, conduct evidence reviews, help health care systems apply evidence- and value-based clinical quality improvement methods including special help for work groups such as clinical guideline development teams, pharmacy & therapeutics and medical technology assessment committees, clinical quality improvement teams and more. They also train physicians and others in communicating with patients.

Mike Stuart MD, Co-founder, President & Medical Director

- Family physician and clinical faculty at University of Washington
- Former Director of the Department of Clinical Improvement and Education at Group Health Cooperative in Seattle, Washington, where he led development of more than 35 evidence-based clinical guidelines and other clinical improvements, chaired the Pharmacy & Therapeutics and Medical Technology Assessment Committees
- His work has received praise from prominent health care leaders such as David Eddy MD, Don Berwick MD, Health Ministry of New Zealand and the US Navy Bureau of Medicine.

Sheri Strite, Co-founder, Principal & Managing Partner

- Initiated many Delfini health care improvement strategies, tools and training programs including their popular critical appraisal training program
- Former Associate Director, Program Development, University of California, San Diego, (UCSD) Family & Preventive Medicine, School of Medicine, where she taught faculty, physicians, residents, and medical and pharmacy students
- Past member of the UCSD Family Medicine Research Leaders and faculty for their Research Fellowship in the Department of Family & Preventive Medicine
- Prior to UCSD, Ms. Strite worked in clinical improvement, education and research at Group Health Cooperative in Seattle, Washington, where she held various positions including leadership and research management and administration.

Topics upon which Delfini has written and taught include critical appraisal of medical literature, evidence-based committee processes, health care content development, technology assessment, population-based care, projecting economic and health outcomes, performance measurement, patient decision-making, facilitating provider behavior change, physician/patient communications, developing and implementing clinical guidelines, and creating information, decision and action aids for clinical care.

The Delfini Mission

To improve health care quality and use of resources by assisting medical leaders, health care professionals and others interested in and affected by health care decisions 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, clinical improvement and other topics.
- Building competencies and confidence in improving medical care through our consultations, training programs and tools.
- Providing inspiration to others to improve medical care and help bring about needed change.

Also—

- Textbook contributors
- Website featured on Oxford's CEBM: Centre for Evidence Based Medicine
- Editorial Board, DynaMed
- Pharmacy & therapeutics and medical technology assessment committee membership experience
- Featured in bestseller, *Overtreated: Why Too Much Medicine is Making Us Sicker and Poorer,* by Shannon Brownlee

Delfini Pearls
Critical Appraisal Matters: Quick Facts

References available at http://www.delfini.org/delfiniFactsCriticalAppraisal.htm.

1. As of December 10, 2012 at least 37 deaths have been linked to fungal meningitis thought to be caused by contaminated epidural steroids, and 590 cases in 19 states have been identified with a clinical picture consistent with fungal infection. This may be yet one more example of healthcare professionals basing decisions on poor quality evidence and intervening with unproven—yet potentially risky treatments. **Issues: epidural steroids have been used for more than 50 years to treat low back pain and sciatica and are the most common intervention in pain clinics throughout the world. And yet, despite their widespread use, their efficacy remains unproven.**

2. Lack of critical appraisal skills by physicians and a **misreading of one paragraph** in an abstract about a Vioxx study may have contributed to 27,785 heart attacks and sudden cardiac deaths between 1999 and 2003. To put this in perspective, roughly 58,000 US lives were lost in the Vietnam War. **Issues: absolute versus relative risk reduction; insufficient critical appraisal skills to detect potential spin**

3. Lack of critical appraisal skills by physicians resulted in roughly 63,000 preventable deaths were due to encainide/flecainide for premature ventricular contractions (PVCs) after acute myocardial infarction. **Issues: intermediate markers; observational design**

4. Roughly 42,000 women with advanced breast cancer were subjected to treatment with autologous bone marrow transplant and high dose chemotherapy. It is estimated that over 9,000 died from treatment. Yet, RCTs showed no benefit. Costs have been estimated at $3.4 billion. **Issues: observational design**

5. Leading experts estimate that 20 to 50 percent of all healthcare in the United States is inappropriate.

6. Training in medical schools and other schools for allied health professionals in the United States is shockingly poor when it comes to training in science. This affects the quality of medical research and the quality of medical care. **Roughly 70 percent of physicians and clinical pharmacists fail our basic pre-test**.

7. We have long estimated that **less than 10 percent of all medical research—regardless of source—is reliable or clinical useful**. Others agree. Professor John Ioannidis "...charges that as much as 90 percent of the published medical information that doctors rely on is flawed." In one survey of 60,352 studies, a meager 7% passed criteria of high quality methods and clinical relevancy, and fewer than 5% passed a validity screening for an evidence-based journal.

8. FDA approval is not sufficient for establishing scientific validity and usefulness. We know of no fully "trustable" healthcare information sources, and sources that claim to be "evidence-based" frequently are not. Some of the best and "most trusted" sources have frequently failed our critical appraisal audits. Most secondary sources are based on invalid studies or studies that do not have clinically meaningful outcomes. This includes reviews, meta-analyses, performance measures, compendia, clinical recommendations, health care economic studies, disease management protocols and more. Clinical guidelines vary in quality and the majority may be invalid, including many from professional societies.

9. **Bias in studies tends to favor the intervention under investigation**. Certain kinds of bias have been shown to distort research results up to a relative 50 percent or more—for each flaw.

10. Most physicians rely on abstracts which are frequently inaccurate. **One study found that 18-68 percent of abstracts in 6 top-tier medical journals contained information not verifiable in the body of the article**. One study concluded that there may be considerable bias in p-values reported in abstracts. Physicians and others who understand critical appraisal know it cannot be determined whether a study is valid by reading the abstract.

11. Physicians and others who do not understand issues with **findings that are not statistically significant frequently mistakenly interpret these findings as meaning there is no meaningful difference between the groups**. Those with critical appraisal skills understand how to use confidence intervals to avoid these erroneous interpretations.

12. Key skills required to critically appraise the medical literature are not difficult to learn. We believe all healthcare professionals should be competent in evaluating primary and secondary studies and secondary sources.

Delfini Pearls
Basics of Evaluating Evidence in Superiority Trials for Therapies

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Quick Assessment

If the results are reliable, are they useful and usable? Would they change your practice? Do they apply to your situation considering your patients and circumstances of care? Consider effects on your patients including benefits, harms, risks, costs, uncertainties, alternatives, applicability, satisfaction, abuse and dependency issues. Consider conflicts of interest.

1. Are the results in **clinically significant areas** (morbidity, mortality, symptom relief, emotional/physical functioning and health-related quality of life)? If not, is there a reliable causal chain of evidence to support use of an intermediate marker?
2. Were research questions, outcomes and populations for analyses determined **in advance**?
3. Are **definitions of outcomes** such as success/failure, improvement/no improvement, etc. reasonable?
4. Are the **confidence intervals** wholly inclusive of clinical benefit? If **non-significant**, are the confidence intervals wholly exclusive of clinical benefit?
5. Is this a **new intervention**? If yes, safety is likely to be unknown.

Study Design Considerations for Usability

1. **Randomized controlled trials** (RCTs) for efficacy and safety (tip: **choice of intervention was not made** by patient or patient's physician or by other means that would render study observational)
2. Possibly observation studies with **all-or-none results** (very rare)
3. Observational studies for **safety** if lacking quality information from RCTs

Validity Considerations to Assess Potential Distortion of Results Due to Bias, Confounding or Chance
Assess methodologic details and outcomes in the 4 Phases of a Study

I. Selection of Subjects

1. **Sufficient number of participants**
2. **Random** allocation of study subjects to their groups (minimization may be acceptable)
3. Adequate methods for blinding the allocation of subjects to their groups (aka "**concealment of allocation**")
4. **Balanced distribution of prognostic variables** as assessed through review of baseline characteristics

II. Performance

1. **Comparisons** are reasonable
2. **Execution** is successful, **adherence** was achieved, **duration** of treatment is reasonable
3. Everything is the **same between the groups** except for the subject of interest (e.g., groups are **concurrent** and **balanced**, use of **co-interventions** is the same, same **care experiences**, **adherence** is balanced, **protocol deviations** are balanced, etc.) and no bias is present affecting the groups as a whole (e.g., measurement problems, changes due to time, etc.)
4. **Blinding** of subjects and all working with subjects and their data was performed and success was likely

III. Data

1. Are **measurement methods** valid and the same between groups? "Validated" may not really be valid. Consider **duration** of treatment and follow-up.
2. Could high **discontinuation rates** distort the outcomes resulting in under reporting of safety problems or otherwise create a distortion due to such issues as subjects using other interventions?
3. Are **missing data** likely to distort results? Are missing data imbalanced between the groups?

IV. Assessment of Outcomes

1. Was **assessment blind**?
2. Were **analysis methods** appropriate including predefined groups for analysis?
3. If **composite outcomes** were utilized, were they reasonable?
4. If appropriate, was analysis done by **Intention-to-Treat** (all patients evaluated in assigned groups) with **missing variables assigned** by reasonable methods which will not favor the intervention?
5. Were assumptions used for **modeling** reasonable?
6. Was reporting likely to have been **selective**?
7. Was **safety** assessed and reported?
8. Have results been **confirmed** in other valid studies?

Delfini Pearls: Evidence Grading

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Rationale for Evidence Grading
Effective critical appraisal requires assessing both validity and usefulness of studies or study results. An evidence grade rates a study or outcome. Higher grades of evidence reflect higher quality which is more likely to report more accurate estimates of effect.

Evidence Grading Systems
It is important to examine the criteria used in the various grading systems because some systems assign misleading quality grades by inflating lower quality or invalid studies.

Delfini Evidence Grading Scale & Strength of Evidence Considerations

Grades can be applied to individual studies, to conclusions within studies, a body of evidence or to secondary sources such as guidelines or clinical recommendations. General advice is provided below. (Due to complexities with studies of diagnostic tests, no recommendations for them are provided here.) All-or-none studies (observational) may be an exception and occur rarely.

Grade A: Useful
The evidence is strong and appears sufficient to use in making health care decisions—it is both valid and useful (e.g., meets standards for clinical significance, sufficient magnitude of effect size, physician and patient acceptability, etc.). Studies achieving this grade should be outstanding in design, methodology, execution and reporting and have successful study performance outcomes, providing useful information to aid clinical decision-making, enabling reasonable certitude in drawing conclusions.

- For a body of evidence: Several well-designed and conducted studies that consistently show similar results.
- For therapy, screening and prevention: RCTs. In some cases a single, large Grade A RCT may be sufficient; however, without confirmation from other studies, results could be due to chance, undetected significant biases, fraud, etc. In such instance, the SOE should include a cautionary note.
- For natural history and prognosis: Cohort studies
- Grade A should be rarely assigned to any study. ("Extra points" are not given for challenge or difficulty in answering the research question. Authors should not be given extra points by second-guessing them. Transparency is required.)

Grade B: Possibly Useful
Grade B studies should be very well designed and executed and meet most of the requirements that it takes to achieve a Grade A. Grade B evidence appears potentially strong and is probably sufficient to use in making health care decisions—some threats to validity have been identified. Studies achieving this grade should be of high quality and contain only non-lethal threats to validity and with sufficiently useful information to aid clinical decision-making, enabling reasonable certitude in drawing conclusions.

- For a body of evidence: The evidence is strong enough to conclude that the results are probably valid and useful (see above); however, study results from multiple studies are inconsistent or the studies may have some (but not lethal) threats to validity.
- For therapy, screening and prevention: RCTs. In some cases a single, large Grade B RCT may be sufficient; however, without confirmation from other studies results could be due to chance, undetected significant biases, fraud, etc. In such instance, the SOE should include a cautionary note.
- For natural history and prognosis: Cohort studies
- Grade B is more frequent than Grade A, but is still a difficult grade to achieve.

Grade B-U: Possible to uncertain usefulness

The evidence might be sufficient to use in making health care decisions; however, there remains sufficient uncertainty that the evidence cannot fully reach a Grade B, and the uncertainty is not great enough to fully warrant a Grade U.

Grade U: Uncertain Validity and/or Usefulness
There is sufficient uncertainty that caution is urged regarding its use in making health care decisions. Grade U should be assigned when there is sufficient uncertainty about the accuracy of the estimates of effect resulting in an inability to comfortably draw conclusions from the research and in comfortably applying results.

- We end up assigning most studies a Grade U. As stated, we generally never use Grade U studies to inform efficacy decisions, but we will use Grade U evidence for safety, being very careful to describe that the evidence is of low quality.

The Agency for Healthcare Research and Quality (AHRQ) has a simple, useful system for grading evidence of individual studies and the overall strength of evidence (SOE)considering all included studies:

- Individual study risk of bias ratings: high risk of bia, medium risk of bias, low risk of bias
- Overall SOE ratings: High, Moderate, Low, Insufficient

Delfini Modifications: Overall level of evidence (LOE) ratings: High, Moderate, Borderline, Inconclusive

Delfini Pearls
Assessing Results: Considerations for Intention-to-Treat Analysis

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Superiority Trials

Intention-to-Treat (ITT) analysis should generally be the primary method for analyzing results of superiority trials for **efficacy** of therapies (not safety), when outcomes are dichotomous, to keep randomization intact and to deal with missing data.

- ITT analysis requires that—
  - Subjects are analyzed in the groups to which they were randomized; and,
  - Some value is included for each subject in the analysis.
- **Assignment of missing values** (imputation of data) generally is done through—
  - Attempts to estimate truth, or
  - Challenges against the intervention to discern if statistical significance is maintained.
- Methods which attempt to **estimate truth** include—
  - Mixed effects models (e.g., mixed linear, two-stage random effects or random coefficient models), multiple imputation models (software program), cumulative change approach; however, these are models, models are not truth and assumptions used in creating models are infrequently reported and so are usually unevaluable.
  - LOCF (last observation carried forward) is prone to bias and should not be used. However, in cases of progressive conditions may be conservative and reasonable at least to determine efficacy, if not actual estimate of effect.
  - Applying the mean results in the same answer as a completer analysis and should not be employed.
  - Baseline carried forward may be a possibly acceptable method
- Methods which put results through a **challenge test** are not attempts to estimate truth, but present a hurdle, that if met, can provide confidence in the direction of the results.
  - Extreme-case analysis puts the intervention through the toughest test (e.g., missing in intervention group are counted as "treatment failures" and missing in comparison group are counted as "treatment successes").
  - Apply control subject recovery rate to all groups for imputation.
- **Sensitivity analyses** (what-if scenarios) test the strength of the data.

Equivalence & Non-Inferiority Trials

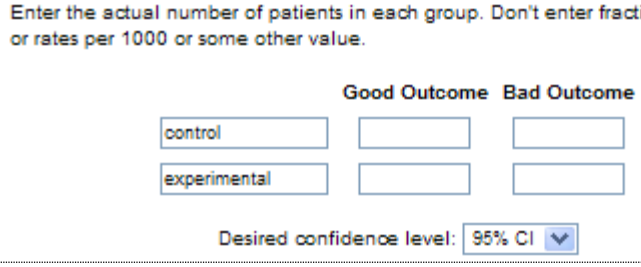
- **Intention-to-Treat (ITT) analysis** is generally a conservative imputation method and should **NOT** be used as the primary analysis for analyzing results of equivalence and non-inferiority trials if it biases the results towards equivalence.

Considerations & Critical Appraisal Issues

- Was analysis an appropriate method or not?
- If imputation was performed, was it appropriate or not?
  - Methods should not favor the intervention.
- If ITT has not been done, do missing values exceed your threshold? If yes, if the study would otherwise get a passing grade, consider doing a re-analysis.
  - Prepare to create a 2 x 2 table which requires the number in each group to be analyzed based on positive or negative outcomes.
  - Determine the number of subjects in each group with positive outcomes, negative outcomes and indeterminate outcomes. Distribute the indeterminate outcomes to each group as desired.
  - Compute the p-value and/or confidence intervals.

Web Link for Computing Confidence Intervals

http://www.graphpad.com/quickcalcs/nnt1.cfm



Delfini Pearls
Analyzing Results: Time-To-Event Analysis— Kaplan Meier Survival Curves & Hazard Ratios

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Survival Curves measure the length of time to an outcome of interest, (e.g., time-to-pregnancy, time-to-cancer progression). **Synonyms**: Life table analysis and survival analysis which refers to the method regardless of whether survival is the outcome. Kaplan-Meier methodology is the most commonly used survival analysis in healthcare.

- Because a bias could result from subjects spending different amounts of time in the study (e.g., a subject being enrolled near the end of the study), "censoring" is almost always utilized in time-to-event analysis.

Censoring is the practice of removing the patient from the curve at a specific point in time. Examples of censoring:
1) Patients who don't experience the event (administrative censoring or right censoring and which is considered acceptable); and , 2) Other reasons determined by the investigators and called "censoring rules" (non-administrative censoring such as lost to follow-up or dying before a non-mortality outcome of interest is reached). These latter censoring rules should be evaluated for potential bias.

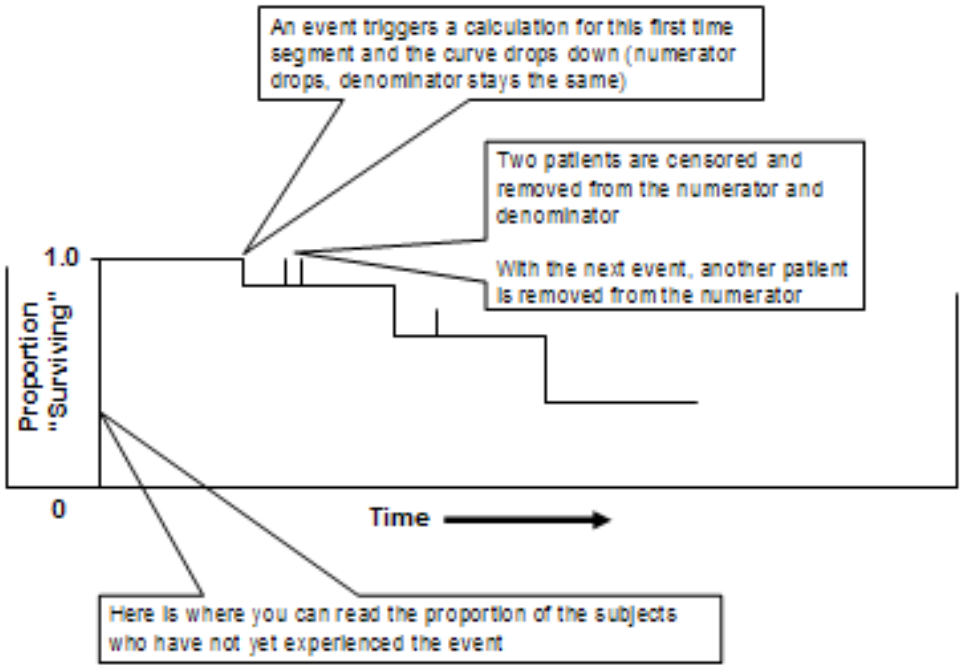
- Censored data is assumed to occur randomly (may not be a valid assumption).
- Censoring reduces sample size which may reduce reliability of results.
- Censored subjects may differ from subjects remaining in the study and may create bias.

Creation of the curve involves computing the number of people who experience the outcome at a certain time point, divided by the number of people who were still in the study at that time taking into account the censored patients.

- When a patient's data are censored, the number of patients "at risk" (numerator and denominator decrease) is reduced by one when the calculation for that time segment is performed. When a patient experiences the outcome, the "survival" for the interval is calculated (numerator decreases) according to the number remaining at risk at the time of event. (Denominator is decreased for the next interval.)

Hazards, Hazard Rates and Hazard Ratios

- A hazard is an incidence rate. A hazard rate (slope of the survival curve) is a measure of how rapidly subjects are experiencing the endpoint. A hazard ratio (calculated using Cox proportional hazards model) approximates the relative risk in the intervention group compared to the control group and is assumed to remain constant (may be an invalid assumption). Median survival is usually presented with hazard ratios. Example: With an HR of 2, a patient who has not yet experienced the outcome at a certain time has twice the chance of experiencing the outcome at the next point in time compared to a subject in the control group.



