The Critical Appraisal

- Overview of the critical appraisal for primary research study designs
- Critical Appraisal Question Guide & CA Example

Barbara M. Sullivan, Ph.D.
Department of Research, NUHS

Benefits of a Critical Appraisal

- An analytical summary and evaluation of a research study
- Standard approach: recognize important information
- Standard format: easily digested, a quick read
- Usable by professionals in busy practices as summarized, synthesized evidence
- Links practitioner to primary research
- Rapidly accessible, archived for your use in busy practices
Critical Appraisals: EBP in Action

- Several critically appraised primary research papers focused on the same patient oriented clinical question = Critically Appraised Topic (CAT)
- Several summarized Critical Appraisals focused on the same topic = Best Evidence for a Topic (BET)

Really useful places to find accessible, patient focused CATs and BETs

- Critically Appraised Topics
  - University of North Carolina – Chapel Hill Dept. of Internal Medicine
    http://www.med.unc.edu/medicine/edursrc/lcatlist.htm
  - Center for Evidence Based Medicine Oxford University
    http://www.minervation.com/cebм2/cats/alcats.html

- Best Evidence Topics (and linked Critical Appraisals)
  - Emergency Department of Manchester Royal Infirmary, UK
    http://www.bestbets.org/

- Journal Clubs
  - American College of Physicians (ACP) Journal Club http://www.acpjc.org/
    available through NUHS EBSCOhost (search & browse)
The Critical Appraisal

Let's look at a Critical Appraisal…

Back Pain Therapies: Patient Scenario

- Evidence based practice begins and ends with a patient
- Describe the case or problem that focused your clinical question and structured search
- Present a patient focused clinical question (PICO)
**Back Pain Therapies: Patient Scenario**

James, 32 year old male, technical sales manager who drives long distances (or at least for long periods of time) and is a frequent flier for his job, has been seeing you for neck and back issues for over two years on a fairly regular basis.

In addition to working from the car and plane, he has a home office and uses a laptop. You and he have discussed work place ergonomics as well as exercise and stretching to alleviate chronic neck and back pain.

On a recent visit, he tells you he heard on the TV news that chronic back and neck pain causes depression -- or was it vice versa? -- and that chiropractic and alternative care that relieves the pain can relieve depression. He asks, “Do you think that this back and neck thing could be causing me to feel blue lately? Or do you think mid-winter blues are causing this pain in my lower back and neck? My colleague just had lumbar disk surgery and feels great. I’m not so excited about surgery. Should we be doing something different?”

He says after hearing that news report, he’s been looking on the internet to see what might help with the chronic pain and lift his mood.

**Critical Appraisal: Back Pain Therapies**

**PICO Question**

<table>
<thead>
<tr>
<th>Patient, population, problem</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consider</strong> 32 yo WM chronic neck / back pain spine, lumbar disk, Stenosis, sciatica, etc.</td>
<td>CAM therapy chiropractic acupuncture “physical therapy” nonsurgical nonoperative botanical herbal massage</td>
<td>Prescription drugs Opioid compounds surgery massage… alternative therapy treatment</td>
<td>to treat / relieve chronic (neck / back) pain (mild) depression spine conditions</td>
</tr>
</tbody>
</table>

For **P=** adult patients with chronic (neck / back / spine / lumbar) pain / specific diagnosis], is **I=** conservative / nonoperative treatment / botanical therapy / acupuncture] as effective as **C=** surgery] to **O=** alleviate pain / treat symptoms of mild depression?”
Include the search strategy, results & evaluate:

- Searching, finding, accessing is essential to the evidence-based practitioner.
- Communication skills are essential to applying and assessing evidence.
- List separate searches, queries.
- Explain what you did, summarize.
- Bullet point how full text was located.

Starting an effective search

Go to the source your patient recalls ...

- www.google.com
- NBC 5 health
  - Chiropractic low back pain
Read, follow …

According to an overview of back pain from the National Institute of Health, the back is uniquely complicated as it is made up of dozens of bones, muscles and nerves, all of which must work in harmony. A slight strain in one muscle or a slight misalignment between two bones can cause a new or existing back pain.

Chiropractors such as Lederer take X-rays to ensure that patients’ spines are in need of realignment.

