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:**	

Use of this tool implies agreement to the legal terms and conditions at www.delfini.org.

© 2002-2019 Delfini Group, LLC. All Rights Reserved Worldwide.

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. ▪ 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, excepting in certain instances of all-or-none results (e.g., very large response rates). **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

Use of this tool implies agreement to the legal terms and conditions at www.delfini.org.
© 2002-2019 Delfini Group, LLC. All Rights Reserved Worldwide.

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

		systematic review limit in Clinical Queries (PubMed). The RCT limit along with 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.,	

Use of this tool implies agreement to the legal terms and conditions at www.delfini.org.

© 2002-2019 Delfini Group, LLC. All Rights Reserved Worldwide.

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

	p-value >.05, similar point estimates, overlapping CI's, etc.)? However, this test is susceptible to problems 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.	

Use of this tool implies agreement to the legal terms and conditions at www.delfini.org.

© 2002-2019 Delfini Group, LLC. All Rights Reserved Worldwide.

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:**	

Your Overall Assessment:

Use of this tool implies agreement to the legal terms and conditions at www.delfini.org.

© 2002-2019 Delfini Group, LLC. All Rights Reserved Worldwide.