Considerations & Critical Appraisal Issues

- KM models assume on average the likelihood of experiencing an endpoint is the same for early enrolled subjects and subjects enrolled later (may not be valid).
- KM models assume that the likelihood of experiencing an endpoint is the same for censored and non-censored patients (may not be valid).
- The average HR (usual way of reporting HRs) ignores the distribution of events over time.
- Period-specific HRs are also biased in that susceptible subjects are removed over time resulting in a study population that may have different prognostic variables.
- Appraisers need details of censoring—how many subjects censored in each time segment and why; without this information appraisers cannot evaluate the possible impact of censoring or perform sensitivity analyses
- Survival analysis should not be applied to reoccurring rates so need to ensure double-counting does not occur (e.g., composite endpoint of mortality and MI).
- If any data are available at all for each patient in a study, the investigators frequently state that they analyzed the data according to "the Intention-to-Treat (ITT) principle." However, because the patient's future information is effectively removed at the point at which they have been censored, this is technically not ITT analysis, plus there is no imputation of missing values.
- Censoring reduces sample size which reduces reliability.
- Censoring may not occur at random.
- Censoring assumes that subjects lost to follow-up are similar to those who are not lost — they may not be, so amount of loss and loss difference between groups matters.
- Outcomes in completers may be different from what outcomes would have been without data loss (i.e., censoring may result in attrition bias).
- Even without differential loss between the groups overall, a differential loss could occur in prognostic variables.
- Assessing outcomes through models (e.g., Kaplan Meier estimates) has been reported to potentially erroneously misrepresent outcomes by a relative 50% or higher (Lachin: PMID 11018568)

Delfini Pearls
Assessing Results: Point Estimates, P-Values, Power & Confidence Intervals

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Meaningful Clinical Benefit is a combination of—

- **Clinically significant areas** (morbidity, mortality, symptom relief, emotional/physical functioning and health-related quality of life) +
- **Effect size** which means the difference in the size of outcomes between groups reported as measures of outcomes.

Measures of Outcomes measure the event outcome differences in the groups and should always be associated with the study time period. Most of these are measures of probability that an event will occur. **Synonyms for expressing effect size:** estimates of effect, point estimates.

- **Risk With and Without Treatment** is the number or percent of outcomes in each group (if number = 2 x 2 table data)
- **Absolute Risk Reduction (ARR)** is the percentage difference in outcomes between groups.
- **Number-needed-to-Treat (NNT)** is the reciprocal of the ARR and expresses the number of people needed to treat for 1 person to benefit over the comparator.
- **Relative Risk Reduction (RRR)** is the relative percentage difference in size between outcomes.
- **Relative Risk (RR)**, also known as Risk Ratio or Relative Risk Ratio, is the probability of the risk in the intervention group to the probability in the control group. Example: Probability of drawing an ace = 4/52.
- **Odds ratios (OR)** express the odds of an event occurring compared to not occurring and, therefore, cannot be as specific as probability measures. Example: Odds of drawing an ace = 4/48. Odds and RR are usually similar if the event rate is low, but compare both the odds ratio reduction and the RRR to check and see if the difference is clinically meaningful.
- Many of these measures can be used to express **harms** (e.g., Absolute Risk Increase or ARI, Number-needed-to-harm or NNH, etc.), prevention, screening, etc.

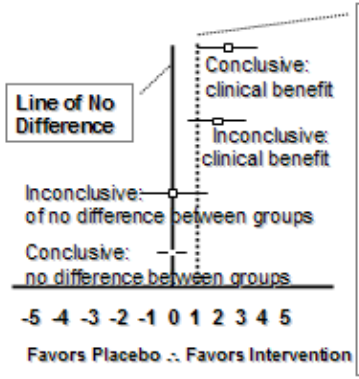
P-values: Assuming there truly is no difference between the groups studied, the P-value is a calculated probability of observing a difference as big as or bigger than the one you observed in a study based on compatibility with an assumed standard distribution. Problems include the P-value cannot tell you the chance the results are true or even how likely they are to be due to chance, you do not know if the null hypothesis is true or not, and you do not know if the sample is truly random and/or representative of the population.

- **Non-significant results** in a valid study arise either because it is true there is no difference between groups, it is a chance effect or there were insufficient people studied for the outcome to happen (e.g., the study was not sufficiently "**powered**").
- **Trials stopped early** present a high risk of outcomes due to chance even if stopping rules are applied.
- **Multiple outcomes** and **multiple analyses points** increase the likelihood of chance effects as high as x the number.
- ***A priori*** should also be research questions, populations for analysis and outcomes for measurement to reduce risk of chance effects

Confidence Intervals (CIs) represent a range of statistically plausible results consistent with an outcome from a single study. Example: ARI 1.3%, 95% CI (-0.21% to 2.8%), p=0.11.

Confidence intervals have some practical limitations similar to P-values. Although the CIs can project a range of results consistent with the study results, they cannot tell you the truth of the outcomes. CIs cannot replace the need to critically appraise the study. **Clinical significance** can be determined by whether values are wholly within or outside values judged clinically meaningful.

Applying CIs to Decide Meaningful Clinical Benefit Size: Hypothetical Endpoint—Reduction in Mortality



For valid studies, compare the confidence intervals to *your line* to evaluate the possibilities for or against what *you* define as meaningful clinical benefit

Given the margin for the play of chance –

- If wholly within your margin, results can be considered conclusive for meaningful clinical benefit
- If overlapping your line, results are inconclusive
- If wholly outside, results can be considered conclusive for no difference between groups

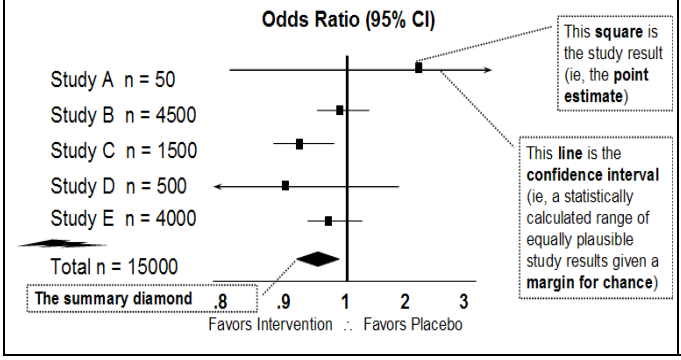
Delfini Pearls

How to Read a Forest Plot

1

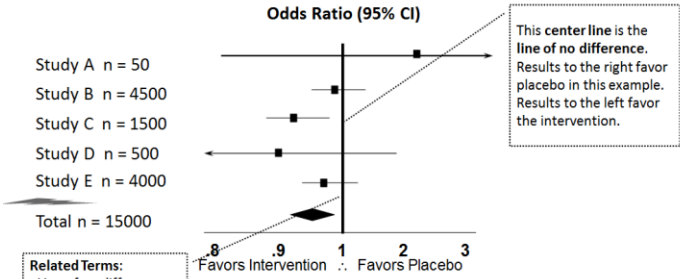
Graphic Display: Point Estimate, CI and Summary Diamond

These are several studies reported in a meta-analysis (some studies are removed, so this is not meant to total correctly) — this is just a sampler.



2

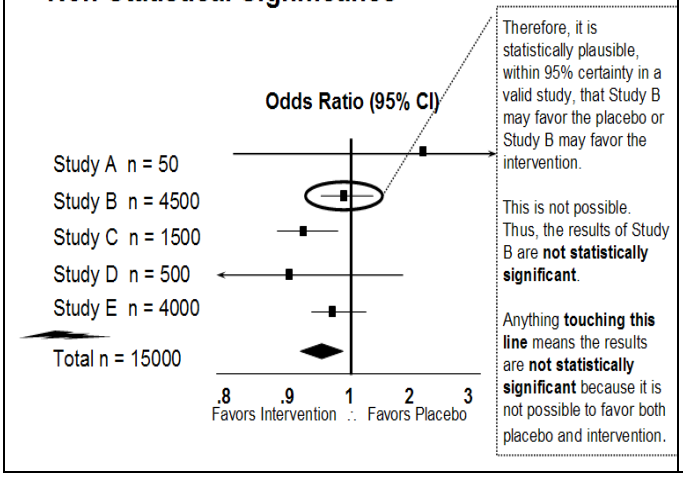
Favors Intervention ∴ Favors Placebo & The Line of No Difference



Related Terms:
- Line of no difference
- Line of no effect
- Infinity (NNT etc)
- Unity (ratios)

3

Non-Statistical Significance



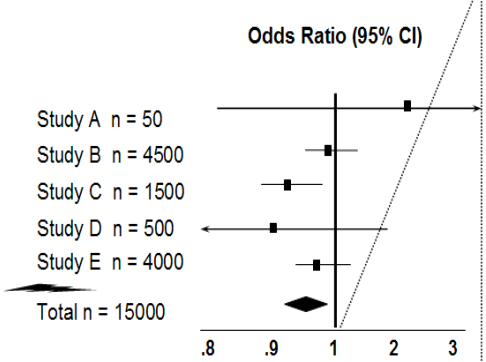
Therefore, it is statistically plausible, within 95% certainty in a valid study, that Study B may favor the placebo or Study B may favor the intervention.

This is not possible. Thus, the results of Study B are **not statistically significant**.

Anything **touching this line** means the results are **not statistically significant** because it is not possible to favor both placebo and intervention.

4

You Need to Know the Numerical Value for "No Difference"



The line of no difference equals 0 or 1 depending upon the measure of outcome used.

No difference for a **percent** is expressed as **zero**.

ARR and RRR are expressed as percentages. Therefore, if these were used, this number would be zero.

No difference for a **ratio** is 1:1. So for **odds ratio or relative risk ratio (aka relative risk),** this number **equals 1**.

Delfini Pearls
Safety Issues & Considerations

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Key Points About Safety Evidence

- **Safety** issues concern risks and harms which are events that cause problems with meaningful outcomes (morbidity, mortality, quality of life, functioning) or cause other unwanted effects.
- **Terms** "safety, risk, harm, adverse event, adverse effect, ADE"are often used interchangeably.
- Harms are infrequent, hard to find and are usually not the topic of study (not determined *a priori* and therefore there is a greater likelihood that findings are due to chance).
- There are potential limitations of RCTs and systematic reviews of RCTS that are not specifically focused on safety questions when the RCTs —
    - May not have reported or fully reported adverse events
    - May be of insufficient duration
    - May have relied upon small populations (eg, sampling error or power issues)—**BEWARE OF NON-SIGNIFICANT FINDINGS!** Non-significant findings could be chance effects, could be due to insufficient numbers of patients to find differences between groups—or could truly be due to no difference between groups.
- Harms are often reported from weaker science such as case report data, database research, observational studies or low quality RCTs.
    - Reminder: With rare exceptions, cause and effect can only be reliably concluded from valid RCTs.
- If outcome measures are not identified *a priori*, it increases the possibility that the findings are due to chance.
- **High discontinuation rates** in studies may result in agents appearing safer than they actually are.
- When effective interventions are no longer available (eg, have been discontinued by the manufacturer) due to poor safety data – which could be inaccurate – patients may be harmed.

Safety data are, therefore, usually not strong and often likely due to chance.

Where to Look

- Systematic reviews of RCTs dealing with harms should be sought, but harms may not be detected if some of the included trials do not report harms or if harms are described in various ways in different studies.
    - In some cases, systematic reviews may falsely indicate lack of harms that are subsequently detected in large, well-designed and conducted RCTs.

Search for observational studies, keeping in mind that observational studies are prone to bias.

Considerations & Critical Appraisal Issues

- In RCTs, the safety population should be only those who receive the intervention.
- Unless a study is powered for harms, lack of statistically significant differences may mean there is no difference or it may mean it is still unknown if there is a difference. Confidence intervals are useful in evaluating harms. Review confidence intervals (CI) for non-significant findings to discern if there is a clinically meaningful difference between the groups within the confidence interval.
- Review multiple studies. Look for patterns.
- Note if support exists for the harm (eg, biologic plausibility, relatedness in outcomes, dose-response relationship).
- Review the exclusions: Exclusion of patients otherwise likely to experience side-effects may affect generalizability of results of adverse events reporting (eg, may happen if patients are restricted to those who are not naïve or may occur through a run-in and exclusion period).
- Review drop-outs due to adverse effects.
- If composite endpoints are used for efficacy, are they used for safety?
- Caution is especially warranted for new agents.
- Beware of the potential for overreacting to possible harms and the risk of creating unintended consequences.

Bradford Hill Criteria for Supporting Considerations of Causality [Delfini Comments or Paraphrasing]

Caution is urged in applying the criteria below as these are neither requirements, nor guarantees, of causality and may not be reliable—but they may be worthwhile to consider:

1. Strength of Association [aka estimates of effect]; 2. Consistency—has it been repeatedly observed by different persons, in different places, circumstances and times?; 3. Specificity [eg, a specific kind of cancer is seen in more people who smoke than in those who do not]; 4. Temporal relationship; 5. Biological gradient [eg, dose-response relationship]; 6. Plausibility [supportive, but not required as is dependent upon what is currently known]; 7. Coherence—not seriously in conflict with generally known facts of the natural history and biology of the disease; 8. Experiment [experimental support]; 9. Analogy [potential for following a pattern such as a virus known to cause birth defects; therefore, maybe that another does too].

HILL AB. THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? Proc R Soc Med. 1965 May;58:295-300. PubMed PMID: 14283879; PubMed Central PMCID: PMC1898525.

Additions AHRQ: Lack of alternative causes, drug levels in body, resolves or improves after discontinuation, & recurrence with restarting.

http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=327

Delfini Pearls
Analyzing Results: Composite Endpoints

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Composite endpoint refers to individual endpoints grouped together for results reporting to serve as a single outcome measure

- Examples—
  - Major cardiovascular events = consisting of several individual outcome measures = cardiovascular death, nonfatal myocardial infarction, stroke
  - Diabetic nephropathy = decreased renal-function, end-stage renal disease, death
  - In oncology, disease-free survival = No tumor recurrence, alive at time of measurement

Synonyms for Endpoint: Measure or measurement; outcome measure or outcome (eg, cardiovascular mortality, number of pain-free days)

Reasons for composite endpoints—

- Greater frequency for otherwise infrequent events
- Allows for smaller sample size
- May form a more robust picture when dealing with a variety of hoped for outcomes (eg, reduction in mortality from MI + prevention of MI)
- There is also a potential for misleading readings—
  - Point being that you have to **watch out** because an investigator can set up the composite endpoint (intentionally or not) to have a high likelihood of showing a desirable outcome. ↓

Cautions

Watch out for what component of the endpoint is driving the results and determine how clinically significant and valid it is—

- "It will either rain or be dark tomorrow."

Considerations & Critical Appraisal Issues

- Is the combination valid, reasonable, fair and clinically useful? Is there any way that its construction is likely to favor the intervention? Watch out for –
  - Subjective outcomes especially if no blinding
  - Combinations including severe outcomes with mild ones, process measures, intermediate markers without a direct chain of causality to a clinical outcome, items under control or influence of a participant in the research
  - Did the researchers avoid double-counting (eg, if someone dies of stroke, did they get counted in both stroke and death)?
  - How meaningfully-related is the combination?
  - Are there other ways the combination could be misleading?
    - Disease-free survival when a treatment reduces risk of tumor recurrence but increases risk of death
  - Did they report results on the individual components? Without this information, depending upon the combination, a situation could result in which symptoms decreased, but mortality increased, but the composite masks this untoward outcome.
  - All-cause mortality is an important outcome as it is likely to be "unbiased." If randomization is successful and the study is otherwise valid, any non-treatment related deaths should be likely to be balanced between the groups or be the result of chance. Disease-specific mortality provides additional information about death from specific causes, but disease-specific outcomes may be biased, if groups are not balanced at outset or blinding is not successful. Biases could result from group imbalance or bias or errors in assigning cause of death. Mortality outcomes are prone to power problems.

Delfini Pearls
Oncology Outcomes Chart and Key Considerations

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Typical Oncology Outcomes

Endpoint	Description	Comment
Overall Survival	Defined as the time from randomization until death from any cause and is measured in the intent-to-treat population	Preferred overall
Progression-Free Survival (PFS)	Defined as the time from randomization until objective tumor progression or death	Preferred to Time-to-Progression; Used for some accelerated approvals Prone to tumor assessment biases If patients are measured until progression and are still followed until death, there is potential for confounding of results post-progression if other treatment is utilized.
Disease-Free Survival (DFS)	Defined as the time from randomization until recurrence of tumor or death from any cause	Prone to tumor assessment biases
Objective Response Rate (ORR)	Defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period	Prone to tumor assessment biases
Time-to-Progression (TTP)	Defined as the time from randomization until objective tumor progression	Prone to tumor assessment biases
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	Not recommended as a regulatory approval endpoint – likely to report biased outcomes as it does not adequately distinguish efficacy from other variables

Key Points About Oncology Studies & Outcomes

Rank of Endpoint Quality

1. Death
2. Death plus tumor assessment judgments
3. Tumor assessment judgments

- In addition to usual biases in clinical trials, there is a higher likelihood of bias and the risk of potentially misleading results when studies are **small** and **brief** and **when survival is not the primary outcome measure**.
- Progression-free survival (PFS) may be a composite endpoint including tumor response.
- Tumor response may not be a good proxy for survival even if assessment is blinded.
    - Tumor may shrink, but may otherwise have increased metastatic disease or other tumor growth as tumors do not grow at the same rate.
    - Toxicity of treatment may be so great that patients die from it even if tumor is stable or shrinking.
- Quality of life and functioning may be important endpoints to study in absence of true survival information.
- Overall survival differences even when statistically significant may be small.

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Screening is the process of identifying a disease, condition or risk factor in asymptomatic patients regardless of setting (practical definition).

- It is useful to think of screening as a type of therapeutic intervention, but screening embodies elements of both diagnosis and treatment.
- Screening may appear to be a good thing, when, in fact, harms might outweigh benefits.
- In addition to usual considerations for interventions, clinically meaningful screening requires that early detection and treatment improve outcomes more than later (symptomatic) diagnosis and treatment.

Screening categories are—

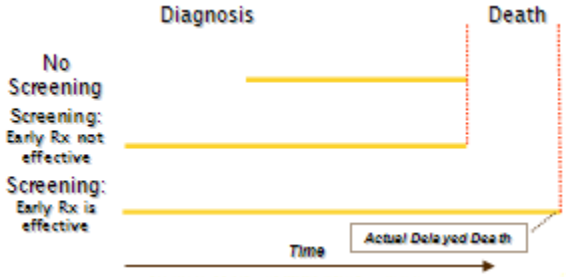
- **Primary Prevention**: Prevention of disease by eliminating causes OR interrupting disease processes before they become established or symptomatic.
- **Secondary Prevention**: Limiting the harms (symptoms, functioning, mortality) done by established disease processes.

Special bias issues in screening—

- **Lead Time Bias** occurs when early detection makes it look like there is longer survival time, but the date death occurs is no different.
- **Length Bias** is a "disease-spectrum" bias and occurs when screening "appears" to improve survival due to missing the most deadly tumors and finding tumors that people are more likely to live with or live a long time with. Screening is more likely to find slower growing tumors that may not be harmful, or as harmful (aka "overdiagnosis bias").
- **Volunteers** participating in screening have been shown to have better outcomes than those who don't (i.e., those who are persuaded to participate), possibly due to the healthy user effect.

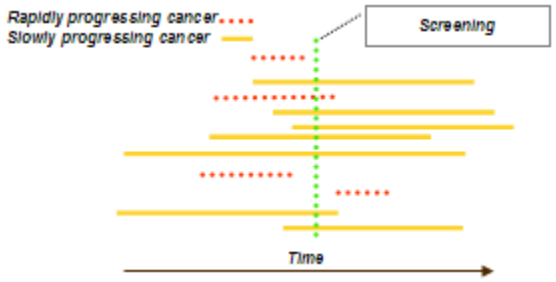
Lead Time Bias

Early diagnosis always increases "survival" (dealth from time of diagnosis):
survival appears longer from time cancer is found.



Length Bias (Disease-Spectrum Bias or Overdiagnosis Bias)

Disease progressing rapidly is likely to be missed by screening.



Considerations & Critical Appraisal Issues

- Requires valid studies of diagnostic tests and interventions demonstrating improved outcomes with early interventions compared with intervening after symptoms develop.
- Assess potential for lead-time bias, length bias and volunteer bias.

Delfini Pearls
Basics of Evaluating Evidence in Diagnostic Studies

Healthcare Information & Decision Equation: **Information→Decision →Action→Outcome**
Is the diagnostic test accurate→Is it useful →Is it usable?

Quick Assessment:
The goal of diagnostic test is to identify individuals who could potentially benefit from other interventions (Cochrane Handbook). Important considerations in diagnostic testing include the following:

Net Benefit
- Does the new test provide improved accuracy and predictive value over existing tests?
- Will adoption lead to improved clinically meaningful outcomes?
- Do benefits outweigh harms?

Measures of Test Function (aka Estimates of Test Performance)
What are the accuracy and predictive capabilities of the test (from 2 x 2 table)?

	Disease	
Test Result	**Present**	**Absent**
+	a True Positives	b False Positives
-	c False Negatives	d True Negatives

Sensitivity (proportion of true positives)=a/a+c
Specificity (proportion of true negatives)=d/b+d
Practical usefulness is limited because these measures are dependent upon people known to have or not have the disease.

+Predictive Value (chance of having disease if test is positive)=a/a+b
- Predictive Value (chance of not having disease if test is negative)=d/c+d
More practical for use in patients in whom disease is unknown.

