Fresno Test of Evidence Based Medicine

Grading Rubrics (Form A)

The practice of Evidence-Based Medicine (EBM) involves some basic knowledge and skills related to searching and evaluating medical literature. This UCSF-Fresno Medical Education tool is designed to assess the level at which you are already utilizing EBM skills. Please complete the entire test in one sitting. There are 7 short answer questions, 2 questions that require a series of mathematical calculations, and three fill-in-the-blank questions. Allow yourself at least 30 minutes to complete the test.

Answer questions 1-4 based on the following clinical scenarios:

- You have just seen Lydia who recently delivered a healthy baby. She plans to breastfeed, but also wants to start oral contraception. You generally prefer to prescribe combination oral contraceptives (estrogen + progesterone) but you have been told that these might more negatively affect her breastmilk production than progesterone only pills.
- John is an 11 year old boy who presents with primary enuresis. He has grown frustrated with the inconvenience and embarrassment of his problem. You have excluded the possibility of urinary tract anomalies and infection as possible causes. You consider recommending a bedwetting alarm, but a colleague tells you he thinks they're "worthless" and suggests that you treat with imiprimine or desmopressin.

1. Write a focused clinical question for each of these patient encounters that will help you organize a search of the clinical literature for an answer and choose the best article from among those you find.

   Scoring Rubric for breast-feeding/contraception question. (When in doubt, consider whether what is written will contribute to an optimally specific search of the clinical literature.)

	Population	Intervention	Comparison	Outcome
Excellent (3 pts)	Multiple relevant descriptors e.g., "post partum woman," "breast feeding/lactating mother" or "breastfeeding mom desiring contraception," or "breast	Includes specific intervention of interest; e.g. *combined* contraceptives (estrogen and	Identifies specific alternative of interest since pt. wants to use *oral* contraception e.g. *progesterone only contraception*	Outcome that is objective and meaningful to patient e.g. infant growth rate, number of lactation "drop outs," or maternal satisfaction with infant satiety or milk flow

	fed newborn" *Note: "breastfeeding woman" is considered two descriptors.*	progesterone), or specific individual components of contraception such as "estrogen"		
Strong (2 pts)	One appropriate descriptor as above examples e.g. "woman," or "infant" or "breastfeeding"	Mentions *contraception* or type of intervention, e.g. *oral contraceptives* , or *hormones*	Mentions a specific comparison group e.g. placebo, or a specific form of contraception, or mothers not taking OCP's	Non-specific outcome e.g. "milk" or "breast feeding" OR Disease oriented outcome such as milk volume without accompanying measure of clinical relevance e.g. "milk volume" or "chemical composition of milk" or "breastmilk production"
Limited (1 point)	A single general descriptor unlikely to contribute to search e.g. "patient"	Mentions intervention but unlikely to contribute to search e.g. "methods" "options" "treatments"	Mentions comparison but unlikely to contribute to search e.g. "compared to other methods" (Note: Using a plural non-specific term, e.g. "various treatment options," should only be counted once, in the Intervention column)	Reference to outcome, but so general as to be unlikely to contribute to search e.g. "effects" "change the outcome"
Not Evident (0 pts)	None of the above present	None of the above present	None of the above present	None of the above present

Scoring Rubric for bedwetting question. (When in doubt, consider whether what is written will contribute to an optimally specific search of the clinical literature.)

	Population	Intervention	Comparison	Outcome
Excellent (3 pts)	Multiple relevant descriptors e.g., "boy with primary enuresis" specific age group, gender, exclude infection or anatomic anomalies *Note: "primary enuresis" is considered two descriptors.*	Specific intervention of interest; e.g. bedwetting alarm	Identifies specific alternative treatment of interest e.g. "desmopressin acetate" or "imipramine" or "anti-depressants"	Outcome that is objective and meaningful to patient e.g. dry nights
Strong (2 pts)	One appropriate descriptor as above examples e.g. "enuresis" or "child"	Mentions type of intervention without specifics e.g. "Behavior modification"	Mentions a specific comparison group e.g. "placebo" or "medical treatment" or "no treatment"	Disease oriented outcome without accompanying measure of clinical relevance e.g. "urine output"
Limited (1 pt)	A single general descriptor unlikely to contribute to search e.g. "patient"	Mentions intervention but unlikely to contribute to search e.g. "methods" "options" "treatment"	Mentions comparison but unlikely to contribute to search e.g. "compared to other methods" *Note: Using a plural non-specific term, e.g. "various treatment options" should only be counted once, in the Intervention column.*	Reference to outcome, but so general as to be unlikely to contribute to search e.g., "effective," "improvement," "success" "change the outcome"