“While we see most often in the X-rays we take are degenerative arthritis, degenerative disc disease or other conditions,” he said.

Treating The Pain

Surgery may be necessary for more serious cases of back pain, but Lederer emphasized that it is only an option after more conservative options have been exhausted. He cited a recent article in the Journal of the American Medical Association that showed back surgery and certain exercise programs yielded about the same results.

Manipulation from a chiropractor is a much less drastic method for helping reduce the imbalance that causes the pain. Chiropractors go through years of training to learn how to adjust joints enough to restore a normal range of motion.

Spinal manipulation — often accompanied by the “pop” or click of air being moved from a joint — is not a full solution by itself. According to Lederer, “chiropractic care now and exercise later” is the best approach. He gives patients exercises to perform that will strengthen back muscles, and he also emphasizes the need for periodic adjustments.

Long-Term Solutions

Of course, if your job is contributing to your back problems, it may be necessary to change the conditions in which

Follow a link to Website #3:
Follow a link to a research article:
http://jama.ama-assn.org/cgi/content/abstract/296/20/2451

Search strategy and results:

Search Engines / Program(s) & Databases searched
1) Google, NBC5.com
2) Natural Standards (www.naturalstandards.com)
3) Entrez PubMed
4) EBSCOhost: CINAHL, AMED

Query used (Key Search Terms, Operators used and limits)
2) Conditions: lumbar, low back, cervical and neck key word search
3 & 4) (back OR lumbar OR neck) pain surgical > (Limits: human, date: 2005-2007, peer reviewed)
4) PubMed Clinical Query: therapy, narrow, specific

Limits and Special Techniques:
- Patient info; local TV channel website Google search back pain
- MeSH for “surgery” led to nonoperative;
- Boolean operators: included OR for multiple conditions
- Limits used to revise search: published in the last 2 years, Humans, English, core clinical journals, complementary medicine, adult:19-44 yrs
Search strategy and results:

**Search results:**
- Google search TV website: links to 1 website, 1 article
- PubMed with limits: 69 articles, 6 reviews
- EBSCOhost 149 articles, 14 reviews

**Selection rationale:** (JTASS)
- Journal of the American Medical Association (JAMA) is peer reviewed, strong publishing history (1060), professional association AMA;
- Title key words focus on surgery versus non-operative for lumbar disc herniation;
- Authors have a publishing track record in spine research, surgery
- Large subject population from surgical centers is generalizable;
- Outcomes measured: pain (patient), physical disability; secondary sciatica, return to work, quality of life;
- Strong study design (randomized clinical trial)

**How full text was accessed:**
- Website link to JAMA related article, available as free full text; PubMed
- Related articles in NEJM, JAMA available through NUHS EBSCOhost and LRC password list

---

**Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial.**

Weinstein JN, Tosteson TD, Lurie JD, Tosteson AN, Hanscom B, Skinner JS, Abdu WA, Hilibrand AS, Boden SD, Deyo RA.


**Type of study:** Therapy

**Study design:** Randomized clinical trial

Study objectives and hypothesis

- State the purpose, objectives and hypothesis
- Using your words, what was the research question and objective(s) of the study?
- Was the purpose of the study conveyed clearly and rationally?
- Were the objectives of the study clearly stated?
- Was the hypothesis / null hypothesis explained (RCTs)

Type of study, study design, strength

- What is the stated study design?
- Was the study design stated and adequately described?
- Considering the strengths and limitations of the study design, is it suitable for the objectives?
“Best” study design

- **Therapy:**
  - randomized controlled trial (RCT), randomized clinical trial (comparison, no zero control, placebo), strong cohort with defined control
  - Other study designs are valid, not as “strong”

- **Prognosis:**
  - cohort studies with untreated / exposed control, case control design, strong, well-defined case series

- **Diagnosis:**
  - cohort study with strong reference standard, strong all-or-none case series

- **Etiology / Harm:**
  - RCT, prospective cohort, case control with well defined control / comparison
Methods: Subjects / Participants / Patient / Population

- Focus on PICO components, but don’t leave out info that might affect validity
- “Real life” circumstances of study? (relevance)
- Population: large group of people (should be described)
- Sample: population subset selected from a larger population
- Selection population. Bias?
- Method of selection...
  - Why selected?
  - How were patients recruited? Bias?