Likelihood Ratios (change from pretest to post test odds): The likelihood ratio combines information from sensitivity and specificity and indicates **how much the odds of disease change based on a positive or a negative result**. It is used together with the pre-test odds, which can be derived from prevalence information of the disease found in the study or by clinical judgment. By multiplying the pre-test odds by the likelihood ratio the post-test odds can be calculated:

+Likelihood Ratio (+LR) (positive test)=sens/1-spec
-Likelihood Ratio (-LR) (negative test)=1-sens/spec

Heavily dependent upon judgment and risky to apply unless pre-test odds are uncertain (~50 percent or less).

General Considerations
- Diagnostic testing is based on use of intermediate outcomes which raises possibility that test may not truly result in clinical significance.
- Although observational studies are acceptable for accuracy, RCTs are needed to demonstrate benefits for people exposed to testing.
- Typically there are trade-offs between the paired test function values. For example, increased specificity often comes at the cost of decreased sensitivity.
- Values of statistics from 2 x2 table are likely to vary with different populations. Disease prevalence, for example, affects predictive value.
- Frequently there is no single, accurate test for diagnosis. For example the diagnosis of rheumatoid arthritis involves history, physical exam plus laboratory testing.
- Frequently there is a clinical need to choose a less accurate method due to cost or risk (e.g., chest x-ray vs lung bx).

Critical Appraisal Considerations
1. Was the index test compared to a reasonable gold standard (reference) test?
2. Were the tests compared together sufficiently close in time to prevent a change in condition to affect test results?
3. Did the study include an appropriate population?
4. Was the reference test applied to all patients, or a random sample of patients, with and without the disease?
5. Were assessors blinded to the results of the comparison test?
6. Does the new test find the same spectrum of disease as the reference test?
7. Were the number of withdrawals and indeterminate tests acceptable?
8. Was assessment blind?

Delfini Pearls
Basics of Evaluating Cross-over Designs

Healthcare Information & Decision Equation: **Information➔Decision ➔Action➔Outcome**
Is the diagnostic test accurate➔Is it useful ➔Is it usable?

Terminology

- "Cross-over" is sometimes used to describe when patients end up in a treatment group other than the group to which they were assigned
- Synonyms: migration; exposure
- Can happen by accident or chance
- A patient or physician choosing to cross-over renders outcomes **"observational"**
- **Cross-over study design is one in which a patient is intentionally assigned to one intervention and then crossed over to another intervention (including placebo)**

Critical Appraisal Considerations

1. **Randomization** — patients are randomized to an intervention sequence – needs to vary so all patients are not receiving the same intervention at the same time
2. **Blinding** — including concealed cross-over points, risk of unblinding due to familiarity with intervention or comparator
3. **Timing** — including pre-specification of reasonable cross-over points, carry-over effects of intervention or non-intervention elements, disease issues (e.g., considering issues relating to curative potential, disease fluctuations, rebound, seasonal effect, etc.)
4. **Results** calculations can be exceedingly complex
5. **Loss** (magnified since patient serves as subject and control)
6. **Choice** versus assignment to crossover — choice to cross-over renders outcomes "observational"
7. **Inappropriate application** —
8. When an intervention has a lasting effect, such as irreversibility, because of the carry-over effect in the subsequent time period(s)
9. Unstable conditions such as rapidly progressing conditions because disease progression creates a confounding effect for the subsequent time period(s)

Delfini Pearls
Correlation Analysis

Healthcare Information & Decision Equation: **Information➔Decision ➔Action➔Outcome**
Is it true➔Is it useful ➔Is it usable?

Correlation analysis is a mechanism to analyze how different variables relate to each other.

Types of Variables: statistical tests are chosen based on type of variables; the 4 main types are—

- **Nominal** (named categories without any measurable scale such as ethnic groups)
- **Dichotomous** or binary (two mutually exclusive categories resulting in "either this or that" such as "death" or "survival")
- **Ordinal** or ranked (three or more variables that can be "ordered" or ranked such as good/better/best or satisfied/neutral/unsatisfied)
- **Continuous** (can be anywhere along a continuum, e.g., blood glucose readings)
- Variables under study are also classed as "**dependent**" (the outcome under study) or "**independent**" (all others that might affect the "dependent" variable)

Correlation Analysis includes the following analysis categories—

Analysis Type	Purpose	Analysis Methods
Univariate Analysis	Methods for analyzing data on a single variable	Frequency distribution
Bivariate Analysis	Assess relationship of two variables	Correlation analysis Linear regression
Multivariable Analysis	Assess relationship of multiple variables to a single outcome	Multiple regression Proportional hazards
Multivariate Analysis	Assess relationship of multiple variables to multiple outcomes	(not reviewed)

Sometimes "-variate" and "-variable" get misapplied

Pearson Correlation Coefficient

- Commonly used correlation analysis method
- Extent of the linear relationship (how independent and dependent variables change together) is calculated for the two variables by calculating the **Pearson correlation coefficient,** referred to as the **r value**
- Pearson correlation coefficient (r) is frequently used when both variables are continuous to show **how variables change together**, e.g., salt intake and blood pressure
- The *correlation coefficient* has a range of possible values from -1 to +1
- 0 indicates no relationship between the dependent and independent variables
- Positive correlation coefficients indicate that as the value of the independent variable increases, the value of the dependent variable increases
- Negative correlation coefficients indicate that as the value of the independent variable increases, the value of the dependent variable decreases
- r^2 (square of the correlation coefficient) represents the proportion of variation in y (on an x-y plot) explained by x (or vice versa)
    - Example: "…A moderately strong inverse criterion validity correlation (Pearson correlation coefficient = -0.68) was shown when preoperative patients were administered both the AOFAS and FFI questionnaires, and the resultant scores were compared."

Critical Appraisal Considerations

- It may be incorrect to draw cause/effect conclusions from correlations
    - Example: Height/weight are correlated, but height does not cause weight

Delfini Pearls
Comparative Study Designs of Clinical Trials

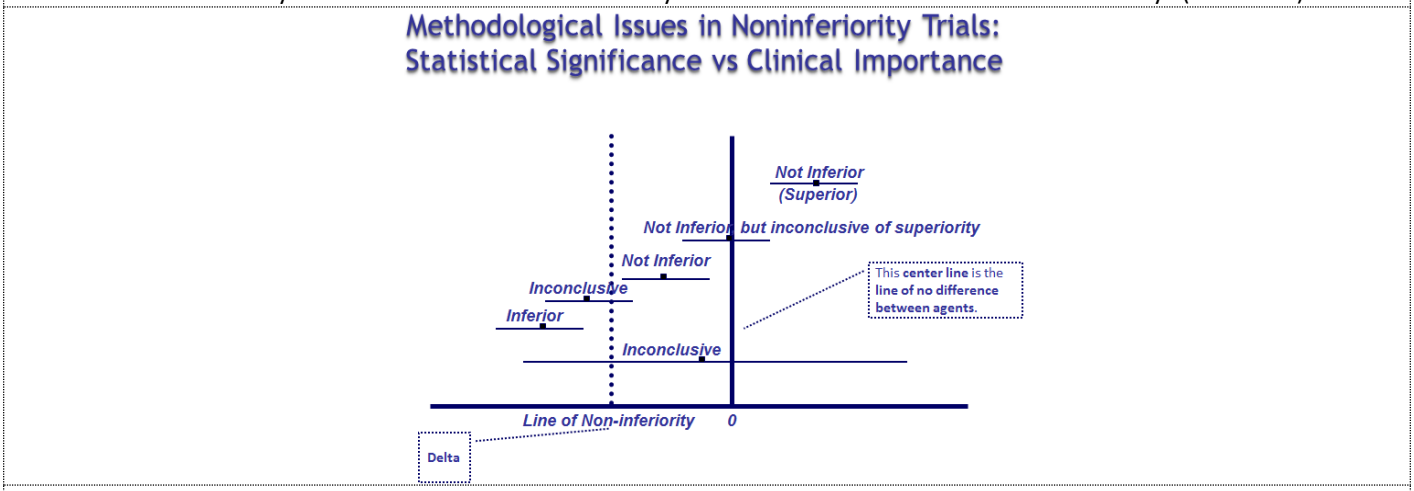
Healthcare Information & Decision Equation: **Information➔Decision ➔Action➔Outcome**
Is it true➔Is it useful ➔Is it usable?

Superiority is the typical aim of an RCT. Ideally, a non-inferiority test is included in superiority trials.
Equivalence trials aim to determine whether one (typically new) intervention is therapeutically similar to an existing treatment.
Non-inferiority trials seek to determine whether a new treatment is no worse than a reference treatment.
Delta: Because proof of exact equality is impossible, a margin of non-inferiority or equivalence ("Delta") for the treatment effect is defined. Establishing Delta requires statistical and/or clinical judgment. (GraphPad: "…determine your zone of scientific or clinical indifference…")

- For **equivalence trials**, two lines are established to define equivalence so that equivalence is defined as the treatment effect being between − delta and + delta: the confidence interval for the comparison of the new treatment to the old must be within this range. (Pictured below.)
- For **non-inferiority trials**, one line is established which represents the smallest amount of clinical benefit acceptable: the smallest boundary of the confidence interval (CI) for the comparison of the new treatment to the old must be above this line.

Terminology

- "New" refers to the treatment being tested.
- The comparison or "reference treatment" is often called an "active control" or "positive control."
- We refer to the study or studies that determined efficacy of the "active control" as the "referent study" (or studies).



Considerations & Critical Appraisal Issues For Non-Inferiority and Equivalence Trials

- Is the referent truly efficacious in area studied? Strongly recommended to obtain the referent study and critically appraise it as well as determine if the study of the new agent is sufficiently similar to the referent study. Review key details such as population, dosing, duration, co-interventions, adherence, endpoints, etc. Comparison is limited to the specific outcomes chosen—"equivalence" does not equate with "me too." Even if studies are well-done, true equivalence or non-inferiority cannot be directly established—there may be unaccounted for differences between agents.
- If the new agent has not been compared to placebo, then superiority to placebo can only be indirectly assumed even if the referent agent is superior to placebo.
- Superiority claim may, in a noninferiority or equivalence trial, be valid using an appropriate test with confidence intervals (not just point estimate): groups that agree superiority can be claimed under the right circumstances include CONSORT 06, FDA, EMEA. Multiplicity adjustment is not needed. Population should be ITT.
- Lacking direct comparison to placebo risks creating confusion about benefits and harms.
- Time may have affected efficacy for even the referent agent—such as changes in resistance patterns to antibiotics or in patient behaviors such as dietary changes due to public health interventions.
- Anything that diminishes effect size favors equivalence and non-inferiority (e.g., conservative application of ITT; insufficient power, which is determined by CIs, that result in "inconclusive" or "uncertain" outcomes, not blinded to study design without hard outcomes, etc.)
- Is the Delta clinically reasonable?
- **IMPORTANT: Claims of equivalence or non-inferiority may not be appropriate in superiority trials where delta is established *post hoc*. If prespecified and valid, claims can ONLY be made for the outcomes compared.**

Delfini Pearls
Basics of Evaluating Secondary Studies

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Definitions: Secondary study is a study of studies

- **Systematic Reviews**: A formal method for summarizing results of more than one study
  - Meta-analysis: systematic reviews that use statistical techniques to do this quantitatively
    - Meta-analyses either combine study results or pool actual study data
- **Overviews**: An informal method for summarizing results of more than one study (synonyms: narrative review, review)
  - Lack some or all of the necessary components of systematic reviews (e.g., a priori questions, systematic search, validity assessments, application of statistical tests) and present big opportunity for bias

Quick Assessment:

If the results are reliable, are they useful and usable? Would they change your practice? Do they apply to your situation considering your patients and circumstances of care? Consider effects on your patients including benefits, harms, risks, costs, uncertainties, alternatives, applicability, satisfaction, abuse and dependency issues. Consider conflicts of interest.

1. Are the results in **clinically significant areas** (morbidity, mortality, symptom relief, emotional/physical functioning and health-related quality of life)? If not, is there a reliable causal chain of evidence to support use of an intermediate marker?
2. Were outcomes and analyses determined **in advance**?
3. Are **definitions of outcomes** such as success/failure, improvement/no improvement, etc. reasonable?
4. Are the **confidence intervals** wholly inclusive of clinical benefit? If **non-significant**, are the confidence intervals wholly exclusive of clinical benefit?
5. Is this a **new intervention**? If yes, safety is likely to be unknown.

Study Design Considerations for Usability

1. **Randomized controlled trials** (RCTs) for efficacy and safety (tip: **choice of intervention was not made** by patient or patient's physician or by other means that would render study observational)
2. Possibly observation studies with **all-or-none results** (very rare)
3. Observational studies for **safety** if lacking quality information from RCTs

Validity Considerations

1. Research Question: Clearly stated and meaningful questions to the literature?
2. Study Selection: Explicit, documented and appropriate selection criteria chosen in advance for included studies that are sufficiently similar?
3. Study Design: If this is a question of therapy, screening or prevention, and observational studies are used to answer questions of efficacy, Delfini suggests not using the review.
4. Search Strategy: Documented systematic and comprehensive search strategy that is well thought out and executed? (Needs to include search terms, sources, filters used and dates covered and to include a search from the National Library of Medicine.)
5. Patient Population Assessment: Is the population appropriate for this question?
6. Critical Appraisal: What is the quality of included studies? (The Jadad Scale is not a good measure of study quality.
7. Missing Outcomes Data: Assessment of how loss to follow-up is handled and is it done appropriately?
8. Homo-/heterogeneity: If results of the studies were combined, such as in a meta-analyses, did the authors apply tests of homogeneity/heterogeneity to assure that the variation between studies is due to chance (i.e., p-value >.05, similar point estimates, overlapping CI's, I^2 statistic [I2 0-25% is good, to 50% moderate, to 75% not good]. Random effects models are often used when greater inconsistencies, but can overvalue small studies. Combining Results: If results were combined, was it done in a reasonable and appropriate manner?
9. Data Collection: Did more than one author extract and combine data?
10. Weighting: If weighting was employed, was a reasonable approach taken (e.g., larger or higher quality studies)?
11. Author's Discussion: Well executed sensitivity analyses, discussion of limitations, explanations of differences in studies and their results, etc.?
12. Other Issues (eg, potential conflict of interest)?
13. Author's Conclusion: Conclusions are supported by the evidence?
14. Transparency: Is sufficient detail provided that enables a through quality assessment of this review and such that this review could be replicated?

Delfini Pearls
Basics of Evaluating Secondary Sources

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Definitions

- **Secondary Source**: An information source that applies primary and/or secondary studies (e.g., guidelines, disease management protocols, cost-effectiveness studies)
- **Clinical guidelines**: Systematically developed statements to assist practitioners and patients in choosing appropriate healthcare for specific conditions per the Institute of Medicine (IOM). The best guidelines are based on evidence-based principles. Accurately predicting outcomes requires reliable information.

IOM's 8 Desirable Attributes Of Clinical Guidelines

• Validity	• Clarity
• Reproducibility	• Documentation
• Clinical Applicability	• Multidisciplinary Process
• Clinical Flexibility	• Plans for Review

Key Considerations for Critical Appraisal of Clinical Guidelines

1. Relevance to patients (clinically meaningful outcomes in mortality, morbidity, functioning, health-related quality of life, symptom relief)
2. Currency of information
3. Development involved people with appropriate perspectives
4. Evidence-based using systematic methods, and evidence was rigorously critically appraised as appropriate to the clinical question
5. Strength of evidence is disclosed
6. Recommendations and options are provided along with key information
7. Meets patients needs and accommodates different values and preferences
8. Limitations are disclosed
9. No other issues (eg, ethical issues, external requirements, etc) that would preclude adoption of the guideline
10. Likelihood of successful implementation and sustainability
11. Measurability
12. Impacts are reasonable (eg, patient outcomes, organizational impacts)
13. Mechanisms for updating the guideline if new evidence is available

Beware of low quality guidelines - review of 431 clinical guidelines:

- 82% did not apply explicit criteria to grade evidence
- 87% did not report whether a systematic search of the literature was performed
- 67% did not describe the type of professionals involved in the development of the guideline

PMID: 10675167

You will need to critically appraise and audit all secondary sources prior to adopting. Look for—

- Method for evidence grading and rating of recommendations. Common problem is "upgrading" of evidence, i.e., rating lower quality evidence as of higher quality. Example: one guidelines group rates the evidence from multiple low-quality studies as Level II (Strong Evidence)
- Systematic analysis of evidence (obtaining, critically appraising, grading and synthesizing evidence) to minimize bias

See **Delfini Pearls & Tools** for critically appraising and for auditing.

Delfini Pearls
Suggestions for Auditing Secondary Studies & Sources

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Auditing a Secondary Study or Secondary Source

It is important to critically appraise all secondary studies and secondary sources for how well they have been done, but it is also important to critically appraise or audit the science upon which the study or source is based.

Conservative Approach
- Critically appraise all studies utilized in the secondary study or source.

Minimal Approach (risky)
- One or two of included primary studies considered to be of the **highest quality** are critically appraised for validity and usefulness.

In either event—
- If the source passes for validity and usefulness, it may be reasonable to use the source's efficacy and safety conclusions in the evidence synthesis.

- If the source fails for validity and usefulness, but has utilized a **sound search strategy** and sound criteria for **excluding efficacy studies** lacking relevance, validity or for other problems, **all the primary studies** selected for inclusion by the source are **critically appraised**.

In all cases—
- **Update** with any new valid and clinically useful primary studies published since the date of the secondary source's search

Background: Evidence-based medicine and comparative effectiveness research movements have increased interest and activity in relative effectiveness of interventions.

- There is an expressed need for innovative approaches to clinical trials to be conducted under conditions of actual practice, enabling estimates of real-world effectiveness.
- There is an expressed need for statistical and epidemiological methods to predict patient responses to interventions.
- Key Requirements: **Transparency** so studies, data, conclusions can be assessed. Sufficient detail regarding methods, PICOTS (patients, interventions, comparators, outcomes, timing setting).

RCTs

Good internal validity if well designed and conducted; however, are time consuming, expensive and may lack external validity.

Pragmatic or Practical Trials

Danger of "buzz term."

Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in 6 therapeutical trials. J Chronic Dis 1967;20:637-48. PMID 4860352

- Referred to as "explanatory" trials if investigators attempted to establish causality.
- Referred to as "pragmatic" trials if designed to help choose options for clinical care.
- Pragmatic trials were NOT introduced as **a new trial design**, but rather an **"attitude"** to clinical trial design.

❖	Explanatory Trials	Pragmatic Trials
1	Efficacy	Effectiveness
2	Ideal conditions	Normal practice conditions
3	Highly selected (compliant, likely to benefit) subjects	Minimal selection criteria beyond clinical condition
4	Enforced, monitored interventions	Flexibility in interventions to reflect normal practice
5	Outcomes: short-term, intermediate	Outcomes relevant to end-users

- Need to critically appraise trials for validity first.
- The term "pragmatic" should not be assumed to be more valid or more useful.
- Explanatory trials may have good external validity.
- Pragmatic trials may have serious threats to internal validity.

Other Designs with Control Groups

- Interrupted time series or delayed treatment design: several units are studied with before/after intervention and progressively delayed starting times.
- Propensity scores
  - Start with observational study and assume equal groups using propensity scores (note—assumption likely to be wrong)
  - Then perform regression analysis providing estimate of effect
  - Scores can only account for the factors measured and only as well as the instruments can measure them (selection bias). Problems with differing dosages and other care experiences (performance bias). Requires modeling (assessment bias)
- Network meta-analyses
  - Assess the comparative effects of more than two alternative interventions for the same condition that have not been studied in head-to-head trials--they must have one intervention in common
  - Include both direct and indirect evidence (mixed comparisons)
  - Indirect evidence is derived from statistical inference—not direct comparisons—which requires multiple complex assumptions and complex statistical models to adjust for the inclusion of both direct and indirect evidence and multiple clinical and methodological differences in the included trials
  - The combination of direct and indirect evidence may be more likely to result in distorted estimates of effect size if there is inconsistency in effect sizes between direct and indirect comparisons
  - Network meta-analyses rank different treatments according to the probability of being the best treatment—rankings may be misleading because differences may be quite small or inaccurate if the quality of the meta-analysis is not high

Observational & Administrative Claims Data, Surveys, Medical Records

Can use these sources to (examples)—

- Identify populations for further study
- Evaluate implementation of intervention
- Generate hypotheses
- Current condition scenarios (e.g., who, what, where in QI projects)
- Safety signals
- Extend findings from RCTs, meta-analyses (e.g., registry data)
- Economic projections (e.g., balance sheets, models)

Need for more information on costs and benefits of data collection, transparency, skills in modeling

Positive Predictive Value by Study Type

Well-done RCT	0.85
Meta-analysis of well-done RCTs	0.85
Meta-analysis of small, inconclusive RCTs	0.41
Well-done epidemiological (observational) study	0.20
Epidemiologic study with threats to validity	0.12
Discovery-oriented exploratory research	0.0010

Ioannidis JPA. Why Most Published Research Findings are False. PLoS Med 2005; 2(8):696-701 PMID: 16060722

Delfini Pearls
Overview of Evidence-Synthesis

Healthcare Information & Decision Equation: Information→Decision →Action→Outcome
Is it true→Is it useful →Is it usable?