Not Evident (0 pts)	None of the above present	None of the above present	None of the above present	None of the above present

2. Where might clinicians go to find an answer to questions like these? Name as many possible types or categories of information sources as you can. You may feel that some are better than others, but discuss as many as you can to demonstrate your awareness of the strengths and weaknesses of common information sources in clinical practice. Describe the most important advantages and disadvantages for each type of information source you list.

	Variety of Sources	Convenience	Clinical Relevance	Validity

Excellent (6 points)	At least four types of sources listed. Types include: • electronic databases of original literature (Medline, Embase, CINAHL) • journals (JAMA, NEJM) • text book (Merck, Harrisons, monographs) • Systematic Reviews (Cochrane) • EBM publications or databases of pre-appraised information (Best Evidence, InfoRetriever, DynaMed, EBM, ACPJC, EBP, Clinical Evidence) • Medical website (MDConsult, PraxisMD, SumSearch) • General internet search (google, yahoo) • Clinical Guidelines (Guideline Clearinghouse, • Professional Organization (AAFP, La Leche League, NIH website) • People (colleague, consultant, attending, librarian)	Discussion includes at least 2 specific issues related to convenience, or mentions the same issue while discussing two different sources. Issues may include: • Cost (e.g. "free," "subscription only") • Speed (e.g. "fast," "takes time") • Ease of search (e.g. "must know how to narrow search," "easy to navigate") • Ease of use (e.g. "concise" and "NNTs already calculated") • Availability (e.g. "readily available online")	Discussion includes at least 2 specific issues related to relevance, or mentions the same issue while discussing two different sources. Issues may include: • Clinically relevant outcomes • Written for clinical application (e.g. "pertinent" "info on adverse effects" or "has patient information sheets") • Appropriate specialty focus (e.g. "directed at FPs") • Information applicable to patient in question (e.g. "can go over details of this particular patient" or "most of studies are from Europe") • Includes specific interventions in question • Specificity (overview vs. targeted) (e.g. "can get basic information" or "more specialized") • Comprehensiveness of source (likelihood of finding an answer in that source) (e.g. "she can find anything" or "contains usable references" or "not likely to have answer to this question")	Discussion includes at least 2 specific issues related to validity, or mentions the same issue while discussing two different sources. Issues may include: • Certainty of validity (e.g. quality is uncertain" or "has not been screened" or "needs to be critically appraised") • Evidence Based approach (e.g. "evidence based" or "Grade 1 Evidence" or "no references provided") • Expert bias (e.g. "usually just someone's opinion") • Systematic approach • Peer review • Ability to verify • Standard of care (e.g. "accepted in medical community") • Enough information provided to critique validity (e.g. "abstract only" or "not available full-text") • Up-to-date/outdated (e.g. "most recent research")
Strong (4 points)	Three types of sources listed.	Includes 1 specific issue/explanation related to convenience	Includes 1 specific issue/explanation related to relevance	Includes 1 specific issue/explanation related to validity

Limited (2 points)	Two types of sources listed.	Mentions convenience involved in using one or more source, but without explanation e.g. "convenient" or "easy" or "difficult"	Mentions relevance of using one or more source, but without explanation e.g. "relevant"	Mentions validity of using one or more source, but without explanation e.g. "good" "junk"
Not Evident	No variety. Only one source listed, or all sources of same type.	No mention of convenience	No mention of relevance	No mention of validity

3. If you were to search Medline for original research on one of these questions, describe what your search strategy would be. Be as specific as you can about which topics and search categories (fields) you would search. Explain your rationale for taking this approach. Describe how you might limit your search if necessary and explain your reasoning.