Selection bias:
- Differences in intervention and comparison groups due to incomplete randomization
  specific allocation
decision of when a “case” is a “case”

Inclusion / exclusion criteria:
- Why subjects are enrolled in a study or left out
- Broad? Narrow? Generalizable?

Confounders:
- Characteristics or factors not under study or not included in criteria
- May affect the outcome of the intervention or disease
Methods: Subjects / Participants / Patient / Population

- Baseline differences?
  - Experimental and control groups start with similar prognosis
  - More homogeneity is stronger
- Did the population, experimental and control or comparison groups start with the same baseline demographics and prognostic factors?
- How homogeneous is the population selected?

Confounders: 2 or more factors that are “associated” (age and weight) and may affect (confuse, distort, augment?) the effect of the other factors on the outcome (onset of diabetes)
Methods: Subjects / Participants / Patient / Population

- Sample size adequate to support measurement of outcomes?
  - Rare event?
  - Likelihood of staying in the study, following through?
  - Size based on previous studies, outcomes?
  - Rationale for choosing sample size?

Methods: Subjects / Participants / Patient / Population

- Sample size adequate to analyze statistically?
  - Rule of thumb:
    - 30 subjects per group
  - Case series: at least 10 subjects with well defined characteristics, baseline, histories
  - Likelihood of staying in the study, following through?
  - Rationale for choosing sample size?
Follow-up / Accountability

- Were all study participants or subjects accounted for at the end of the study?
- Rule-of thumb: >20% drop-out, non-adherence affects validity
- Unintended cross-over
- Cross-over not accounted for affects validity
- Are the reasons why patients withdraw from clinical trials included in the follow-up information?
- Exclusion bias: systematic differences in withdrawals from a study
  - between groups
  - of certain subsets

Follow-up / Accountability

Intent to treat analysis

- Include / analyze all patients in the group to which they were assigned
- regardless of whether or not they finished the study
- regardless of compliance
- Accounts for drop-out, not necessarily crossover
- High rate of crossover dilutes power of intervention
Ethical Approval

- Institutional Review Board (IRB)
- Informed consent
- Disclosure of methods, intervention, risks, predicted benefits
- Different from affiliation and support

Methods: Intervention

- Describe intervention
  - Relate to PICO question
  - Described sufficiently so that the reader (practitioner) can adequately deliver the same intervention?
  - Adequate length in experimentation / observation / trial and measurement?
  - Adequate number of visits provided at appropriate intervals and frequency?
- With what was the investigated or experimental intervention compared?
  - Gold standard, alternate, placebo, sham?
  - Was the comparison valid? Realistic?
  - Why was the comparison selected?
Methods: Intervention

➢ Performance Bias
  • Difference in care provided to intervention and comparison groups case versus non-case or control
  • Other than difference in intervention / comparison
  • Population presenting at hospital versus control presenting at private practice
  • Systematic

Methods: Randomization

➢ Incomplete randomization and crossover threatens validity
➢ By computer upon enrollment
➢ By lottery / blind drawing
➢ No involvement of study investigator, recruiter, enroller, etc.
➢ Allocation to a particular group: cohort study
  ➢ Patient choice, provider’s choice, expertise
**Methods: Blinding**

- Multiple points where investigators and subjects don’t know...
- Provider determining eligibility does not know to what group a subject is randomized
- Subjects are blind to whether they receive treatment or comparison (placebo, sham, conventional treatment)
- Person providing treatment, dispensing does not know what they are providing (experimental, real, sham)
- Assessors are blind to randomized group, treatment, exposure, whether “case” or “control”

**Methods: Outcomes, measurement, observation**

**Outcome:**
- “Outcomes” and results are different
- Outcome: what is accomplished; what is measured
  - Clinical event or accomplishment of interest, desired effect, end product or consequence following an intervention or exposure
- Clinically relevant
  - “a reduction in blood pressure,”
  - “reduced mortality,”
  - “better quality of life,”
  - “management of blood glucose levels,”
  - “resolution of pain,” etc
- Result: measurement of an outcome is reported as a result
Methods: Outcomes, measurement, observation

Outcome:

- Biologic outcomes or surrogate endpoints
  - decrease in blood glucose levels, decrease in serum IgE levels, half-life of a drug in serum samples
- may not singularly correlate with a clinical outcome
  - control of diabetes, death, recovery from a disease, decrease in blood pressure
- “Flaw” to make a “claim” regarding a clinical outcome when a biological outcome or surrogate endpoint is assessed
- Watch for validation of quantitative biological endpoint with accepted, subjective, qualitative, clinical outcome endpoint or measurement tool

Outcomes Measured

- Primary outcomes: changes from baseline bodily pain and physical function:
  - the Medical Outcomes Study (a 36-item Short-Form Health Survey scales
  - the modified Oswestry Disability Index (American Academy of Orthopaedic Surgeons MODEMS version)
  - Measured at 6 weeks, 3 months, 6 months, 1 and 2 years from enrollment.
- Secondary outcomes:
  - sciatica severity (Sciatica Othersomeness Index)
  - satisfaction of self-reported improvement of symptoms
  - employment status and quality of life function assessment
Results: calculation of measurement

➢ Primary outcomes: changes from baseline bodily pain and physical function

➢ Results:
  - Pain decreased in intervention group 72%
  - Disability index score decreased from 12 to 4
  - Pain measured by Visual Analogue Scale decreased in sham by 35%

Statistical Analysis
“p value”

- Probability that any particular outcome would have arisen by chance.
- “Standard scientific practice” (often considered somewhat arbitrary):
  - $p < 0.05$ (p value less than one in twenty) is "statistically significant"
  - $p < 0.01$ (p value less than one in one hundred) is "statistically highly significant"
  - $p$ values $> 0.05$ (e.g., 0.49, or 0.30) are not considered statistically significant

“p value”

- Statistically significant: reject the “null hypothesis”
  - $p$ values in the non-significant range indicate that either there is not difference between groups OR there are too few subjects to demonstrate a difference (if a difference exists).
  - Does not determine which circumstance the $p$ value reflects.
  - Typically, "positive trials" show a statistically significant difference between groups or arms of a trial, and "negative trials" appear to show no significant difference between groups or arms.
- Look for statement of cutoff chosen for the study (e.g., $p < 0.05$ or $p < 0.01$) and why
A typical clinically relevant confidence interval of 95%.

The wider the confidence interval, the more likely that a certain result will fall within that interval. Strong evidence will have a wider confidence interval.
**Statistical Analysis**

- Sample size allowed up to 20% missing data
- Analyses for primary and secondary outcomes used all available data
- Predetermined outcomes
- Predetermined endpoint measurement times.
- Adjustments, analysis made for missing data
- P<0.05 used to determine statistical significance
- Confidence intervals (CI) of 95% for mean treatment effects at each designated time
- Global tests of joint hypothesis of no treatment effect at any designated time performed

---

**Validity & Limitations**

- Representative population?
- Bias in selection, prognostic factors, confounding factors
- Follow-up: drop-outs threaten validity
- Ignoring withdraws typically favors intervention
- Comparison should be equivalent
- Non-adherence (cross-over) threatens validity when >20%
- "Intention-to-treat" analysis adjusts for drop-outs, not cross-over
- Dilution of effect of intervention (surgery)
- "As-treated" analysis may compensate for cross-over, but may exaggerate effect of intervention if unmeasured or differing baseline factors favor intervention
Clinical Impact & Significance

- Do the studies add anything to the body of evidence?
- What is your evaluation of the strength of the evidence presented in these selected papers?
- Does your appraisal of the papers indicate studies are as strong as / stronger than the “CEBM” designations indicate?
- Is the evidence presented strong, moderately strong, neutral or weak if therapy, prognosis or etiology papers were selected?
- Does the evidence support the therapy, diagnosis, procedure or diagnostic tool discussed?
- What is the clinical significance in light of your patient?
- Form a “Clinical Impact Statement” referring to your patient
Clinical Impact & Significance

Impact statement:

Using this study and related articles from the SPORT trial, patients with LDH, bodily pain and disability may try conservative care unless the pain and disability are too much to bear. Further research on “reverse hypothesis” looking at specific conservative therapies compared to “standard,” efficacious surgical intervention should be done.