Evidence Synthesis is the process of summarizing systematically obtained and critically appraised evidence. Frequently clinical recommendations are created from evidence syntheses.

Creating Your Synthesis

Summarize the best available evidence. This may be a **text statement** or a **table** documenting characteristics of the evidence you have identified as being the best available. **There is no one correct way to summarize the evidence —** you will have to apply **judgment. Quantitate** as you can.

Evidence Grading

For individual studies, you grade the study or conclusions. For summaries of the evidence, you rate the level of evidence.

Level of Evidence (LOE) or Strength of Evidence (SOE) Example

- **High**: More than one grade A or B study
- **Moderate**: At least one grade A or B study
- **Borderline**: At least two grade B-U studies with consistent results
- **Inconclusive**: Single grade B-U study, B-U studies with conflicting results or Grade U studies

Elements You May Choose to Summarize

- Key clinical question
- Quality of the evidence
  - Key threats
- Type, number and size of studies (the "n")
- PICPOTS: population/condition, intervention, comparators, study performance outcomes (i.e., on-study adherence, on-study use of co-interventions, etc.), outcomes, timing, setting
  - Population description (see inclusions, exclusions and baseline characteristics)
- Interventions used and how
- How measured, successful outcome as defined as what?
- Results
- Limitations
- Reviewer conclusions and/or comments

Format Suggestion: Supporting Documentation

- Background
- Drug information
- FDA information
- Representation in Guidelines
- Expert Commentary
- Balance Sheet Information (Triangulations)
- Measurement Instruments and Interpretations
- Ideal study parameters
- Evidence synthesis tables
- Search & filtering strategy (efficacy, harms, other)
- Selection criteria for studies
- Methods used to determine validity and usability
- Grading scheme
- Table of included studies
- Critical appraisals of included studies
- Table of excluded studies
- References
- Glossary
- Conflicts of interest
- Reviewers
- Preparers
- Date

Example of Evidence Synthesis: MRI Use for Women At High Risk of Breast Cancer

- The strength of the evidence (SOE) is insufficient to conclude that, in high risk women, the addition of MRI to mammographic screening reduces the need for mammography or ultrasound.
- Adding MRI will change treatment plans and result in more extensive surgery for some women (SOE: Borderline), but may not change incomplete excision rates or breast cancer recurrence rates (SOE: Inconclusive).
- We found no evidence that adding MRI to conventional screening in women at high risk of breast cancer will reduce mortality rates SOE: Inconclusive).

Delfini Evidence Tool Kit

Short Critical Appraisal Checklist: Interventions for Prevention, Screening & Therapy

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

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Study Design Assessment **POTENTIAL EXCEPTION: ALL-OR-NONE RESULTS**	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** (allocation by minimization may be acceptable) ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?
Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ If variables are dichotomous, **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values). (May not be an issue if missing values are very few.) ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
External Validity	**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**.
Patient Perspective	☐ Consider benefits, harms, risks, costs, uncertainties, alternatives and satisfaction
Provider Perspective	☐ Satisfaction, acceptability (includes adherence issues, potential for abuse, dependency issues), likely appropriate application and actionability (e.g., FDA approval, affordability, external relevance, circumstances of care, able to apply, tools available)

☐ **Non-Inferiority & Equivalence Supplement:** Absence of the following problems: lack of sufficient evidence confirming efficacy of referent treatment; study not sufficiently similar to referent study; inappropriate Deltas; or significant biases or analysis methods which would tend to diminish an effect size (e.g., conservative application of ITT analysis, insufficient power, etc.)

☐ **Diagnostic Test Supplement:** New test requires better outcomes or value. Test is compared to gold standard or reasonable comparator and finds same abnormality and within time period that does not result in a change in diagnosis. Test is applied to all or random sample of subjects with and without disease. Assessors are blinded. There is minimal bias from indeterminate results. Measures of test function are useful.

☐ **Screening Supplement:** Early diagnosis and treatments determined to be effective will improve outcomes more than later diagnosis and treatment. Beneficial outcomes are not explained by bias (e.g., lead time, length, overdiagnosis or volunteer bias).

Delfini Evidence Tool Kit

Study Validity & Evidence Usability: Tool and Primer for Secondary Studies (Including Systematic Reviews & Meta-analyses)

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

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Systematic Review Study Details	
PICOTS (population, intervention, comparator, outcomes, timing, setting):	
Number of studies included / Number of subjects included:	
	Reported Results **Primary outcome measures:**
	Secondary outcome measures:
	Authors' conclusions:

Systematic Review Validity Assessment		
1.	**Best Sources:** ▪ If from a "best source" (see ***Delfini*** **Searching & Sources Tool**) — o We still recommend that you critically appraising the review and perform an audit (see ***Delfini*** **Searching Tool** for tips on working with best sources and audit recommendations) o Ensure they are not drawing cause and effect conclusions from poor evidence **Your Assessment:**	
2.	DARE Review: Is there an assessment of this study from DARE (see ***Delfini*** **Searching & Sources Tool**)? If yes, and DARE says use with "caution," probably the review should not be used for drawing cause and effect conclusions about efficacy. **Your Assessment:**	
3.	**Commentaries:** Documentation of any flaws or pertinent information found in study "commentaries" in PubMed. **Your Assessment:**	
4.	**Research Question:** Clearly stated and meaningful questions to the literature? For example, can you tell from the questions they pose to the literature that they will be capturing the right information for population, condition, intervention or exposure and outcome. **Your Assessment:**	
	Poor Quality Answer: We retrieved all studies dealing with pimecrolimus therapy for atopic dermatitis in the last 5 years. (Having many questions or many outcomes assessed is a red flag.)	**Good Quality Answer:** We utilized a two part question to the medical literature including the condition and the intervention. In PubMed the search terms were: atopic dermatitis, pimecrolimus OR Elidel OR SDZ ASM 981.
5.	**Clinical Significance of Question:** Does the research question address morbidity, mortality, symptom relief, emotional and/or physical functioning or health-related quality of life? **Your Assessment:**	

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

	Poor Quality Answer: Outcome measure was skin thickness by ultrasound.	**Good Quality Answer:** A priori stated outcome measures of pruritis score, percent days using topical steroids, and overall rating of disease control.
6.	**Study Selection:** Explicit, documented and appropriate selection criteria chosen in advance for included studies that are sufficiently similar? For example, needs to specify study type (eg, RCT, cohort, etc.), population, methods, inventions or exposures and outcomes. ▪ Sufficiently similar means similar in methods, population, intervention or exposures or characteristics, follow-up period, outcomes, etc. ▪ Preferably more than one author selecting studies? **Your Assessment:**	
	Poor Quality Answer: (For question of therapy.) RCTs were sought. Observational studies were used when RCT information was not available.	**Good Quality Answer:** For efficacy, effectiveness and adverse events we included valid and useful systematic review and meta-analysis data, and randomized controlled trials using antihypertensive medications dealing with the following clinically meaningful health and health care outcomes: mortality, morbidity, quality of life, functioning, and symptom relief. We excluded observational studies, editorials, opinion pieces, narrative reviews, animal studies, and studies with clinically non-useful outcomes.
7.	**Study Design:** If this is a question of therapy, screening or prevention, and observational studies are used to answer questions of efficacy, ***Delfini*** suggests not using the review. **Your Assessment:**	
	Poor Quality Answer: (For question of therapy.) RCTs were sought. Observational studies were used when RCT information was not available.	**Good Quality Answer:** Only RCTs judged to be valid were included.
8.	**Search Strategy:** Documented systematic and comprehensive search strategy that is well thought out and executed? ▪ Needs to include search terms, sources, filters used and dates covered ▪ Needs to include a search from the National Library of Medicine ▪ Textbooks are generally not considered to have relevant scientific information **Your Assessment:**	
	Poor Quality Answer: Medline search through 1995. References, abstracts, *Current Contents*, textbooks were evaluated for relevant information.	**Good Quality Answer:** Cochrane Database, Clinical Evidence and PubMed (National Library of Medicine) were systematically searched on March 1, 2005 and April 9, 2005 using the following terms: atopic dermatitis, pimecrolimus OR Elidel OR SDZ ASM 981. We searched using the RCT and metaanalysis limits. We also used the systematic review limit in Clinical Queries (PubMed). The RCT limit along with

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

		a limit of studies from Jan 1, 2004 through April 9, 2005 was used for updating. An additional search for adverse events utilized the search terms: pimecrolimus OR Elidel OR SDZ ASM 981 AND included terms for harms: harms, adverse effects, adverse events, adverse reactions, adverse reaction monitoring, ADR, pharmacovigilance (singular and plural as appropriate).
9.	**Patient Population Assessment:** Is the population appropriate for this question? **Your Assessment:**	
	Poor Quality Answer: We included all studies with a control group.	**Good Quality Answer:** We included only studies of patients with condition X as defined by the following criteria in patients ages 18 and older.
10.	**Critical Appraisal:** What is the quality of included studies? ▪ Did the authors use an explicit and quality method for determining validity of individual studies? ▪ Is there more than one author appraising studies? ▪ How were disagreements resolved? ▪ NOTE: The Jadad Scale is frequently employed by reviewers for determining study quality. The Jadad Scale is referred to as a "validated" scoring system; however, it is **not** a good measure of study quality. If the Jadad Scale is used, is there some assurance that the reviewers went beyond the Jadad Scale criteria to critically examine the studies so that only valid and clinically useful studies are used to draw conclusions about efficacy, for example? Our advice is to audit the review. See ***Delfini* Searching & Sources Tool** for recommended approach**.** **Assessment:**	
	Poor Quality Answer: Conclusions are referenced. Comments or notes regarding study designs are included (e.g., whether studies are crossover, double-blind, randomized, single-blind, whether Rx was for atrial fib of onset <24 hours or >24 hours).	**Good Quality Answer:** The authors used validity criteria from the JAMA Users Guides to the Medical Literature. They then applied the Delfini evidence/usability grading scale and excluded all X and U studies (studies with lethal threats to validity or where validity was uncertain or where usefulness was uncertain). They included studies rated A and B (clinically meaningful outcomes with few threats to validity). Two authors reviewed all articles for validity and meaningful clinical significance. Any differences were resolved by discussion and reaching 100 percent consensus.
11.	**Missing Outcomes Data:** Assessment of how loss to follow-up is handled and is it done appropriately? **Your Assessment:**	
	Poor Quality Answer: The authors quantitate the loss to follow-up, but do not discuss how loss to follow-up was handled.	**Good Quality Answer:** Three of 15 studies assessed loss to follow-up and in these studies there was no significant difference in drop-out rates between the groups. All three studies performed an ITT analysis using worst case scenario and in all three instances the outcomes were similar to the completer analysis with statistical significance.
12.	**Homo-/heterogeneity:** If results of the studies were combined, such as in a meta-analyses, did the authors apply tests of homogeneity/heterogeneity to assure that the variation between studies is due to chance (i.e., p-value >.05, similar point estimates, overlapping CI's, etc.)? However, this test is susceptible to problems	

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

	depending upon the number of trials combined. Ideally a test for inconsistency is run — I2 statistic — which reports percent of total variation due to heterogeneity instead of chance: [I2 0-25% is good, to 50% moderate, to 75% not good]. Fixed-effects model assumes each study as the same treatment effect. Random-effects model assumes effects of treatment vary around an overall average treatment effect. Random effects models are often used when greater inconsistencies, but can overvalue small studies. **Your Assessment:**	
	Poor Quality Answer: For studies in which results are combined, the authors do not state how homogeneity/heterogeneity was assessed.	**Good Quality Answer:** Individual studies showed similar results, reflected in the P values of the test of heterogeneity (P 0.99 for vertebral and 0.88 for nonvertebral fractures).
13.	**Combining Results**: If results were combined, was it done in a reasonable and appropriate manner? ▪ If results were combined, were the authors explicit about how they did so and did they employ quality methods? (For example, were authors explicit about how they summarized the data such as in percentages or ratios; did authors make reasonable choices for grouping or stratifying outcomes of interest using such variables as age, duration of treatment, dosage, etc.) ▪ Did more than one author extract and combine data? **Your Assessment:**	
	Poor Quality Answer: The authors do not state how results were combined.	**Good Quality Answer:** Two reviewers extracted data onto an Excel spreadsheet. All of the reviewers were involved in resolving differences through discussion. Data were extracted for all variables reporting at least one of the outcomes of interest (survival to discharge or immediate survival) for patients with and without the characteristic (e.g. the rate of survival to discharge for patients with and without metastatic cancer). If available in the original literature dichotomous outcomes were also presented as continuous variables (i.e. age, haematocrit and serum creatinine levels). If more than one dichotomous cutpoint was used for a variable, both results were extracted. Immediate survival and survival to discharge were plotted against sample size using funnel plots in order to assess the degree of publication bias. The outcome rates were also plotted against the year of publication in order to identify any longitudinal trends. For dichotomous variables summary odds ratios (ORs) were calculated using the DerSimonian and Laird random effects model. For continuous variables summary effect sizes, standard errors (SE) and 95% confidence intervals (CI) were calculated.
14.	**Weighting:** If weighting was employed, was a reasonable approach taken? ▪ Weighting is generally used to favor larger studies or higher quality studies and reduce potential bias from smaller studies or those of lower quality. Be aware, however, that larger studies are not necessarily higher quality so both size and quality need to be considered, and weighting from flawed studies could distort results. ▪ Consider sensitivity analyses where results of higher quality studies are compared with lower quality studies.	

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

	Your Assessment:	
	Poor Quality Answer: The authors weighted the studies by number of deaths.	**Good Quality Answer:** Authors weighted the studies by study size.
15.	**Author's Discussion**: Well executed sensitivity analyses, discussion of limitations, explanations of differences in studies and their results, etc.? **Your Assessment:**	
	Poor Quality Answer: The authors did not provide information about sensitivity analysis or study limitations.	**Good Quality Answer:** We performed two sensitivity analyses. First we excluded the postcoital study (Author X 1990) and then we excluded those studies that included patients who had only two infections in the 12 months prior to enrollment instead of three, and those that had as inclusion criteria "history of recurrent UTI." The overall effect remained unchanged. Limitations of our review stem primarily from including studies of short duration.
16.	**Other Issues (eg, potential conflict of interest):** **Your Assessment:**	
17.	**Author's Conclusion:** Conclusions are supported by the evidence? **Your Assessment:**	
	Poor Quality Answer: The authors state that the evidence suggests benefit from the use of tricyclic antidepressants in preventing postnatal depression.	**Good Quality Answer:** This systematic review found only two studies of antidepressant prophylaxis of postnatal depression. Nortriptyline was not significantly more effective at preventing postnatal depression than placebo, but one small study found sertraline was significantly more effective than placebo at preventing postnatal depression. It is not possible from these two studies to draw any clear conclusions about the effectiveness of antidepressants in preventing postnatal depression. Furthermore, there has been no research into starting antidepressant prophylaxis during pregnancy. Therefore, the evidence does not allow us to make any recommendations about the role of antidepressants in preventing postpartum depression.
18.	**Transparency:** Is sufficient detail provided that enables a through quality assessment of this review and such that this review could be replicated? ▪ Does the review provide a list of the specific studies included for drawing conclusions? **Your Assessment:**	
19.	**Biostatistics:** Do you need a biostatistical consult? **Your Assessment:**	

Delfini Tools
Available at www.delfini.org (http://www.delfini.org/delfiniTools.htm)

Delfini

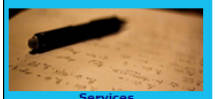
Evidence- and Value-based Solutions for Health Care™
Clinical Improvement Consults, Content Development, Training & Seminars, Tools

The EBM Information Quest: Is it true? Is it useful? Is it usable?™

Delfini Co-founders: Michael E Stuart MD, President & Medical Director . Sheri Ann Strite, Principal & Managing Partner

Quick Picks

Why Critical Appraisal Matters



Services

Seminars

Books

Updates & Contact Info

PRACTICAL EVIDENCE-BASED PRACTICE RESOURCES & SERVICES FOR HEALTHCARE QUALITY

Helping people improve **healthcare decision-making** through **"applied" evidence-based clinical quality improvement** and **communications** with a focus on **value** and **patient-centered care**.

DelfiniGram™: GET ON OUR UPDATE LIST

- Resources
  We are a public service entrepreneurship and make much of our work **freely available** online.
- Solutions
  Read about the problems with healthcare information and our solutions.
- Help
  See our services including evidence reviews, evidence-based clinical improvement training workshops, committee help (P&T, medical technology assessment, quality improvement teams, etc.), clinical guidelines, evidence-based quality improvement project facilitation and more.

To navigate our website—

- See our **Quick Picks** to selected sections to the left.
- See our **Main Menu** to the right.
- Beneath our Main Menu, you will find our **Quick Navigator** to some of our key resources.

Also see Delfini Group Publishing.

Workshops

Critical Appraisal of the Medical Literature for Therapies
Selected Topics
July 10 & 11, 2014 ~ Portland, Oregon

Or we come to you—click here to learn about other workshop topics.

At Home

Read Our Blog...

Menu

Use of our website implies agreement to our Notices. Citations for references available upon request.

Home
What's New
Blog
Seminars
Services
Delfini Group Publishing
Resources
Sample Projects
Notices
About Us & Our Work
Testimonials
Other
Site Search
Contact Info/Updates

Quick Navigator to Selected Resources

- Tools & Educational Library
- The Healthcare Information Problem
- Tutorials
- DelfiniClick™—commentaries
- On the Same Page™—patient decision support
- Evidence Messaging Scripts
- Delfini Publications
- Medical Leader Interviews
- Glossary
- Recommended Reading
- Recommended Web Links
- Delfini Critical Appraisal Flier
- Get DelfiniGram™ Updates

Delfini Blog

DelfiniClick

transparent. Delfini average training program scores are 4.8 out of 5.0.

"... gifted facilitators and passionate educators... have rarely encountered presenters so well-prepared and well-versed in their subject matter... both are engaging, up-beat and empowering... extremely generous with their knowledge and expertise..."

Tools & Educational Library

The EBM Information Quest: Is it true? Is it useful? Is it usable?™

Delfini Co-founders: Michael E Stuart MD, President & Medical Director . Sheri Ann Strite, Principal & Managing Partner

Quick Picks

Why Critical Appraisal Matters

Services

Seminars

Page Navigation.....

1. Important Reading
2. Searching, Sources & Terms Utility Tools
3. Critical Appraisal Tools
4. Critical Appraisal Primers/Papers
5. EB Committee Tools
6. Quality Improvement Project Tools
7. Patient Communication Tools
8. EBM Organization

DelfiniGram™: GET ON OUR UPDATE LIST

Tools & Educational Materials by Delfini

Basics For Evaluating Medical Research Studies:
A Simplified Approach
And Why Your Patients Need You To Know This

Delfini Group Evidence-based Practice Series
Short How-to Guide Book

"Best help with evidence-based medicine available."
Marty Gabica, MD, Chief Medical Officer, Healthwise

This book is about how to evaluate the reliability and

At Tools & Educational Library

Read Our Blog...

Menu

Use of our website implies agreement to our Notices. Citations for references available upon request.

Home
What's New
Blog

Delfini Critical Appraisal Case Study

HYPOTHETICAL CASE STUDY: MYOCEPTIMAB PREVENTS CARDIOVASCULAR MORBIDITY

Critical Appraisers & Date: Sheri A. Strite & Michael E. Stuart MD, Delfini Group; April Fool's Day 2000-Any Year

PUBLISHED ABSTRACT

Background

Elevated myoreactive protein has been demonstrated to be associated with increased risk of myocardial infarction (MI). Myoceptimab is an inhibitor of myoreactive protein and has been shown to reduce myoreactive protein levels.

Methods

We conducted a randomized, double-blind trial in the Beaverton University Heart Care Center to assess the efficacy and safety in patients ages 55 and older who were at increased risk for cardiovascular events and had elevated myoreactive protein levels above 4 mg/L on two separate occasions. Patients were randomly assigned to receive 60 mg of myoceptimab (29 patients) or placebo (35 patients) daily for 6 months.

The study outcomes were the mean difference in levels of myoreactive protein between the groups and mean difference in cardiovascular morbidity between groups. Cardiovascular morbidity was defined as a composite of episodes of new onset angina, admission to the hospital for any cardiovascular-related condition, myocardial infarction, stroke, new onset of claudication, heart failure or cardiovascular death).

Results

At 6 months, active treatment resulted in a 37% reduction of myoreactive protein in the myoceptimab group compared to placebo. Cardiovascular morbidity was n = 19 [65.5%] in the placebo group vs. n = 7 [20%] in the myoceptimab group, ARR 45.5% ; P = 0.0003. Fifty percent more patients in the myoceptimab group reported an increase in quality of life.

Conclusions

Treatment with myoceptimab reduced cardiovascular morbidity and was associated with significant beneficial effects in reported quality of life. Myoceptimab offers a safe and effective therapeutic option for patients who are at increased risk for cardiovascular events.