	Search Terms	Tags	Delimiters
Excellent (8 pts)	3 or more terms that reflect **patient, intervention, comparison, and outcome** (PICO) being considered	Description of search strategy reflects understanding that articles in database are indexed by more than one field. Discusses one or more field/index/tag by name (*MeSH, Title Word, Publication Title, language, Keyword, author, Journal title, etc.)* and provides plausible rationale for search strategy using 1 or more of these indices e.g. "keyword is less specific than MESH"	Describes more than one approach to limiting search (e.g., "*limit to human*" or "*adult*" or "*English*"), names a specific publication type, or describes of Clinical Queries in PubMed, or the use of Boolean operators or search combinations or includes terms related to an optimal study design (e.g. *randomized*) or suggests use of subheadings * NOTE: If the subject includes the name of the index when describing a delimiter (e.g. "check language as English") then we give credit for a tag as

			well as a method of delimiting.
Strong (6 pts)	2 terms from PICO	Names 1 or more field or index category but does not provide plausible defense of search strategy based on this knowledge e.g. "I would do a keyword search…"	Describes only 1 common method of limiting search
Limited (3 pts)	1 term from PICO	NA	NA
Not evident (0 pts)	Not present	No evident understanding that articles "tagged" by different fields or indices	No valid techniques for limiting a search listed

4. Choose to focus on one of the clinical scenarios (breastfeeding and oral contraceptives, or bedwetting alarm). What type of study (study design) would best be able to address this question? Why?

	Study Design	Justification
Excellent (12 pts)	Names one of the best sources: *Randomized Controlled Trial* or *Randomized Trial,* *Systematic Review or Meta-Analysis of RCTS,* *Randomized, Double Blinded Clinical Trial*	Includes well-reasoned justification that reflects understanding of the importance of randomization and/or blinding. Explicitly connects randomization to reduction of confounding and/or blinding to observer or measurement bias. e.g. "An RCT will attempt to avoid any bias which would influence the outcome of the study through randomization" OR "best suited for therapy questions because it reduces bias and controls for confounding factors."
Strong (9 pts)	Describes but does not call by name one of the best sources as above e.g. "comparing two groups, one gets treatment, other	Justification is present, and touches on issues related to randomization and/or blinding, but less clearly articulated e.g. "groups should be similar" or "try to eliminate confounding factors"

	gets placebo…"	or "avoid selection bias" or "to be objective" or "to eliminate bias"
Limited (6 pts)	Describes or names a **less desirable study design**: e.g. "Cohort study" or "Prospective clinical trial" or meta-analysis of such studies, "longitudinal" or "prospective"	Justification is present, and raises **legitimate issues** unrelated to randomization or blinding, such as cost effectiveness, ethical concerns, recall bias. May mention randomization or blinding but without explanation. (e.g. "best in a random and blind setting") e.g. "impossible to recruit women to get a placebo instead of birth control" or "chart reviews provide lots of data without much cost"
Minimal (3 pts)	Describes or names **a poor study design** to answer a treatment question: e.g. case control, cross sectional study, case report, "retrospective" Or describes a study with insufficient detail to identify a design: e.g. a comparative study	**Attempted justification**, but arguments are non-specific and do not demonstrate understanding of the relationship between the design and various threats to validity May mention randomization or blinding but without explanation. (e.g. "best in a random and blind setting") e.g. "to ensure quality" or "to reduce potential conflicts" or "to compare"
Not Evident (0 pts)	None of above present	None of above present

5. When you find a report of original research on these questions, what characteristics of the study will you consider to determine if it is relevant? Include examples. (Questions 6 and 7 will ask how to determine if the study is valid, and how important the findings are....for this question, focus on how to determine if it is really relevant to your practice.)

(Questions 5-7 address critical review of literature divided into relevance, validity, and magnitude of effect size. These may be arbitrary subdivisions of the process of critical review. Therefore respondents may describe issues of relevance in answers to any of these 3 questions. Consider the responses to all 3 questions as one response when applying the criteria in the following rubric.)