ADDITIONAL REVIEWER COMMENTS

Safety: No reported differences in safety outcomes.

CRITICAL APPRAISAL

Design

- Primary endpoint: questionable composite

Selection

- Study size: small
- Randomization: not truly randomized; patients assigned to groups by study consent date
- Concealment of allocation: no details
- Baseline characteristics: slightly higher rate of angina in the placebo group

Performance

- Blinding: insufficient details and no indication of blind assessment
- Intergroup differences: participating cardiologists were not restricted in patient management so as to replicate real-world conditions; no details of co-interventions reported between groups

Data Collection/Missing Data

- Attrition: less than 1 percent

Assessment

- Safety, including long term harms, is uncertain
- Results: questionable clinical significance, selective reporting and *post-hoc* results

Grade U: Uncertain

Delfini Critical Appraisal Exercise Template
Superiority Trials

Study: Fictional Primary Randomized Controlled Trial
What a Sample Critical Appraisal, Using the Checklist, Might Look Like

Instructions

Your job is to critically appraise this study for **internal validity**. In order to help you become accustomed to critical appraisal, we have used our short critical appraisal checklist to frame some questions for you to answer to both guide you and to give you an example of how you will wish to mentally frame questions for yourself when conducting such evaluations.

When we do critical appraisal, we usually only grade until we are confident about a study grade. For this exercise, **identify as many threats to validity as you can find.**

Important: While these questions are meant to guide you through a fairly complete review, **do** add comments, questions and observations as you believe relevant for assessing internal validity.

After you have done each section's Bias Assessment, decide how you rate risk of bias for that section and enter your rating in the column to the left. Risk of Bias Ratings: Low, Medium, High, Uncertain. At the end, you will grade the study (or you may wish to grade study outcomes if there is a mix). Grades are A, B, U = uncertain—or BU = probably true, but sufficient uncertainty to make borderline.

Appraisal — Sample answers are in white area below. Feel free to keep your answers short—ours are sometimes a little more lengthy to make a few teaching points.

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Notes: NIH Grant, university directed study

Study Design Assessment	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination used. ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Study Design Risk of Bias Rating: Low	**Bias Assessment** • No biases identified considering items above. **Notes** •
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Selection Risk of Bias Rating: Medium	**Bias Assessment** • Title says randomized; however, no details of randomization provided. Baseline characteristics appear balanced between groups, suggestive that randomization was performed successfully. • No other biases identified considering items above. **Notes** • Allocation to treatment groups was concealed through use of a call-in center. However, it is not known if the center staff had a random method for allocation.
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)

Performance Risk of Bias Rating: Low, but see Assessment	**Bias Assessment** • No biases identified considering items above (however, uncertainty about blinded assessment—see Assessment). **Notes** • A review of side effects shows a reasonable enough balance between groups and so unlikely to become unblinded due to side effects.
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?

Data/Attrition Risk of Bias Rating: Medium	**Bias Assessment**
	• No biases identified considering items above; however, see below.
	Notes
	• Discontinuations were high at 30 percent. Despite this high attrition, however, attrition bias seems unlikely.
	  ○ Randomization was likely to have been achieved, concealment of allocation method was appropriate, blinding at least before assessment appears likely to have been successful, high adherence, low protocol violations, groups were not treated differently except for the intervention, co-interventions were balanced between groups, reasons for discontinuations were balanced between groups (6 categories listed), data imputation was appropriate, and patterns in outcomes make chance unlikely.
	  ○ Given the above, it seems unlikely that a significant number of discontinued people in the comparator group would have had good outcomes, had they completed the study, in sufficient numbers to reverse the results or render statistically significant findings non-significant.

Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values which puts the intervention through a challenging trial or reasonable sensitivity analysis) or missing values are very small. ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Assessment Risk of Bias Rating: Uncertain due to assessment blinding; otherwise low Risk of Chance Results: Low	**Bias Assessment** • No specific mention of blinded assessment or that all working with subjects' data were blinded. • Non-significance in primary outcome of all-cause mortality is probably due to lack of power for that outcome (i.e., rare event may mean that too few people were studied to show a true difference)—that the problem may be a power issue is supported by the confidence intervals which include a clinically meaningful difference. This is also supported by a pattern of statistical significance in the individual outcomes (all pre-specified) of reduction in new onset of heart failure, reduction in non-fatal myocardial infarction plus reduction in hospital admissions for myocardial infarction or heart failure. • No other biases identified considering items above.

	Notes • Efficacy analysis was by intention-to-treat (all subjects by groups). Mixed effects model is appropriate for data imputation. • All key outcomes reported, so appears to be no problems with selective reporting. • Patterns in outcomes make chance unlikely.
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
Efficacy Evaluation **Safety Evaluation**	**Efficacy Results Assessment** • Clinically significant areas with clinically meaningful effect sizes. If the study had been conducted with more people, reduction in mortality is likely. **Safety Assessment** • Population was appropriate for safety (patients not as randomized, but as actually treated provided they had at least one exposure to the study drug). • No significant safety issues reported.
Overall Grade and Summary	**Grade B to BU** **Summary** • Results are clinically meaningful and likely to be true. Confirmatory studies desirable.

JOURNALIA MEDICUS HYPOTHETICALIA

Vol. 57, No. 5, April 1, 2014, pp 1005-1009
Delfini Group Publishing, LLC

Use of Yoroceptimab, a New Trigeminal Nerve Cell Inhibitor, for Prevention of Sphenopalatine Ganglioneuralgia

Stewart ME, Streit SA and Pinglo CP. University of Delphinidae.

BACKGROUND

Optimal management of patients prone to sphenopalatine ganglioneuralgia (SG), commonly referred to as "ice-cream headache" or "brain freeze," is challenging especially among those who are fast eaters or drinkers. Previous surveys have shown that 99% of SG patients would prefer to experience icy treats without the specific pain they associate with “brain freeze” [refs]. Yoroceptimab is a recently developed trigeminal nerve cell inhibitor which has been reported to provide immediate relief from SG in case series [refs]. This is the first RCT designed to evaluate the efficacy of yoroceptimab in preventing SG in susceptible patients who enjoy ice cream and other icy desserts and beverages.

STUDY DESIGN

Double-blind, multicenter, randomized controlled superiority trial.

OBJECTIVES

To compare efficacy and safety of yoroceptimab to placebo in patients highly prone to experiencing SG especially following rapid consumption of frozen beverage or frozen dessert-type treat (collectively referred to as "icy treat"). The primary end point, assessed at 4 weeks was the absolute risk reduction in SG in-clinic episodes at the end of week 4.

METHODS

We conducted a randomized, double-blind trial in 3 centers (details in online appendix), with similar populations, in the US to assess the efficacy and safety of yoroceptimab in patients aged 18 to 66 years of age who had a history of "SG *in extremis*" and were referred for enrollment by participating neuorologists after meeting SG criteria of the International Society of Neurology. Patients with a history of migraine, cluster headache or other headache types were excluded by referring neurologists.

Using a computerized random number generator, a total of 201 patients underwent 1:1 randomization and were assigned, via a centralized call-in center, to receive 5 mg of yoroceptimab or placebo twice weekly, during in-clinic testing, for 4 weeks.

Baseline characteristics were similar in the two groups (Table 1).

Table 1. Baseline Characteristics by Group

Characteristic	Yoroceptimab	Placebo
n	101	100
Age, yrs	41 (18–65)	41 (18–64)
Women (n, %)	47, 47%	48, 48%
White	71%	72%
Black	13%	12%
Hispanic	13%	13%
Asian	2%	1%
Other (unknown)	1%	2%
Patient-estimated frequency of SG episodes per rapidly consumed icy treat (30 seconds or less)	80%	80%
Self-identified rapid eater or drinker	62%	63%

Yoroceptimab and placebo were provided by the manufacturer of yoroceptimab, manufactured to be identical in all ways except for active agency, and were placed in identical capsules. All study subjects, physicians and staff in each center as well as all working with patient data were blinded.

An "SG event" was defined as an SG experience of any severity within 1 minute of any icy treat consumption.

Upon twice weekly visits to the research clinic, patients were escorted into the observation laboratory for purposes of a timed trial with a blinded observer. Patients were instructed to take their medication, observed to take assigned medication by the observer and the time was noted.

Patients were provided with an icy drink (7 ounces, standardized for temperature, texture and taste) in a controlled environment and scheduled exactly 1 hour post-medication intake upon which patients were instructed to consume the entire drink in 30 seconds or less under observed conditions.

Episodes of SG were documented during this clinic visit by both patients and blinded observers, observing reactions and facial expressions, and compared for concordance which was nearly 95 percent.

Safety outcomes were not predefined, except for all-cause mortality—an important outcome for all interventional studies—but included all reported adverse events. A screening panel was obtained at each visit to assess potential organ system/metabolic effects [ref].

Patients were evaluated for the primary endpoint, adherence to the study protocol and side effects at every clinic visit.

RESULTS

Three percent of subjects in the yoroceptimab group and 2 percent of patients in the placebo group failed to complete the trial. Reasons for loss to follow-up or withdrawal were similar between groups (Table 2) except for 1 death in the placebo group due to automobile accident (relationship to study medication, unknown).

Table 2. Disposition of Subjects

Disposition	Yoroceptimab	Placebo
RANDOMIZED	**101**	**100**
TOTAL DISCONTINUED	**3 (3%)**	**2 (2%)**
• Withdrawn consent	2 (2%)	1 (1%)
• Adverse events	0 (0%)	0 (0%)
• Loss to follow-up	1 (1%)	0 (0%)
• Death	0 (0%)	1 (1%)

Accounting for non-completers, of the yoroceptimab patients on-study, a total of 785 potential in-clinic events were possible; and of the placebo patients, a total of 782 potential in-clinic events were possible.

The safety population consisted of subjects who filled at least one prescription of study medication (100 percent of patients) analyzed as treated (equal to all patients as randomized as no patients received incorrect medication).

There was 97% adherence to yoroceptimab and 98% to placebo (non-adherence due to discontinuations).

Study results are summarized in Table 3. The ARR in SG in-clinic episodes was 55% at 4 weeks, P<0.0001.

Table 3. Primary End Point

Absolute risk reduction in SG in-clinic episodes at the end of week 4.	Yoroceptimab n = 101	Placebo n = 100	ARR	P-Value for Difference; 95% CI
	120 episodes	552 episodes	55%	P<0.0001; 95% CI (50.87 to 59.01)

ADVERSE EVENTS

Through 4 weeks, reported adverse events were similar in both groups (Table 4). There were no significant differences with regard to the incidence of clinical or laboratory adverse events, discontinuations due to any adverse events and serious adverse events. No subjects in either group discontinued treatment due to drug-related clinical adverse events.

Table 4. Overall Summary of Safety Events N (%)

Adverse Event	Yoroceptimab n = 101	Placebo n = 100
Number of patients with abnormalities in comprehensive screening metabolic panel*	1 (1%) (1 subject with total protein 8.0—normal range 6.3 to 7.9)	0 (0%)
Number of patients with clinical AEs	0 (0%)	0 (0%)
Treatment-related clinical AEs	0 (0%)	0 (0%)
Discontinuations due to any AEs	0 (0%)	0 (0%)
All-cause mortality	0 (0%)	1 (1%) p=1.0

*Albumin, Alkaline phosphatase, ALT (alanine transaminase), AST (aspartate aminotransferase), BUN (blood urea nitrogen), Calcium, Chloride, CO2 (carbon dioxide), Creatinine, Direct bilirubin, Gamma-GT (gamma-glutamyl transpeptidase), Glucose-serum, LDH (lactate dehydrogenase), Phosphorus, Potassium, Sodium, Total bilirubin, Total cholesterol, Total protein, Uric acid.

DISCUSSION: Omitted for exercise brevity.

CONCLUSIONS

Yoroceptimab offers an effective, clinically meaningful and safe therapeutic option for patients susceptible to SG, or icy brain-freeze, post-rapid icy treat consumption. For those who prefer icy treats *with* pain, placebo is clearly superior. (ClinicalTrials.gov #NCT09125612).

Sponsorship
This study was funded by a grant from Balint Pharmaceuticals, maker of yoroceptimab. The lead author reports serving as a consultant to Balint Pharmaceuticals.

JOURNALIA MEDICUS HYPOTHETICALIA

Vol. 55, No. 2, April 1, 2014, pp 1225-1234
Delfini Group Publishing, LLC

Use of Myoceptimab, a New Smooth Muscle Surface Protein Inhibitor, for Treatment of Patients with Exercise-Related Coronary Artery Angina on Individualized Medical Therapy

Stewart ME, Streit SA and Pinglo CP. University of Delphinidae.

BACKGROUND

Optimal management of patients with multivessel coronary artery disease untreatable by percutaneous coronary intervention (PCI) is challenging. Myoceptimab is a smooth muscle cell proliferation inhibitor previously demonstrated to improve coronary arterial blood flow and provide relief from angina [refs]. This is the first RCT evaluating the efficacy of myoceptimab in patients with angina pectoris untreatable by PCI.

STUDY DESIGN

Double-blind, multicenter, randomized controlled superiority trial.

OBJECTIVE

To compare efficacy and safety of myoceptimab and placebo in patients with exercise-induced angina pectoris and multivessel coronary artery disease untreatable by PCI.

METHODS

We conducted a randomized, double-blind trial in 18 centers, with similar populations, in the US to assess the efficacy and safety of myoceptimab in patients aged 55 and older who had a history of coronary heart disease (CHD), exercise-related angina and who had been determined to be untreatable by PCI by a review panel of cardiologists who had participated in developing the ACC/AHA angina pectoris guideline.

The review boards of all participating institutions approved the study, which was conducted according to the provisions of the Declaration of Helsinki and the Good Practice Guidelines of the International Conference on Harmonization. All patients provided written informed consent to participate in the study.

Patients with a history of heart failure were excluded. Eligible patients had a documented diagnosis of angina pectoris for at least 3 months and experienced at least 3 episodes of angina per week, were on optimal medical therapy as determined by current ACC/AHA clinical guidelines for the management of angina pectoris, were naïve to myoceptimab and had multivessel coronary artery disease untreatable by PCI.

Patients were allowed to take their usual antihypertensive medications and cardiac medications which included aspirin, individualized therapy of nitrates, beta-blockers, ACE or ARB, calcium channel blockers and statins. Efforts to prevent loss to follow-up included an explanation to patients about the importance of keeping in contact with study staff and supportive messages at each follow-up visit. All patients were advised of the importance of providing follow-up information for the duration of the trial regardless of their participation status and to adhere to protocol requirements regarding exercise and diet.

Patients were randomly assigned to receive 60 mg of myoceptimab or placebo daily for 36 weeks.

Previous studies have demonstrated that titration is unnecessary for myoceptimab, and both myoceptimab and placebo were administered as a single daily morning dose. Participating health care professionals in each center were trained in details of the study protocol.

Table 1. Baseline Characteristics by Group

Characteristic	Myoceptimab	Placebo
n	2662	2660
Age, yrs	66 (61–72)	66 (60–71)
Women (n, %)	1065, 40%	1037, 39%
White	71%	72%
Black	13%	12%
Hispanic	13%	12%
Asian	2%	2%
Other (unknown)	2%	2%
BMI (kg/m2)	29 (25–33)	30 (26–34)
Mean no. episodes angina/wk	10	10
Systolic BP, mm Hg	137 (125–148)	136 (124–147)
Diastolic BP, mm	86 (71–95)	87 (72–96)
Current smoker	9%	9%
Co-interventions	Use of aspirin, nitrates, beta-blockers, ACEIs, ARBs, calcium channel blockers and statins were similar*	
Family history of CHD	11%	12%
LDL-C, mg/dl	119 (91–130)	118 (90–129)
HDL-C, mg/dl	46 (35–61)	47 (36–62)
Triglycerides, mg/dl	118 (86–170)	115 (82–164)
Total cholesterol, mg/dl	190 (175–203)	191 (176–204)
Glucose, mg/dl	93 (87–101)	94 (88–102)
Hemoglobin A1C, %	5.7 (5.5–5.8)	5.7 (5.5–5.9)
eGFR, ml/min/1.73 m2	75 (65–87)	74 (65–84)

Data are presented as n, median (interquartile range), or n (%).
*For this exercise data summarized for brevity.

The primary end point, assessed at 36 weeks through self-reports recorded daily in diaries provided to each study subject and medical record evaluation, was the difference between groups in the mean number of anginal episodes per week. Anginal episodes were as defined by the ACC/AHA guidelines.

Secondary predefined endpoints included a combined cardiovascular disease endpoint (cardiovascular mortality, non-cardiovascular mortality, non-fatal myocardial infarction and refractory angina—as defined by the ACC/AHA guidelines—and change from baseline in exercise tolerance on a standard Bruce treadmill test. Results for each component of the secondary combined outcome were also reported.

Safety outcomes were not predefined, but included all reported adverse events.

Patients were evaluated for the primary and secondary endpoints, adherence to the study protocol, co-interventions and side effects every 3 weeks at each medical center.

Evaluations included dosages of all cardiovascular medications including aspirin, beta blockers, nitrates, statins, calcium channel blockers, antihypertensives and all other medications prescribed by their physicians as well as over the counter drugs and supplements.

Alpha was set at < 0.05.

RESULTS

A total of 5322 patients underwent randomization. Baseline characteristics were similar, including medications, dosages, clinical and demographic variables in the two groups (Table 1). Four percent of subjects in both groups failed to complete the trial. Reasons for loss to follow-up or withdrawal were similar between groups (Table 3).

The safety population consisted of subjects who filled at least one prescription of study medication. Three patients in the myoceptimab group and 2 patients in the placebo group received the incorrect study medication. In the safety evaluation, these subjects were analyzed as treated and not as randomized.

Study results are summarized in Table 2. The primary outcome (the mean number of anginal episodes per week) had decreased from 10 to 4 in the myoceptimab group, as compared with a reduction from 10 to 7 in the placebo group, a 57% relative risk (RR) with myoceptimab compared with placebo (P<0.01). The secondary combined outcome measure occurred in 5.5% of the myoceptimab group and 9.5% receiving placebo, relative risk (RR) 0.58, 95% confidence interval 0.51 to 0.73, P<0.001, absolute risk reduction (ARR) 4%, 95% CI 2.53% to 5.36%, number needed to treat (NNT) 25, 95% CI 19 to 40. There was a nonsignificant decrease in cardiovascular mortality with myoceptimab (1.7% compared to 4.3% in the placebo group, (RR) 0.60, P=0.07). There was no significant difference in noncardiac mortality between the groups (3.4% vs. 3.7%, (RR) 0.92,

P=0.26). Myoceptimab was associated with reduced risk of nonfatal myocardial infarction (3% vs 8.7%, (RR) 0.34, P=0.009, NNT 18), refractory angina (5.1% vs 13%, (RR) 0.39, P=0.002, NNT 13).

In addition, exercise tolerance increased from 2.5 minutes to 4.1 minutes in the myoceptimab group as compared to an increase from 2.3 minutes to 3.1 minutes in the placebo group (P<0.01), a 50% relative increase with myoceptimab compared to placebo.

Table 2. Primary and Secondary Outcomes

Outcome Measures at 36 Weeks	Myoceptimab n =2662	Placebo n =2660	RR (%)	P Value
Primary Outcome Measure (mean number of anginal episodes per week)	4	7	0.57	P<0.01
Secondary Combined Outcome Measure (overall mortality, cardiovascular mortality, non-cardiovascular mortality, non-fatal myocardial infarction and refractory angina)	5.5%	9%	0.58	P<0.001
Cardiovascular mortality	1.7%	4.3%	0.60	P=0.07
Non-cardiac mortality	3.4%	3.7%	0.92	P=0.26
Nonfatal myocardial infarction	3%	8.7%	0.34	P=0.009
Refractory angina	5.1%	13%	0.39	P=0.002
Change in exercise tolerance increased	1.6 min	0.8 min	0.50	P<0.001

Table 3. Disposition of Subjects

	Myoceptimab	Placebo
RANDOMIZED	2662	2660
TOTAL DISCONTINUED (4%)	106 (4%)	106 (4%)
• Withdrawn consent	14 (0.5%)	13 (0.5%)
• Adverse events	59 (2%)	53 (2%)
• Loss to follow-up	21 (1%)	26 (1%)
• Protocol deviation	12 (0.5%)	14 (0.5%)

ADVERSE EVENTS

Through 36 weeks, reported adverse events were similar in both groups (Table 4). There were no significant differences with regard to the incidence of clinical or laboratory adverse events, discontinuations due to any adverse events and serious adverse events. No subjects in either group discontinued treatment due to drug-related clinical adverse events.

None of the subjects had consecutive elevations in ALT and AST values >/= 3 times ULN (upper limit of normal). There were no cases of hepatitis, jaundice or other clinical signs of hepatic dysfunction reported. No patients had elevations in CK levels (>/= 5 times ULN) and there were no reported cases of myopathy or rhabdomyolysis. Results of other laboratory tests, including routine serum chemistries, renal and hematologic parameters as well as vital signs and findings on physical examinations revealed no evidence of additional safety concerns with myoceptimab.

Table 4. Overall Summary of Safety Events N (%)

Adverse Event	Myoceptimab n =2661*	Placebo n =2661*
Number of patients with laboratory AEs	24 (0.9%)	24 (0.9%)
Number of patients with clinical AEs	360 (13.5%)	359 (13.5%)
Treatment-related clinical AEs	35 (1.3%)	35 (1.3%)
Serious clinical AEs	38 (1.4%)	34 (1.3%)
Serious treatment-related clinical AEs	0	0
Discontinuations due to any AEs	59 (2.2%)	53 (2.0%)
Discontinuations due to any treatment-related AEs	0	0
Discontinuations due to any serious AEs	0	0
Consecutive 3 x ULN elevations in ALT and/or AST	0	0
CK >/=5 x ULN	0	0

AEs, adverse events; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CK, creatine kinase; n, number of patients receiving agent; ULN, upper limit of normal.
*Myoceptimab randomized = 2662, less 3 receiving placebo in error, plus 2 receiving myoceptimab in the placebo group. Placebo randomized = 2660, less 2 receiving myoceptimab in error, plus 3 receiving placebo in the myoceptimab group.

DISCUSSION

[Excluded for this exercise]

….

CONCLUSIONS

In this trial, treatment with myoceptimab improved cardiovascular outcomes, had significant beneficial effects on anginal events and exercise tolerance, and had an acceptable side-effect pro-

file in patients with stable coronary artery disease on optimal medical therapy. Myoceptimab offers an effective therapeutic option for patients with symptomatic angina who are ineligible for PCI. (ClinicalTrials.gov #NCT91256122).

Sponsorship
This study was funded by a grant from Balint Pharmaceuticals, maker of myoceptimab. The lead author reports serving as a consultant to Balint Pharmaceuticals.

Myoceptimab Compared to Drug Y for Squamous Cell Carcinoma of the Head and Neck

Background

In a trial of 1200 people with stage III and IV squamous cell carcinoma of the head and neck without distant metastases, comparing myoceptimab to drug Y, the primary outcome was overall survival. Patients were assigned 10 mg of either myoceptimab or drug Y daily for 5 courses of treatment. It was prespecified that study would terminate upon reaching 521 events.

Methods

Critical appraisal yields rating of low risk of bias for—

- Selection bias (randomization including no issues with 2:1 ratio, concealed allocation and baseline characteristics)
- Performance bias (blinding, reasonable comparator, adherence, balanced co-interventions, etc.)
- Measurement methods
- Blinded assessment
- Selective reporting

Disposition of 1200 Randomized (2:1 Ratio) Subjects

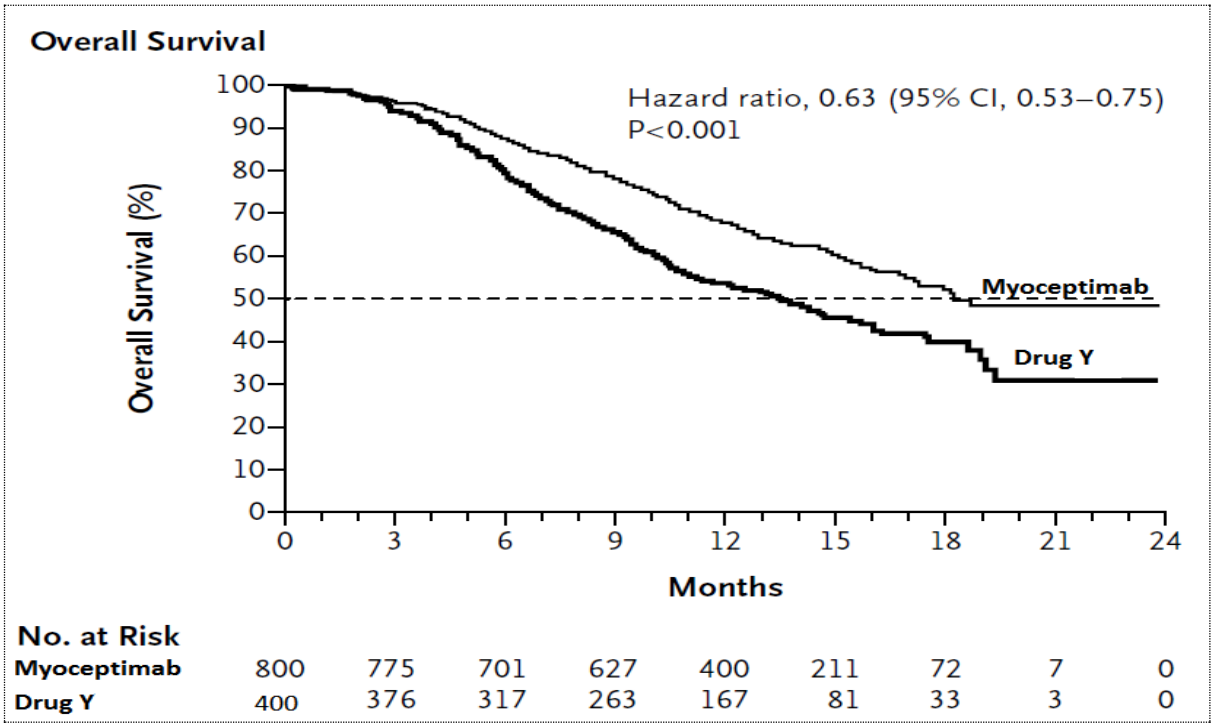
SUBJECT DISPOSITION	Myoceptimab	%	Drug Y	%
Randomized	**800**		**400**	
Included In ITT Efficacy Analysis	**800**	**100%**	**400**	**100%**
Died	**308**	**39%**	**213**	**53%**
Total Discontinued Due to Reasons Other Than Death	**430**	**54%**	**237**	**59%**
· Loss to follow-up	25	3%	12	3%
· Progressive disease	320	40%	180	45%
· Adverse events	61	8%	32	8%
· Withdrawn due to consent/protocol violation	24	3%	13	3%

Censoring rules

If death did not occur before the cutoff date, data were censored at the date the subject was last known to be alive or date lost to follow-up since randomization.

Results

The median overall survival was 18.5 months in the myoceptimab group versus 13.6 months in the drug Y group (hazard ratio for death in the myoceptimab group, HR 0.63; 95% CI, 0.53 to 0.75; P<0.001). 308 subjects in the ITT (intention-to-treat) myoceptimab population died. 213 subjects in the ITT drug Y population died.



1. How do you interpret the statement that 100 percent of patients randomized in each group were included in the ITT efficacy analysis (assume that ***all* results data** are presented in this exercise)?

2. Do the censoring rules seem biased or not? If yes, why? If no, why?

3. Regardless of whether this censoring rule is biased or not, what could be the likely rationale for censoring patient data at the date the subject was last known to be alive if death did not occur before the cutoff date?

4. A subject is enrolled late into the study. He is assigned to the myoceptimab group. He is still alive when 521 events are reached and the study terminated. His total time enrolled in the study is 1.5 months. A death occurs at 2 months. Where are the first subject's data represented on the curve? Where is the death at 2 months represented?

5. Explain what is happening to the patient numbers reported under the curve.

6. Plot median survival on the graph.

7. Might you have made a different choice in time other than median survival to compare the curves and why?

8. Define the HR and describe how you would use this, or would you not use it and why?

9. Discuss varying critical appraisal considerations when evaluating TTE.

Bonus Attrition Question

Attrition in this study is high. High attrition has potential problems related to study size such as greater potential for non-significant results due to diminished power and smaller pool for safety data. When it comes to **efficacy**, however, attrition may not equal attrition bias and simply result in a smaller sample size.

What might you conclude regarding the potential of attrition bias to affect efficacy results in this study?

Keep in mind that there are 4 potential realities:

1. Intervention A > Intervention B
2. Intervention A < Intervention B
3. Intervention A = Intervention B
4. Confounded intervention trumps

Some helpful questions to ask are—

- What would be required for any scenario to be true?
- Were those conditions met?

Therefore, a key question in exploring attrition is, "Considering attrition, what would be required for it to be **false** that myoceptimab is superior in efficacy to drug y in this indication and for this population, and were those conditions met?"

Features of This Study Including Above Risk of Bias Points

1. No confounding treatments included in the analysis;
2. Unlikely imbalanced groups or treatment of groups due to likely successful randomization, effective concealment of allocation and likely success of blinding, as supported by blinding methods plus review of safety data suggesting that adverse events would be unlikely to unblind participants;
3. Very high degree of adherence;
4. Very low incidence of protocol violations;
5. Balance in numbers and reasons for discontinuations in each of the many categories reported with the potential exception of progression of disease—which is informative about the potential efficacy of myoceptimab; and,
6. Patterns through similar outcomes across studies.

Aside from fraud, are there other study features that would have to be present for attrition to result in attrition bias when considering efficacy?

Bombardier C, Laine L, Reicin A, Shapiro D, Burgos-Vargas R, Davis B, Day R, Ferraz MB, Hawkey CJ, Hochberg MC, Kvien TK, Schnitzer TJ; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. VIGOR Study Group. N Engl J Med. 2000 Nov 23;343(21):1520-8, 2 p following 1528. PubMed PMID: 11087881.

BACKGROUND
Each year, clinical upper gastrointestinal events occur in 2 to 4 percent of patients who are taking nonselective nonsteroidal antiinflammatory drugs (NSAIDs). We assessed whether rofecoxib, a selective inhibitor of cyclooxygenase-2, would be associated with a lower incidence of clinically important upper gastrointestinal events than is the nonselective NSAID naproxen among patients with rheumatoid arthritis.

METHODS
We randomly assigned 8076 patients who were at least 50 years of age (or at least 40 years of age and receiving long-term glucocorticoid therapy) and who had rheumatoid arthritis to receive either 50 mg of rofecoxib daily or 500 mg of naproxen twice daily. The primary end point was confirmed clinical upper gastrointestinal events (gastroduodenal perforation or obstruction, upper gastrointestinal bleeding, and symptomatic gastroduodenal ulcers).

RESULTS
Rofecoxib and naproxen had similar efficacy against rheumatoid arthritis. During a median follow-up of 9.0 months, 2.1 confirmed gastrointestinal events per 100 patient-years occurred with rofecoxib, as compared with 4.5 per 100 patient-years with naproxen (relative risk, 0.5; 95 percent confidence interval, 0.3 to 0.6; P<0.001). The respective rates of complicated confirmed events (perforation, obstruction, and severe upper gastrointestinal bleeding) were 0.6 per 100 patient-years and 1.4 per 100 patient-years (relative risk, 0.4; 95 percent confidence interval, 0.2 to 0.8; P=0.005). The incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1 percent vs. 0.4 percent; relative risk, 0.2; 95 percent confidence interval, 0.1 to 0.7); the overall mortality rate and the rate of death from cardiovascular causes were similar in the two groups.

CONCLUSIONS
In patients with rheumatoid arthritis, treatment with rofecoxib, a selective inhibitor of cyclooxygenase-2, is associated with significantly fewer clinically important upper gastrointestinal events than treatment with naproxen, a nonselective inhibitor.

The cost-effectiveness of cyclooxygenase-2 selective inhibitors in the management of chronic arthritis. Spiegel BM, Targownik L, Dulai GS, Gralnek IM. PMID 12755551

Source: Veterans Administration
Greater Los Angeles Healthcare System, David Geffen School of Medicine at University of California, CURE Digestive Diseases Research Center, Los Angeles, CA 90073, USA.

BACKGROUND: Rofecoxib and celecoxib (coxibs) effectively treat chronic arthritis pain and reduce ulcer complications by 50% compared with nonselective nonsteroidal anti-inflammatory drugs (NSAIDs). However, their absolute risk reduction is small and the cost-effectiveness of treatment is uncertain. OBJECTIVE: To determine whether the degree of risk reduction in gastrointestinal complications by coxibs offsets their increased cost compared with a generic nonselective NSAID. DESIGN: Cost-utility analysis.

PERSPECTIVE: Third-party payer.

INTERVENTIONS: Naproxen, 500 mg twice daily, and coxib, once daily. Patients intolerant of naproxen were switched to a coxib.

DATA SOURCES: Systematic review of MEDLINE and published abstracts.

TARGET POPULATION: Patients with osteoarthritis or rheumatoid arthritis who are not taking aspirin and who require long-term NSAID therapy for moderate to severe arthritis pain.

TIME HORIZON: Lifetime.

OUTCOME MEASURES: Incremental cost per quality-adjusted life-year (QALY) gained.

RESULTS OF BASE-CASE ANALYSIS: Using a coxib instead of a nonselective NSAID in average-risk patients cost an incremental 275 809 dollars per year to gain 1 additional QALY.

RESULTS OF SENSITIVITY ANALYSIS: The incremental cost per QALY gained decreased to 55 803 dollars when the analysis was limited to the subset of patients with a history of bleeding ulcers. The coxib strategy became dominant when the cost of coxibs was reduced by 90% of the current average wholesale price. In probabilistic sensitivity analysis, if a third-party payer was willing to pay 150 000 dollars per QALY gained, then 4.3% of average-risk patients would fall within the budget.

CONCLUSIONS: The risk reduction seen with coxibs does not offset their increased costs compared with nonselective NSAIDs in the management of average-risk patients with chronic arthritis. However, coxibs may provide an acceptable incremental cost-effectiveness ratio in the subgroup of patients with a history of bleeding ulcers.

"Does spiral CT screening save lives?"

Background
The lung cancer five-year survival rate (~16%) is lower than many other leading cancer sites, such as the colon (~65%), breast (~90%) and prostate (99%) Reference: U.S. National Institutes of Health. National Cancer Institute: SEER Cancer Statistics Review, 1973-2008.

Methods
In a large collaborative study of 38 community and academic centers in five countries, investigators screened 31,567 asymptomatic persons at risk (history of cigarette smoking, occupational exposure to asbestos, beryllium, uranium, or radon, or exposure to secondhand smoke without having smoked themselves) for lung cancer using low-dose CT. 27,456 repeated screenings were performed 7 to 18 months after the initial screening. Follow-up algorithms depended heavily on follow-up imaging for detection of growth. No further work-up was done for stable lesions. All new lesions were biopsied. Subsequent treatment of cancer, if detected, was at the discretion of patient and center. Patients with lung cancer were followed annually. The duration of follow-up ranged from 1 to 123 months (median, 40).

Kaplan–Meier curves were calculated for lung cancer– specific survival from the date of diagnosis, irrespective of the type of treatment, including no treatment, for all participants with lung cancer, irrespective of the stage of the cancer. A survival curve was also calculated for the subgroup with clinical stage I cancer. Survival curves were also calculated for participants who underwent resection of clinical stage I cancer within 1 month after diagnosis and those who did not receive treatment. A pathology panel reviewed the surgical specimens obtained from participants who underwent resection.

Results
The majority (405) of the cancers were detected at baseline screening. Positive CT results requiring further workup were found in 13 percent (n=4186) of initial scans and 5 percent (n=1460) of annual scans. Lung cancer was identified in 484 patients; 412 (85%) were stage I. Estimated ten-year survival was 80 percent (95% CI 74 - 85) for all patients regardless of stage and treatment and the estimated 10-year survival rate was 88% (95% CI 84 - 91) for stage I cancer. Among the 302 participants with clinical stage I cancer who underwent surgical resection within 1 month after diagnosis, the survival rate was 92% (95% CI, 88 to 95). Operative mortality was 0.5%. The 8 participants with clinical stage I cancer who did not receive treatment died within 5 years after diagnosis.

DISCUSSION
What are possible threats to validity in this study?

JOURNALIA RHEUMATICA HYPOTHETICALIA
Vol. 55, No. 2, April 1, 2008, pp 1225-1234
Rheumatology Research Publications, Inc.

Efficacy and Safety of Gabagabalin in the Treatment of Fibromyalgia: A Randomized, Double-Blind, Placebo-Controlled Trial

Matt H. Romer[1], Harold S. Pinglo[2], Michael E. Twister[3], Martha E. Stuart[4]

Objective: **To assess the efficacy and safety of gabagabalin in patients with fibromyalgia.**

Methods: **A 13-week, randomized, double-blind study designed to compare gabagabalin (1,000–1,750 mg/day) with placebo for efficacy and safety in the treatment of fibromyalgia. The primary outcome measure was the Brief Pain Inventory (BPI) average pain severity score. Response to treatment was defined as a reduction of >=30% in BPI score.**

Results: **Gabagabalin-treated patients were found to have significantly greater improvement in the BPI average pain severity score (*P* <0.016; estimated difference between groups at week 13 = -0.91; 95% confidence interval [-1.74, -0.72]). Response was achieved in 50% gabagabalin-treated patients versus 30% placebo-treated patients (*P* < 0.014). Gabagabalin was also found to significantly improve the BPI average pain interference score, the Fibromyalgia Impact Questionnaire total score, the Clinical Global Impression of Severity, the Patient Global Impression of Improvement and the MOS Short Form 36 vitality score. The Montgomery Asberg Depression Rating Scale improvement difference was not statistically significant. Gabagabalin was generally well tolerated.**

Conclusion: **Gabagabalin (1,000–1,750 mg/day) taken for up to 13 weeks is safe and effective for the treatment of pain and other symptoms associated with fibromyalgia.**

- Affiliations
  [1] Welsely Medical College, Winnetka, Illinois
  [2] Tafts School of Medicine, Wallace, Idaho
  [3] Belmont Park Hospital, Horseville, Louisiana
  [4] Emmett Health System, Forest Grove, Maryland

- Dr. Romer has received consulting fees from Lily, Phozer, Formost Labs and research grants from Asstra-Zeta, Lily and Foremost Labs. Dr Pinglo has received consulting fees from Lily, Phozer, Formost Labs and research grants from Asstra-Zeta, Lily and Formost Labs. Dr. Twister is a principal or coinvestigator on research studies sponsored by the American Fibromyalgia Group, and has received consulting fees from Lily, Phozer, Formost Labs and research grants from Asstra-Zeta, Lily and Formost Labs. Dr. Stuart reported no potential conflicts of interest.

- This study was supported by a grant from Swiss Health. Medications and analytical support were provided by Formost Labs.

- **References removed.**

Efficacy and Safety of Gabagabalin in the Treatment of Fibromyalgia: A Randomized, Double-Blind, Placebo-Controlled Trial

Romer MH, Pinglo HS, Twister ME, Stuart ME

Background
Fibromyalgia is a pain disorder occurring in at least 2% of the US general population and is associated with substantial morbidity and disability. Fibromyalgia for some patients represents a disabling, chronic musculoskeletal pain disorder. The pathophysiology of fibromyalgia has not been clearly defined but recent evidence suggests that fibromyalgia is associated with aberrant central nervous system (CNS) processing of pain impulses.

Gabagabalin, an agent thought to represent similar actions to the neurotransmitter gamma-aminobutyric acid (GABA), has been postulated to be an effective agent for reducing hypersensitivity created by local inflammation or neural dysfunction and has been demonstrated in RCTs to be safe and effective in diabetic neuropathy, postherpetic neuralgia, migraine prophylaxis and other neurological conditions.

Based on these scientific findings and because there has not been to our knowledge a randomized, controlled trial testing the efficacy of gabagabalin in the treatment of fibromyalgia, we studied whether gabagabalin would be safe and effective in reducing pain and other symptoms in patients with fibromyalgia. We used a randomized, double-blind, placebo-controlled, parallel group, flexible-dose study design to assess the safety and efficacy of gabagabalin with a dosing range (usual dosages) of 1,000–1,750 mg/day, administered in 3 doses) in 150 outpatients.

Setting
The study was conducted in 4 large clinical care centers in the US—two academic centers, one large community hospital and one managed care health system. Enrollment began in October 2004, and the study was completed in February 2007. Institutional Review Boards from all centers approved the protocol. Patients were identified by physician referral or were self-referred via a newspaper notice-of-study.

Inclusion Criteria
Female or male patients were eligible for the study if they were >/=18 years of age and met the American College of Rheumatology (ACR) criteria for fibromyalgia and if their Brief Pain Inventory (BPI) score was 4 or greater.

Exclusion Criteria
Patients were not eligible if they had known rheumatic or medical disorders with symptoms that could mimic fibromyalgia symptoms, symptoms resulting from injuries, autoimmune disorders, significant psychiatric illness, psychosis, dementia, substance abuse, capable of pregnancy with lack of acceptable contraception or were breastfeeding. Patient with prior treatment with gabagabalin, gabapentin or pregabalin were not eligible as were patients deemed by their physicians as likely to be treatment refractory. Medication exclusions included sedating drugs e.g., antidepressants (required 30 day washout), all antihistamines, all analgesics except for acetaminophen or NSAIDs.