	The Question	Description of Subjects

Excellent
 	
(12 points)
	Well-reasoned and thoughtful discussion of the relevance of the independent and dependent variables used in the study including **examples/specific reasons**. May refer to:
	
	• the feasibility of the test or intervention
	
	    e.g. "the test might work but if my practice can't afford to buy the machine it doesn't matter"
	
	• the patient or disease-oriented nature of the outcome
	
	    e.g. "did they measure dry nights after a week or after several months?" or "should measure infant growth, not just amount of milk produced"
	
	• the congruence between the operational definition and the research question *e.g. "whether their method of measuring the outcome is a realistic representation of the outcome we care about"*
		Includes both:
		
		• A clear expression of the importance of the link between the study subjects and target population.
		• At least one example of a relevant disease or demographic characteristic
		
		    e.g. "were the patients similar to mine in terms of age and race?" or "was it a hospital or clinic sample like my patients?" or "did patients have same level of disease severity as my patient?" or "did selection or inappropriate inclusion criteria result in a study population that differs from mine by race, age,etc"

Strong
 	
(9 points)
	Less thoughtful discussion of the relevance of the independent and dependent variables used in the study. May include specific concepts or examples without clear rationale. May refer to:
	
	• the feasibility of the test or intervention
	
	    e.g. "is it feasible?" or "can I actually use it?"
	
	• the patient or disease-oriented nature of the outcome
	
	    e.g. "look for patient-oriented outcomes" or "does the outcome matter to my patient?"
	
	• the congruence between the operational definition and the research question *e.g. "did they measure what they set out to*
		Includes one but not both:
		
		• A clear expression of the importance of the link between the study subjects and target population.
		• At least one example of a relevant disease or demographic characteristic
		
		e.g. "is the patient like mine?" or "education level of population"

	study?" or "what methods were used to determine lactation performance?"	
Limited (5 points)	Response **implies consideration** of how well the study addresses the question at hand, but offers little discussion about why this may be important e.g. "what are the variables?"; "does it answer my question?"; "the outcome measure"; "the purpose of the study"; "will it impact my practice?"; "what type of OCP was used in the study?"; "length of follow-up"	Response implies consideration of the study subjects, but offers no discussion of the connection between study subjects and target population or specific characteristics of the sample e.g. "is it an appropriate sample?" or "what was the response or participation rate?" or "what were the exclusion criteria?" or "selection bias" or "setting" or "where study was conducted"
Not Evident (0 pts)	No discussion of the research question and variables used to answer it.	No discussion of the characteristics of the research subjects.

6. When you find a report of original research on these questions, what characteristics of the study will you consider to determine if its findings are valid? Include examples(You've already addressed relevance, and question 7 will ask how to determine the importance of the findings...for this question, focus on the validity of the study.)

(Questions 5-7 address critical review of literature divided into relevance, validity, and magnitude of effect size. These may be arbitrary subdivisions of the process of critical review. Therefore respondents may describe issues of validity in answers to any of these 3 questions. Consider the responses to all 3 questions as one response when applying the criteria in the following rubric.)

	Internal Validity
Excellent (24 pts)	Lists or describes at least 5 issues important to internal validity, such as: • Appropriateness of study design • Adequacy of blinding • Allocation concealment • Randomization of group assignment • Invalid or biased measurement *("followed own protocol?")* • Importance of comparison or control group

	• Intention to treat analysis • Consideration of appropriate covariates *("were other relevant factors considered?")* • Conclusions consistent with evidence *("do the results make sense?")* • Importance of follow-up of all study participants • Appropriate statistical analysis • Sample size / Power • Sponsorship • When study was conducted • Confirmation with other studies
Strong (18 points)	Identifies 3-4 specific issues as above.
Limited (10 pts)	Identifies 2 specific issues as above.
Minimal (5 points)	Mentions internal validity or lists one specific concept from examples above.
Not Evident (0 pts)	None of the above present

7. When you find a report of original research on these questions, what characteristics of the findings will you consider to determine their magnitude and significance? Include examples. (You've already addressed relevance and validity…for this question, focus on how to determine the size and meaning of an effect reported in the study.)