Study Methods and Details
The study utilized a randomized, double-blind, controlled clinical trial design. Patients meeting entry criteria were randomly assigned to gabagabalin or placebo groups, in a 1:1 ratio. Treatment was double-blind for 13 weeks. Patients were seen weekly for the first 3 weeks. Thereafter, study visits were scheduled at 2-week intervals. Gabagabalin or matching placebo was titrated from 300mg daily to 1,750 mg/day over 7 weeks in a stepwise fashion. If a subject could not tolerate the study dosage of 1,750 mg/day given at bedtime, the dosage was reduced to a minimum of 1,000 mg/day, administered 3 times a day. The study medication dose was held constant for at least the last 4 weeks of the 13 week study. Following the 13 weeks, the dosage was decreased by 300 mg/day until discontinuation.

Patients underwent a physical examination, electrocardiography (EKG), and laboratory tests including hematologic studies, chemistry panel,

urinalysis, serum pregnancy test, urine drug screening, thyroid-stimulating hormone, antinuclear antibody level, erythrocyte sedimentation rate, and rheumatoid factor. At the initial visit, and at prespecified subsequent visits until the end of the 13 week study, the outcome instruments were completed along with checks of vital signs, reviews of adverse events and concomitant medication usage. Weight and height were measured at baseline and at the end of the 13 week period.

The prespecified primary outcome measure was pain severity as measured by the self-reported Brief Pain Inventory (BPI short form), average pain severity score (reference removed), an instrument that assesses average pain severity during the previous 24 hours (0–10 scale, where 0 = no pain and 10 = pain as bad as you can imagine).

Secondary outcome measures included the BPI average pain interference score (0–10 scale, where 0 = does not interfere and 10 = completely interferes), response to treatment defined as a >=30% reduction in the BPI average pain severity score. Fibromyalgia Impact Questionnaire (FIQ), a self-administered questionnaire used to measure health status indicators affected by fibromyalgia over the prior week (total score ranges from 0 to 80, higher scores being more negative impact), the Clinical Global Impression of Severity scale (1–7 scale, where 1 = normal, not at all ill, and 7 =among the most extremely ill patients), the Patient Global Impression of Improvement scale (1–7 scale, where 1 =very much better and 7 =very much worse), the Montgomery Asberg Depression Rating Scale, a 10 item scale completed by clinicians measuring apparent sadness, reported sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. Additional patient-reported health outcomes were measured using the MOS Short Form 36 (SF-36) health survey, which consists of 36 items in 8 health domains (subscales): bodily pain, general health, mental health, physical functioning, role–physical, role–emotional, social function, and vitality.

Statistical Analysis
In this trial, a sample size of 150 subjects was assumed adequate based on previous observational studies to provide a power of 90% or more to detect a 0.60 effect size of gabagabalin in the treatment of fibromyalgia with type 1 error of $\alpha = 0.05$ for the analysis of the BPI average pain severity score. Adjustments were not performed for the secondary measures. We used a longitudinal analysis for the primary analysis comparing the rate of change of the outcome during the treatment period between groups as estimated by random regression methods. We used a model for the mean of the outcome variable. The analyses used all available observations from all time points from patients who completed a baseline evaluation. A secondary analysis, measuring changes from baseline to end point using the last observation carried forward (LOCF) method was conducted. The same model was used for the SF-36 administered at baseline and study completion. Response to treatment and participant ratings of global improvement were analyzed using the Cochran-Mantel-Haenszel test for end point values, using LOCF. Analyses employing LOCF used all available observations of subjects with at least one assessment following enrollment. The primary analysis for all variables was based on the ITT sample, which included observations of participants whether or not they were adherent to study medication treatment. A secondary analysis using only observations from visits while patients were adherent to study medication was also conducted. Treatment effects were tested at a 2-sided significance level of 0.05.

Population
Two hundred and fifty-five patients were screened to identify 150 who were eligible to participate and consented to be enrolled in the study. These 150 patients were randomly assigned to either the gabagabalin (n =75) or the placebo (n =75) group. There were no significant differences between the treatment groups in most demographic or clinical factors. The groups differed significantly in baseline ratings in age, the BPI average pain interference

score and the bodily pain domain of the SF-36 (Table 1).

Thirty-one patients (21%) withdrew during the 13-week trial phase, 18 (24%) from the gabagabalin group and 13 (17%) from the placebo group ($P = 0.42$ by Fisher's exact test). Withdrawals due to adverse events were 12 from gabagabalin group (16%) and 7 from placebo group (9%), lack of efficacy 1 vs 2, loss to follow-up 2 versus 3, withdrawal of consent 2 vs 1 and other 1 vs 0. Of 1,350 possible study visits, the number of visits was 1,212 (90.0%), of which 1112 (82.4% of total possible) were obtained while participants were adherent to study medication treatment. The median dosage at the end point for patients treated with gabagabalin was 1,312 mg/day.

Results

There was a greater mean BPI average pain severity scores reduction over time in the gabagabalin group (Table 2). In the primary longitudinal analysis, the gabagabalin group had a significantly greater improvement in the BPI average pain severity score as compared with the placebo group. There were also significant improvements in the gabagabalin group as compared to placebo in all secondary efficacy measures with the exception of the Montgomery Asberg Depression Rating Scale (Table 2). There was a significant difference in average pain severity score response rates (defined as >=30% reduction from BPI baseline to end point) between patients treated with gabagabalin (38 of 75 [51%]) compared with patients treated with placebo (23 of 75 [31%]) ($P = 0.014$). Gabagabalin was associated with a significantly higher level of global improvement in patient ratings at the end point ($P = 0.001$) as compared with placebo. Of the 8 SF-36 domains listed in Table 1, Vitality was the only domain that improved significantly more in the gabagabalin group ($P = 0.032$).

Safety

Patients treated with gabagabalin reported significantly more dizziness, sedation, lightheadedness, and weight gain than did patients assigned to the placebo group (Table 3). The severity reported for most adverse events was mild to moderate, with no significant group differences in serious adverse events. Further, we found no clinically important safety outcomes in the laboratory results or physical examinations.

Table 1.
Baseline Characteristics and Scores^

Item	Gabagabalin Group	Placebo Group
Age, years	47.2*	42.3
Women (%)	70 (93.3)	65 (86.7)
With major depressive disorder (%)	14 (18.7)	15 (20.0)
With anxiety disorder (%)	8 (10.7)*	6 (8.0)
BPI average pain severity score (0-10)	5.7	6.0
BPI average pain interference score (0-10)	4.7 *	5.3
Fibromyalgia Impact Questionnaire (0-80)	46.3	47.7
CGI Severity scale score (0-7)	4.4	4.5
Montgomery Asberg Depression Rating score (0-60)	15.9	17.1
SF-36 score (range 0-100)		
Physical functioning	47.6	46.1
Role–physical	19.0	11.3
Social functioning	61.7	57.8
Bodily pain	37.0*	32.3
Mental health	67.6	64.3
Role–emotional	60.9	54.2
Vitality	21.7	20.1
General health	52.6	51.3

^Scores except where indicated otherwise, values are the mean values.
*P<0.05 vs placebo

Table 2.

Outcomes in Gabagabalin and Placebo Groups After 13 Weeks of Treatment

Outcome Measure	Gabagabalin (n=57)	Placebo (n=62)	Difference between Groups	
			Estimate (95% CI)	*P*
BPI average pain severity score (0-10)	3.2	4.6	-0.92 (-1.75,-0.71)	0.015
BPI average pain interference score (0-10)	2.2	3.6	-0.81 (-1.56,-0.07)	0.032
Fibromyalgia Impact Questionnaire (0-80)	26.2	37.3	-8.4 (-13.0,-3.3)	0.001
CGI Severity scale score (0-7)	3.1	3.8	-0.66 (-1.08,-0.24)	0.002
Montgomery Asberg Depression Rating score (0-60)	9.1	13.9	-2.79 (-6.13,-0.56)	0.067

BPI=Brief Pain Inventory
CGI=Clinical Global Impression of Severity
Values are the mean values. Difference is mean (week 13 minus baseline) for gabagabalin minus the mean (week 13 minus baseline) for placebo.

Table 3. Adverse Events Reported by >/= 5% of Patients (%)

ADVERSE EVENT	GABAGABALIN (N=75)	PLACEBO (N=75)
Headache	20 (26.7)	16 (21.3)
Nausea	16 (21.3)	16 (21.3)
Lightheadedness	11 (14.7)*	1 (1.3)
Pharyngitis	7 (9.3)	11 (14.7)
Flatulence	6 (8.0)	4 (5.3)
Amblyopia)	5 (6.7	1 (1.3)
Dry mouth	5 (6.7)	3 (4.0)
Dizziness	19 (25.3)*	7 (9.3)
Somnolence	14 (18.7)	6 (8.0)
Insomnia	9 (12.0)	6 (8.0)
Asthenia	6 (8.0)	5 (6.7)
Nervousness	6 (8.0)	1 (1.3)
Anxiety	5 (6.7)	2 (2.7)
Sedation	18 (24.0)*	3 (4.0)
Edema	12 (16.0)	6 (8.0)
Diarrhea	8 (10.7)	5 (6.7)
Depression	6 (8.0)	3 (4.0)
Weight gain	6 (8.0)*	0
Cold virus	5 (6.7)	11 (14.7)

*P<0.05 vs placebo

DISCUSSION

This randomized, double-blind trial of flexible-dose gabagabalin (1,000–1,750 mg/day) versus placebo in the treatment of fibromyalgia demonstrated significantly reduced pain with gabagabalin as compared with placebo. The primary outcome measure was pain as measured by the BPI average pain severity score. Patients taking gabagabalin compared with those taking placebo experienced a significant decrease in their total level of pain interference as measured by the BPI, and a significantly greater proportion of gabagabalin-treated patients compared with placebo treated patients achieved a >=30% reduction in the BPI average pain severity score by the end of 13 weeks. The BPI average pain severity score change represents a validated, meaningful change in pain intensity. Analysis of the secondary outcomes demonstrated that gabagabalin, compared with placebo, significantly improved the vitality domain scores on the SF-36. Therefore, providing gabagabalin to patients with fibromyalgia may result in both pain relief and a significantly improved quality of life. This conclusion is strengthened by the other confirmatory secondary outcomes listed in Table 2, although patients taking gabagabalin did not show significant improvement in the Montgomery Asberg Depression Rating score compared to patients taking placebo.

Like gabapentin and pregabalin, gabagabalin is thought to exert its effects through actions on calcium channels, reducing neurotransmitters such as glutamate, noradrenaline and substance P

involved in nocioceptive neural transmission (ref removed). Our study combined with prior studies of related agents provides substantial evidence that neurotransmitter-modulating agents such as gabagabalin have the potential for significant benefit in patients with fibromyalgia.

Overall, gabagabalin was well-tolerated. Significantly more gabagabalin-treated patients than placebo treated patients reported dizziness, sedation, lightheadedness, and weight gain. Weight gain may have been due to the increased reporting of edema by patients in the gabagabalin group. The adverse effects reported are similar to safety findings in studies of gabagabalin in patients with other pain disorders.

Because there is a continuing debate about the advantages and disadvantages of ITT analysis design compared with analysis of those subjects who remain adherent to assigned treatment, we conducted secondary analyses using a modified ITT design in which we included only outcomes from adherent participants. The results of the secondary analyses did not vary significantly from those of the primary analysis.

Our study had several limitations. This was a 13-week trial and further studies are required to confirm similar results with longer duration of treatment with gabagabalin. Because our trial was small, some non-significant differences between groups may have been due to a lack of power. We were unable to specify a single effective dose of gabagabalin because of the flexible dose-design of the trial although the median dose of gabagabalin was in the usual range for treatment of other chronic pain conditions.

In summary, this double-blind, randomized, placebo-controlled trial is the first trial to show that gabagabalin taken for up to 13 weeks is effective and safe in the treatment of pain and other symptoms associated with fibromyalgia.

AUTHOR CONTRIBUTIONS

Dr. Romer had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs. Pinglo and Twister were involved in designing the study, acquiring data and preparing the manuscript. Dr. Pinglo designed the statistical analysis. Drs Romer, Pinglo, Twister and Stuart participated in the interpretation of data.

Study Summary

Labrie F, et al. Screening Decreases Prostate Cancer Death: First Analysis of the 1998 Quebec Prospective Randomized Controlled Trial. Prostate 1999; 38: 83-89.

Delfini Group ™, LLC

Michael Stuart, MD
President

Sheri Strite, Principal & Managing Partner

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 Summary Tool

Date: 1/25/03 Study Reference: Labrie F, et al. Screening Decreases Prostate Cancer Death: First Analysis of the 1998 Quebec Prospective Randomized Controlled Trial. Prostate 1999; 38: 83-89.
Reviewer: Michael Stuart, MD

Type of study: ☒ **Randomized Controlled Trial** ☐ **Cohort** ☐ **Case-Control** ☐ **Cross-Sectional**

Study Purpose or Hypothesis: To assess the impact of prostate cancer screening on cause-specific death.

Outcomes: Primary: The effect of screening on the incidence of prostate cancer death.

Secondary: Life years gained by early diagnosis and treatment of prostate cancer.

	STUDY CHARACTERISTIC
1.	N = 46,732 men were enrolled and randomized. 46,193 men were determined eligible, following randomization, and studied
2.	Population: All men ages 45-80 registered in the 1985 electoral rolls of Quebec.
3.	Inclusions: Above men traceable in the health registries in Nov 1988.
4.	Exclusions: Men with diagnosis of prostate cancer made before Nov 15, 1988. Men previously screened at Laval University Prostate Cancer Screening Program. The exclusions occurred after randomization.
5.	Power: No mention.
6.	Randomization: Randomly allocated either to the group invited by letter for annual screening or to the control unscreened group at a ratio of 2:1 in favor of screening.
7.	Concealment of Allocation: No information.
8.	Blinding: No blinding.
9.	Intervention or Exposure: Intervention Group: letter inviting annual prostate cancer screening (n=31,300). •Acceptors of annual visit underwent prostate specific antigen (PSA) measurement and digital rectal exam (DRE). • Men with PSA >3.0 ng/ml +/or abnormal DRE underwent transrectal ultrasound (TRUS). •TRUS was performed at f/u if PSA>3.0 or if there was 20% increase above baseline >3.0 in the first year. •Prostate biopsy was performed if TRUS showed hyperechoic abnormality or there was PSA/DRE abnormality. Control Group: Followed according to "current medical practice" (n=15,432).
10.	Data Sources: Death Registry of the Health Dept, Quebec Jan 1, 1989-Dec.31, 1996

Date: 1/25/03 Study Reference: Labrie F, et al. Screening Decreases Prostate Cancer Death: First Analysis of the 1998 Quebec Prospective Randomized Controlled Trial. Prostate 1999; 38: 83-89.
Reviewer: Michael Stuart, MD

11.	**Data Collection Methods: Obtained through existing records**
12.	**Compliance: 23.1% of eligible men were screened**
13.	**Follow-up Length: Exposure to the intervention was calculated from the date of the initiation of the trial for the control group (Nov.15, 1988). Exposure to the intervention for the invited (intervention) group was from the date of their first visit up to the end of 1996.**
14.	**Follow-up Completeness: Not mentioned. Calculations show 10 out of 31,300 missing from intervention group and no subjects missing from the "not-invited" group**
15.	**Reported Protocol Deviations: Not mentioned**
16.	**Adjustments for Possible Confounders: Not mentioned**
17.	**Intention-to-Treat Analysis – (*When patients are omitted from ITT analysis, it is not an ITT analysis even if so called in the article)* : Authors state that, the analysis was made on "an intent-to-treat basis from the time of enrollment," but analysis described in the text was made on the basis of men who were considered "eligible," i.e., after excluding 539 subjects who were determined post-randomization to be ineligible because of previous screening or prior diagnosis of prostate cancer. They refer to this as a "screening effect analysis."**

Reported Results:

Clinical Outcomes Reported in the Text of the Study for 8 Years 1989-1996

Study N = 46,732 men were enrolled and randomized. 46,193 men were determined eligible, following randomization, and studied

Outcome	Screened Men	Unscreened Men
Deaths	**5/8,137**	**137/38,056**
CaP Deaths/ 100,000 Man-Years	**15**	**48.7**
Odds Ratio Favoring Screening and Early Treatment	**3.25 (p<0.01)**	

Date: 1/25/03 Study Reference: Labrie F, et al. Screening Decreases Prostate Cancer Death: First Analysis of the 1998 Quebec Prospective Randomized Controlled Trial. Prostate 1999; 38: 83-89.
Reviewer: Michael Stuart, MD

Author's Conclusions:

"This the first randomized and prospective study on prostate cancer screening shows a 69% decrease in the incidence of deaths due to prostate cancer in the screened compared to the unscreened populations…The data obtained in this study permit, for the first time, to inform men of the estimated risk of death from prostate cancer if not screened and not treated early…Consequently, no valid reason remains to doubt that treatment of clinically localized prostate cancer can prolong survival. In fact, the major benefits observed in the present study in the screened group can only be due to the treatments used."

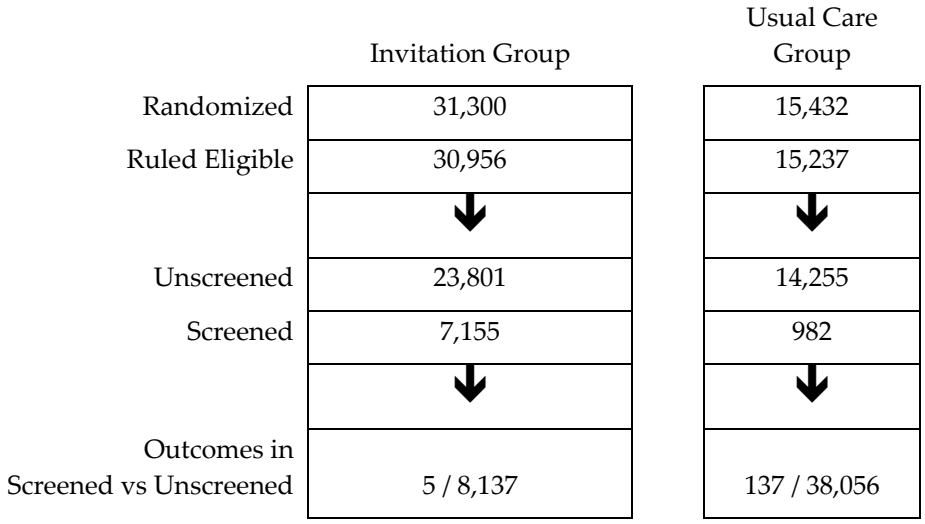
NOTES

Modified CONSORT Diagram – Labrie et al

Randomized to two groups:

Intervention: Receive invitation for screening

Control: No invitation for screening plus usual care



Reported Odds Ratio favoring screening and early treatment = 3.25, p<0.01

EXERCISE

Read the following abstract. Be prepared to discuss your findings about validity and clinical usefulness.

ABSTRACT

Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein.

BACKGROUND

Increased levels of the inflammatory biomarker high-sensitivity C-reactive protein predict cardiovascular events. Since statins lower levels of high-sensitivity C-reactive protein as well as cholesterol, we hypothesized that people with elevated high-sensitivity C-reactive protein levels but without hyperlipidemia might benefit from statin treatment.

METHODS

We randomly assigned 17,802 apparently healthy men and women with low-density lipoprotein (LDL) cholesterol levels of less than 130 mg per deciliter (3.4 mmol per liter) and high-sensitivity C-reactive protein levels of 2.0 mg per liter or higher to rosuvastatin, 20 mg daily, or placebo and followed them for the occurrence of the combined primary end point of myocardial infarction, stroke, arterial revascularization, hospitalization for unstable angina, or death from cardiovascular causes.

RESULTS

The trial was stopped after a median follow-up of 1.9 years (maximum, 5.0). Rosuvastatin reduced LDL cholesterol levels by 50% and high-sensitivity C-reactive protein levels by 37%. The rates of the primary end point were 0.77 and 1.36 per 100 person-years of follow-up in the rosuvastatin and placebo groups, respectively (hazard ratio for rosuvastatin, 0.56; 95% confidence interval [CI], 0.46 to 0.69; P<0.00001), with corresponding rates of 0.17 and 0.37 for myocardial infarction (hazard ratio, 0.46; 95% CI, 0.30 to 0.70; P=0.0002), 0.18 and 0.34 for stroke (hazard ratio, 0.52; 95% CI, 0.34 to 0.79; P=0.002), 0.41 and 0.77 for revascularization or unstable angina (hazard ratio, 0.53; 95% CI, 0.40 to 0.70; P<0.00001), 0.45 and 0.85 for the combined end point of myocardial infarction, stroke, or death from cardiovascular causes (hazard ratio, 0.53; 95% CI, 0.40 to 0.69; P<0.00001), and 1.00 and 1.25 for death from any cause (hazard ratio, 0.80; 95% CI, 0.67 to 0.97; P=0.02). Consistent effects were observed in all subgroups evaluated. The rosuvastatin group did not have a significant increase in myopathy or cancer but did have a higher incidence of physician-reported diabetes.

CONCLUSIONS

In this trial of apparently healthy persons without hyperlipidemia but with elevated high-sensitivity C-reactive protein levels, rosuvastatin significantly reduced the incidence of major cardiovascular events.

REFERENCE

Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ, Koenig W, Libby P, Lorenzatti AJ, MacFadyen JG, Nordestgaard BG, Shepherd J, Willerson JT, Glynn RJ; JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med. 2008 Nov 20;359(21):2195-207. Epub 2008 Nov 9. PubMed PMID: 18997196.

DIRECTIONS

Read abstract and be prepared to discuss. (Note: This abstract has been modified for purposes of this exercise.)

Donnelly BJ, Saliken JC, Brasher PM, Ernst SD, Rewcastle JC, Lau H, Robinson J, Trpkov K. **A randomized trial of external beam radiotherapy versus cryoablation in patients with localized prostate cancer.** Cancer. 2010 Jan 15;116(2):323-30. PubMed PMID: 19937954.

MODIFIED ABSTRACT

BACKGROUND: Localized prostate cancer can be treated several different ways, but head-to-head comparisons of treatments are infrequent. The authors of this report conducted a randomized, unblinded, **noninferiority trial** to **compare cryoablation with external beam radiotherapy in these patients.**

METHODS: From December 1997 through February 2003, 244 men with newly diagnosed localized prostate cancer were assigned randomly to receive either cryoablation or radiotherapy (122 men in each arm). All received neoadjuvant antiandrogen therapy. The primary endpoint was disease progression at 36 months based on a trifecta definition: 1) radiologic evidence of metastatic disease, or 2) initiation of further antineoplastic therapy, or 3) biochemical failure. Two definitions of biochemical failure were used: 1) 2 consecutive rises in prostate-specific antigen (PSA) with a final value >1.0 ng/mL, and 2) a rise above PSA nadir + 2 ng/mL. Non-inferiority was defined as a disease progression rate of 10% or less in disease progression at 36 months.

RESULTS: The median follow-up was 100 months. Disease progression at 36 months was observed in 23.9% (PSA nadir + 2 ng/mL, 17.1%) of men in the cryoablation arm and in 23.7% (PSA nadir + 2 ng/mL, 13.2%) of men in the radiotherapy arm. No difference in overall or disease-specific survival were observed.

CONCLUSIONS: The observed difference in disease progression at 36 months was small, 0.2%. Based on disease progression rates we conclude that cryoablation is noninferior to external beam radiation for the treatment of localized prostate cancer.

STOP. Do not turn page until instructed.

A Hypothetical Clinical Trial Comparing Acupuncture to Sham Acupuncture for Pain Relief in Patients with Lumbar Disc Protrusion Surgery

Hypothesis: classical acupuncture is more effective than sham acupuncture in post-operative pain relief in patients undergoing surgery for lumbar disc protrusion

Patients: 100 post-surgery for lumbar disc protrusion

Diagnosis: History, exam, CT and MRT findings

Primary Outcome Measure: Difference in visual analogue ratings as measured on a 5-point visual analogue scale. Pain was rated as 0 for pain free, 1 for mild, 2 for moderate, 3 for severe, and 4 for very severe pain. A reduction in the pain scale was considered positive if the pain scale at each time point was at least 1 category less severe than at the beginning of the study.

Blinding: Subjects and all working with subjects and subject data were blinded except for acupuncturist.

Intervention: Each patient was administered acupuncture or sham acupuncture in identical fashion by the same experienced physician. True acupuncture points we used for the De-Qi sensation. For the sham procedure, locations 2cm away from actual points were needled so that no De-Qi sensation would occur.

Schedule for Intervention:
Day 1 (immediately following post-operative recovery from surgery): acupuncture
Day 2: sham procedure
Day 3: acupuncture
Day 4: sham procedure

Allowed co-interventions: all patients received the same pain medication at same dose and intervals

Measurements: The visual analogue scale was administered by an independent examiner and was used to assess pain intensity before and after acupuncture at the following intervals 60 minutes, 3 hours, 8 hours, 24 hours. Measurements were conducted between the examiner and the patient solely with no one else allowed in the room during the assessments.

Patient Disposition:
Of the 100 patients enrolled in the study, 89 completed assessments through Day 5.

Findings:
Visual analogue ratings were not statistically different at any measurement point.

Authors Conclusions: Classical acupuncture was not found to be superior to sham acupuncture in the management of pain of patients undergoing lumbar disc protrusion surgery.

Delfini Mini-Case
Intention-to-Treat Reanalysis

In a study comparing myoceptimab to placebo for treatment of mystery Disease X, investigators reported the following information about study completion and patient disposition (patients were not double counted in any category):

Patient Disposition	**Myoceptimab**	**Placebo**
Randomized	927	928
Completed Study 86%	801	802
Discontinued 14%	126	126
Died	2	3
Withdrew consent	17	15
Disease progressed	32	33
Sought other treatment	35	47
Adverse event	18	7
Lost to follow-up	22	21

In an intention-to-treat analysis that analyzed all patients as randomized and imputing values for discontinueds using the respective completer event rate for each group, Myoceptimab was reported to be superior to placebo in curing Disease X (ARR 8.96%, 95% CI (6.10% to 11.82%), P = 0.0001).

- Out of 801 completers, 127 (16%) patients were reported as cured in the myoceptimab group. Assigning the experimental event rate resulted in 20 additional patients, for a total 147 analyzed as cured.
- Out of 802 completers, 55 (7%) patients were reported as cured in the placebo group. Assigning the control event rate resulted in 9 additional patients, for a total 64 analyzed as cured.

Analysis Table	**Myoceptimab**	**Placebo**
Randomized	927	928
Cured	**127**	**55**
Discontinued 14%	126	126
Imputed Cured	**20**	**9**
Total Analyzed as Cured	**147**	**64**

Online calculators are available from GraphPad.* If you were going to compute the p-value and confidence interval yourself using these online calculators, you would enter the following data in their online 2 x 2 table, which would look like this:

	Good Outcome (Cured)	Bad Outcome (Not Cured)
control (placebo)	64	864
experiment (myoceptimab)	147	780

You decide that you wish to perform some of your own sensitivity analyses to recompute the cured/not cured outcomes. Discuss what choices you might make to test the strength of the association. Meaning, what might be some other choices for imputing missing data and testing statistical significance? Pick your favorite based on the context of the information you have here. Prepare to describe your choice and your reasoning. Then complete a new 2 x 2 table using your favored method:

	Good Outcome (Cured)	**Bad Outcome** (Not Cured)	Double Check Your Total N
control placebo			
experiment myoceptimab			

* Confidence interval calculator: http://graphpad.com/quickcalcs/NNT1.cfm
P-value calculator: http://graphpad.com/quickcalcs/contingency1.cfm

Delfini Evidence Tool Kit

Short Critical Appraisal Checklist: Interventions for Prevention, Screening & Therapy

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

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Study Design Assessment **POTENTIAL EXCEPTION: ALL-OR-NONE RESULTS**	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** (allocation by minimization may be acceptable) ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?
Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ If variables are dichotomous, **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values). (May not be an issue if missing values are very few.) ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
External Validity	**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**.
Patient Perspective	☐ Consider benefits, harms, risks, costs, uncertainties, alternatives and satisfaction
Provider Perspective	☐ Satisfaction, acceptability (includes adherence issues, potential for abuse, dependency issues), likely appropriate application and actionability (e.g., FDA approval, affordability, external relevance, circumstances of care, able to apply, tools available)

☐ **Non-Inferiority & Equivalence Supplement:** Absence of the following problems: lack of sufficient evidence confirming efficacy of referent treatment; study not sufficiently similar to referent study; inappropriate Deltas; or significant biases or analysis methods which would tend to diminish an effect size (e.g., conservative application of ITT analysis, insufficient power, etc.)

☐ **Diagnostic Test Supplement:** New test requires better outcomes or value. Test is compared to gold standard or reasonable comparator and finds same abnormality and within time period that does not result in a change in diagnosis. Test is applied to all or random sample of subjects with and without disease. Assessors are blinded. There is minimal bias from indeterminate results. Measures of test function are useful.

☐ **Screening Supplement:** Early diagnosis and treatments determined to be effective will improve outcomes more than later diagnosis and treatment. Beneficial outcomes are not explained by bias (e.g., lead time, length, overdiagnosis or volunteer bias).

Delfini Critical Appraisal Checklist Exercise Worksheet: Superiority Trials

Instructions

Your job is to critically appraise this study for **internal validity**. In order to help you become accustomed to critical appraisal, we have used our short critical appraisal checklist to frame some questions for you to answer to both guide you and to give you an example of how you will wish to mentally frame questions for yourself when conducting such evaluations.

When we do critical appraisal, we usually only grade until we are confident about a study grade. For this exercise, **identify as many threats to validity as you can find.**

Important: While these questions are meant to guide you through a fairly complete review, **do** add comments, questions and observations as you believe relevant for assessing internal validity.

Your Appraisal

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Notes:

Study Design Assessment **POTENTIAL EXCEPTION: ALL-OR-NONE RESULTS**	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Study Design Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** (allocation by minimization may be acceptable) ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Selection Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)
Performance Risk of Bias Rating:	**Bias Assessment**

	• **Notes** •
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?
Data/Attrition Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ If variables are dichotomous, **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values). (May not be an issue if missing values are very few.) ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Assessment Risk of Bias Rating: **Risk of Chance Results:**	**Bias Assessment** • **Notes** •
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
Efficacy Evaluation: **Safety Evaluation:**	**Efficacy Results Assessment** • **Safety Assessment** •
Overall Grade and Summary	

Delfini Critical Appraisal Checklist Exercise Worksheet: Superiority Trials

Instructions

Your job is to critically appraise this study for **internal validity**. In order to help you become accustomed to critical appraisal, we have used our short critical appraisal checklist to frame some questions for you to answer to both guide you and to give you an example of how you will wish to mentally frame questions for yourself when conducting such evaluations.

When we do critical appraisal, we usually only grade until we are confident about a study grade. For this exercise, **identify as many threats to validity as you can find.**

Important: While these questions are meant to guide you through a fairly complete review, **do** add comments, questions and observations as you believe relevant for assessing internal validity.

Your Appraisal

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Notes:

Study Design Assessment **POTENTIAL EXCEPTION: ALL-OR-NONE RESULTS**	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Study Design Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** (allocation by minimization may be acceptable) ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Selection Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)
Performance Risk of Bias Rating:	**Bias Assessment**

	• **Notes** •
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?
Data/Attrition Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ If variables are dichotomous, **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values). (May not be an issue if missing values are very few.) ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Assessment Risk of Bias Rating: **Risk of Chance Results:**	**Bias Assessment** • **Notes** •
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
Efficacy Evaluation: **Safety Evaluation:**	**Efficacy Results Assessment** • **Safety Assessment** •
Overall Grade and Summary	

Delfini Critical Appraisal Checklist Exercise Worksheet: Superiority Trials

Instructions

Your job is to critically appraise this study for **internal validity**. In order to help you become accustomed to critical appraisal, we have used our short critical appraisal checklist to frame some questions for you to answer to both guide you and to give you an example of how you will wish to mentally frame questions for yourself when conducting such evaluations.

When we do critical appraisal, we usually only grade until we are confident about a study grade. For this exercise, **identify as many threats to validity as you can find.**

Important: While these questions are meant to guide you through a fairly complete review, **do** add comments, questions and observations as you believe relevant for assessing internal validity.

Your Appraisal

General: Note sponsorship, funding and affiliations, recognizing that any entity or person involved in research may have a bias.

Notes:

Study Design Assessment **POTENTIAL EXCEPTION: ALL-OR-NONE RESULTS**	☐ **Is the design appropriate to the research question? Is the research question useful?** ☐ For **efficacy**, use of **experimental study design** (meaning there was no choice made to determine intervention) ☐ **Clinically significant area** for study (morbidity, mortality, symptom relief, functioning and health-related quality of life) and reasonable **definitions for clinical outcome such as response, treatment success or failure** ☐ If **composite endpoints** used, reasonable combination ☐ Ensure **prespecified** and **appropriate** 1) research questions, 2) populations to analyze, and 3) outcomes
Study Design Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Internal Validity Assessment: Can bias, confounding or chance explain the study results? See below	
Selection Bias	☐ Groups are **appropriate** for study, of appropriate size, **concurrent** and **similar** in **prognostic variables** ☐ Methods for generating the group assignment sequence are truly **random,** sequencing avoids potential for anyone **affecting assignment** to a study arm and **randomization remains intact** (allocation by minimization may be acceptable) ☐ **Concealment of allocation** strategies are employed to prevent anyone affecting assignment to a study arm
Selection Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Performance Bias	☐ **Double-blinding** methods employed (i.e., subject and all working with the subject or subject's data) and achieved ☐ Reasonable **intervention** and reasonable **comparator** used (e.g., placebo) ☐ **No bias or difference, except for what is under study, between groups during course of study** (e.g., intervention design and execution, care experiences, co-interventions, concomitant medication use, adherence, inappropriate exposure or migration, cross-over threats, protocol deviations, study duration, changes due to time etc.)
Performance Risk of Bias Rating:	**Bias Assessment**

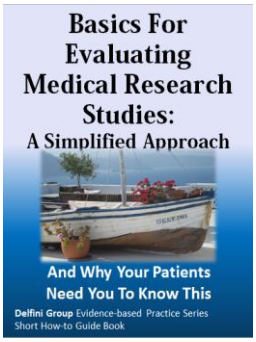
	• **Notes** •
Data/Attrition Bias	☐ Evaluate bias in **measurement activities** ☐ Might **attrition**, including missing data, discontinuations or loss to follow-up, have resulted in distorted outcomes?
Data/Attrition Risk of Bias Rating:	**Bias Assessment** • **Notes** •
Assessment Bias & Chance Assessment	☐ Assessors are **blinded** ☐ Low likelihood of findings due to **chance, false positive and false negative outcomes** ☐ **Non-significant findings** are reported, but the **confidence intervals include clinically meaningful differences** ☐ If variables are dichotomous, **Intention-to-Treat Analysis (ITT)** performed for efficacy (**not safety**) (all people are analyzed as randomized + reasonable method for imputing missing values). (May not be an issue if missing values are very few.) ☐ If **time-to-event analysis** performed, appropriate, transparent and unbiased. Evaluate **censoring** rules. ☐ **Analysis methods** are appropriate and use of **modeling** only with use of reasonable assumptions ☐ No problems of **selective reporting or selective exclusion of outcomes**
Assessment Risk of Bias Rating: **Risk of Chance Results:**	**Bias Assessment** • **Notes** •
Usefulness & Other Considerations	
Meaningful Clinical Benefit	☐ Clinically significant **area** + sufficient benefit **size** = meaningful clinical benefit (consider efficacy vs effectiveness) ☐ **Safety** (caution re: new interventions, caution re: non-significant findings)
Efficacy Evaluation: **Safety Evaluation:**	**Efficacy Results Assessment** • **Safety Assessment** •
Overall Grade and Summary	

Delfini Group Publishing

Delfini Evidence-based Practice Series Guide Books

"Best help with evidence-based medicine available."
Martin Gabica MD, Chief Medical Officer, Healthwise

- **Credit Cards—www.delfinigrouppublishing.com or click on Books to the left at www.delfini.org.**
- **Avoid shipping charges by cash or check. If your shipping address and phone number are not on your check, please write them on the reverse side of this form. Please print legibly! *If we can read it*, your book will reach you in 3 to 4 weeks from Createspace.**



$15 x # copies =

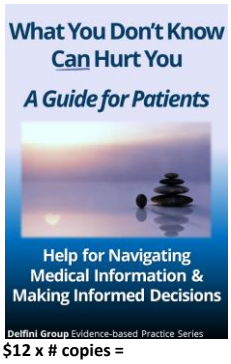
Kindle version $9.99 at Amazon.

"Best help with evidence-based medicine available." *Martin Gabica MD, Chief Medical Officer, Healthwise*

"Highly recommended!..." *John C. Pezzullo, PhD, Biostatistician and Author of Biostatistics for Dummies*

"I am full of admiration for this terrific little book on evaluating medical research studies which is written clearly, simply and appropriately for a starter audience..." *Richard Lehman, MA, BM, BCh, MRCGP, Senior Research Fellow, Oxford, and Blogger, BMJ Journal Watch*

"This book provides a great introduction and guide for anyone who wants to understand how to interpret clinical research but feels intimidated by science or statistics...." *Brian S. Alper, MD, MSPH, FAAFP, Editor-in-Chief, DynaMed*

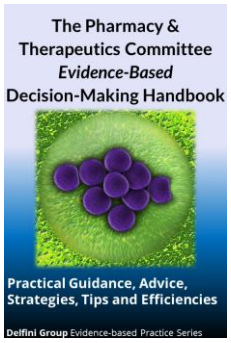


$12 x # copies =

Kindle version $5.99 at Amazon.

From a patient reader: "**A lovely, great work. This is absolutely fantastic information for patients!** As a non-clinical person, I was able to read this without feeling that it was over my head and it held my interest until the very end. I already feel like it is something that I would want to re-read again and again as a resource for my lifelong health care journey. This is something people can draw from throughout their lives. I plan to live with it and really use it, write in it, highlight it and take it along to medical appointments to make sure I get the most from my provider visits."

"I am delighted to see this strong addition to the Delfini library. The authors have created useful models for communicating about health concerns, the ideal patient-physician encounter and decision-making in the world of health care. The content, appendices and examples are strong and relevant." *Gary Schwitzer, Publisher, Healthnewsreview.org*

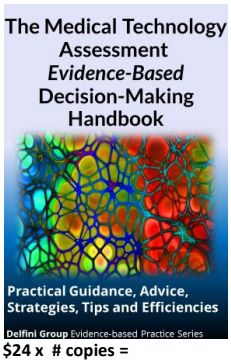


$24 x # copies =

"**The guidelines and tools that you provided have been invaluable** and are used on an almost daily basis as I review journal articles..." *Maribeth M. Bettarelli, PharmD*

"The Delfini Group provided a comprehensive—and practical—approach to interpreting EBM, but most importantly, they provided a tool that I was able to easily incorporate into my practice..." *Chris McFarland, PharmD*

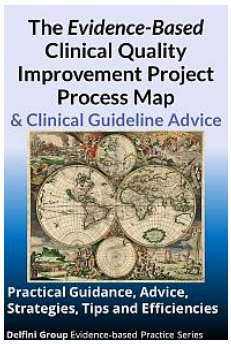
"Delfini has played an instrumental role in the enhancement of our P & T process. Their programs mix together an enthusiasm for evidence-based medicine with factual market-driven ideas to efficiently utilize the available resources..." *Nicole O'Kane, PharmD*



$24 x # copies =

"**Thanks for shouting out to the Washington State Health Technology Clinical Committee**. Along with the help of organizations like Delfini providing us with superb evidence reviews, we are showing a way that the US should follow, in my opinion." *Brian R. Budenholzer, MD, FAAFP, (Former) Health Technology Clinical Committee Chair, Health Care Authority, State of Washington*

"Ours was an important project, and it was successful because each of the members of the workgroup was engaged in the process and participated fully. Having facilitators like Mike and Sheri really made the difference. They guided the group through the process and made the work of EBM easy and fun." *Karen Ching, MD, EBM Director & Nephrologist, Kaiser Permanente Hawaii*

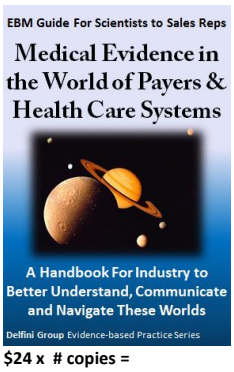


Pre-order: $24 x # copies =

Health care visionaries have had this to say about guidelines accomplished under Mike's direction —

"Always a leader, Group Health has managed to create rigorous, evidence-based guidelines that embody the things I advocate. They use balance sheets to evaluate the benefits, harms and costs of treatments, and use the guidelines in a real-life setting. Anyone who uses these guidelines can expect to achieve better decision-making and improved-outcomes." *David Eddy, MD, PhD*

"Group Health Cooperative's Clinical Practice Guidelines are among the very best available. They are carefully developed, thoroughly documented, assiduously updated, and presented in an easy-to-use form. They provide a superb example of what 'evidence-based medicine' can and should mean." *Donald M. Berwick, MD*



$24 x # copies =

Payer & Health Care Systems Readers React
"I love the mission of this book. The content is great and important. We so need this. This is amazing...." "I absolutely love the book. The book very easy to read with tons of important info!!! I love love love it." "There is a huge gap as you well know, so all I can say is YEAH! All of the questions are on target." "I really like this!"

Industry Readers React
"I love the content; love the style!" "I love it. I am very happy you have taken this on and targeted it to educating industry. Congratulations!" "I think it is fantastic....I do believe that the tide is turning, and has been for years..." "Your mission is great!!! This book will be insightful to all parties..."

Name: _____ **Email:** _____

Address: _____ **Phone:** _____

Sheri Ann Strite & Michael E. Stuart MD
books: www.delfinigrouppublishing.com + resources: www.delfini.org + email: delfini@delfini.org