(Questions 5-7 address critical review of literature divided into relevance, validity, and magnitude of effect size. These may be arbitrary subdivisions of the process of critical review. Therefore respondents may describe issues of magnitude and significance in answers to any of these 3 questions. Consider the responses to all 3 questions as one response when applying the criteria in the following rubric.)

	Magnitude	Statistical Significance
Excellent (12 pts)	Response must clearly discuss both: • clinical significance (*"what is the clinical significance?" or "how large a difference was found?"*) • example(s) of effect size measurements *(e.g., specificity, sensitivity, likelihood ratio of a test, number needed to treat, relative risk, absolute risk reduction, mean difference for continuous outcomes, positive or negative predictive value)*	Well-reasoned and thoughtful discussion of the indices of statistical significance, including at least 2 specific examples of important related concepts such as: • p-values • confidence intervals • power • precision of estimates • Type 1 or Type 2 error
Strong (9 pts)	Response discusses one but not both: • clinical significance (*"what is the clinical significance?" or "how large a difference was found?"*) • example(s) of effect size measurements *(e.g., specificity, sensitivity, likelihood ratio of a test, number needed to treat, relative risk, absolute risk reduction, mean difference for continuous outcomes, positive or negative predictive value)*	Lists more than one concept (as above) with insufficient or absent discussion (e.g. "p-value and confidence intervals") OR Lists and discusses only one concept (e.g. "p-value less than <.05")
Limited (5 pts)	Response only suggests consideration of clinical significance or size of effect. *(e.g. "does it matter?" "will it impact my practice")*	Mentions need to assess statistical significance or names only one concept from above without further discussion (e.g. "p-values")
Not Evident (0 pts)	None of the above present	None of the above present

8. A recent study of the diagnostic accuracy of arterial blood gas in diagnosis of pulmonary embolus included 212 patients with suspected pulmonary embolus, 49 of whom were subsequently determined to have pulmonary embolus. Of those with pulmonary embolus 41 had abnormal alveolar-arterial oxygen gradient ((A-a)DO2). Of the 163 patients determined not to have pulmonary embolus, 118 had abnormal (A-a)DO2.

(4 points each)

- Based on these results, the sensitivity of (A-a)DO2 for pulmonary embolus is .837   OR 41/49
- Based on these results, the specificity of (A-a)DO2 for pulmonary embolus is .276   OR 45/163
- Based on these results, the positive predictive value of (A-a)DO2 for pulmonary embolus is .258 OR 41/159 OR 41/(41+118)
- Based on these results, the negative predictive value of (A-a)DO2 for pulmonary embolus is .849   OR 45/53 OR 45/(8+45)
- Based on these results, the likelihood ratio positive for an abnormal (A-a)DO2 for pulmonary embolus is 1.156 OR .84/(1-.28)

9. A recent randomized trial of found that 29% of diabetics with coronary heart disease (CHD) treated with pravastatin suffered a recurrent coronary event during 5 years of follow-up, while 37% of the placebo group suffered recurrent coronary events.

(4 points each)

- The absolute risk reduction for recurrent events is 8% OR .37-.29
- The relative risk reduction for recurrent events is 22% OR (.37-.29)/.37 OR .08/.37 OR 1-(.29/.37)
- The number needed to treat (NNT) to prevent one recurrent event is 12.5 OR 1/.08 OR 1/(.37-.29)

10. The recent HERS study compared women on estrogen supplements to women on placebo. Results revealed a relative risk of venous thromboembolic events is 2.89 for the women on estrogen. This suggests that estrogen treatment poses a coronary risk, but we wonder if this difference is statistically significant, so we look at the confidence interval. Give an example of a confidence interval that would support the conclusion that the rate of venous thromboembolic events was indeed (statistically) different for these two treatment groups. ___
(anything that encompasses 2.89 and doesn't include 1.0)

(4 points)

11. Which study design is best for a study about diagnosis? cross-sectional study OR "comparison of test with gold standard"

(4 points)

12. Which study design is best for a study about prognosis? cohort studies OR "prospective" OR "longitudinal"

(4 points)