Tufts Clinical and Translational Science Institute

Lori Lyn Price, MAS
Biostatistics, Epidemiology, and Research Design (BERD)
Tufts Clinical and Translational Science Institute (CTSI)

Tufts CTSI’s Mission & Purpose

- Tufts CTSI is based on the conviction that authentic involvement of the entire spectrum of clinical and translational research is critical to meeting the promise and the public’s needs of biomedical science.
- Our mission is to identify, stimulate, and expedite innovative clinical and translational research, with the goal of improving the public’s health.

39 Tufts CTSI Partners

http://ilearn.tuftsctsi.org/

Live seminars are recorded for our I LEARN site. Seminar videos can be viewed at any time, and are free!

How to Request Tufts CTSI Services

- Visit www.tuftsctsi.org and submit a request

Evaluating Scientific Journal Articles

Lori Lyn Price, MAS
Biostatistics, Epidemiology, and Research Design (BERD)
Tufts Clinical and Translational Science Institute (CTSI)
Learning Objectives

• List the questions you should ask yourself when evaluating a scientific journal article
• Identify the specific, testable hypothesis of the paper
• Identify what type of study design was used
• Evaluate whether the results of the study were affected by bias
• Explain why this study was important, what it added to the literature, or how it changed health practice
• Appraise the compatibility of the conclusions of the study with the study objectives

Evaluation of a Scientific Article


Key Sections of a Journal Article

1. Abstract
2. Introduction/Background
3. Methods
4. Results
5. Discussion
6. Conclusions
7. References

Introduction to Evaluating Articles

• In most articles, the authors tell a story based on data that they have collected, analyzed, and interpreted
• The reader should evaluate each of these phases to decide whether to trust the story
• It’s also important to understand why the study was done

Overall Issues in Evaluation

Big picture:
• Strength of the body of literature
• Plausibility of biological/health mechanism
• Effect size and number of people
• Quality of study

Quality methodological details:
• Appropriate hypothesis
• Study design
• Data quality
• Plausible effect estimate or concern about biases
• Generalizability

Where Studies “Fail”

• Inadequate study design
• Biased sample and results
• Uncontrolled confounding
• Study sample too small
### Where Studies “Fail”
- Lack of generalizability
  - Single center
  - Subject recruitment & retention
  - Inclusion/exclusion criteria
- Misinterpretation of results
  - Conclusions don’t match results

### Following the Story Part 1
**Context / Motivation**

What was the motivation for doing this study?

Did the authors conduct this study to:
- Generate descriptive or pilot data or new hypotheses?
- Test a formulated hypothesis?
- Replicate or validate previous findings?

### Motivation
- Evaluate 2 specific hypotheses:
  - Adults with chronic lower back pain (CLBP) treated with Mindfulness-Based Stress Reduction (MBSR) would show greater short- and long-term improvement than adults randomized to usual care
  - Adults with CLBP treated with MBSR would show greater short- and long-term improvement than adults randomized to Cognitive Behavioral Therapy (CBT)

### Why Was the Study Important?
- What information already exists about this topic?
  - Functional status of people with CLBP has decreased over time, despite numerous treatment options and resources
  - Psychosocial factors are a component of pain
  - CBT is known to be effective for a variety of types of chronic pain, but limited access
  - MBSR, another mind-body component, is becoming increasingly available
  - MBSR is “helpful” for chronic pain

- What were the gaps in the literature that this study sought to help fill?
  - Is MBSR effective in treating CLBP?
  - Is MBSR more effective than CBT?

- What other factors make this an important study?

### Does the Paper Present a Clear Research Question or Objective and a Specific, Testable Hypothesis?

**Study objective**

“To evaluate the effectiveness for chronic low back pain of MBSR vs cognitive behavioral therapy or usual care.”

**Testable hypothesis?**
Examples of Hypotheses
1) MBSR is more effective than CBT in treating CLBP.
2) MBSR is more effective than CBT in reducing back pain.
3) MBSR is more effective than CBT in reducing back pain over (pre-specified time frame) using (pre-specified instrument)

Following the Story Part 2
Design & Data Collection

How was the study conducted?
• Inclusion and exclusion criteria
• Recruitment of participants
• Study design
• Definition of outcomes
• Administration of intervention

Inclusion and Exclusion
• 20-70 year olds
• Non-specific CLBP for at least 3 months
• No compensation or litigation issues
• English speaking
• Able to attend classes
• Adequate pain, based on bothersomeness and pain interference questionnaires

Recruitment
• Group Health members
  – Eligibility based on medical record
  – Invited to participate via mail
• Community
  – Random sample of participants
  – Invited to participate via mail

Who Were the Study Subjects?
• What was the source population from which study subjects were recruited?
• Was the subject selection process clearly explained?
• How representative was the sample?
**Study Designs**
- Randomized Controlled Trial (RCT)
- Observational Studies:
  - Cohort – selection based on exposure (smoking status)
  - Case-Control – selection based on disease/outcome (lung disease)
  - Cross-sectional – one snapshot in time
  - Retrospective – exposure collected after disease
  - Prospective – exposure collected before disease

**Methods**

**Design & Data Collection**

**What type of study design was used?**
- Randomized Controlled Trial

**Is this design susceptible to any types of bias?**

**Were potential sources of bias identified and addressed when designing the study?**

**Variables**
- Were independent and dependent variables clearly defined and accurately measured?
  - Potential for misclassification
  - Validation of exposure/outcome status
  - Properties of measurement methods

**Definition of Outcomes**

**Functional limitation related to CLBP**
- Roland Disability Questionnaire (validated)
- One item removed
- Asked about past week rather than only today
- Back pain bothersomeness (0-10 scale)

**Intervention: CBT**
- 2 hour weekly group session for 8 weeks
  - Chronic pain education
  - Changing dysfunctional thoughts
  - Workbooks & CDs
  - Instructions for home practice: Relaxation and imagery
Intervention: MBSR

- 2 hour weekly group session for 8 weeks
  - Didactic content
  - Mindfulness practice
- Workbooks & CDs
- Optional 6 hour retreat
- Instructions for home practice: Mindfulness, meditation and yoga

Following the Story Part 2
Design & Data Collection

Subjects

- Were a significant number of subjects lost to follow-up?
  - Differential between groups (26 weeks)
    - 5% usual care, 18% MBSR, 19% CBT
    - 13 participants each in MBSR & CBT did not attend any classes
  - How missing data were handled
    - imputation
- Intent to treat Analysis

Following the Story Part 3
Statistical Methods

Was a Sample Size/Power Calculation Performed Before Beginning the Study?

- Are all calculation parameters reported so that the calculation could be duplicated?
- Outcome: proportion of participants experiencing meaningful improvement
  - 90% power to detect 25% difference in MBSR (55%) vs. usual care (30%)
  - 80% power to detect 21% difference in MSBR (76%) vs. CBT (55%)

Was the sample adequately powered?
Following the Story Part 4
Reporting and Interpretation of Data

Did the authors present and compare the characteristics of the 3 study groups?
- This information is provided in Table 1
- Are there any clinically meaningful differences in Table 1?
  - More women in usual care (77% vs ~60%)
  - Fewer college grads in MBSR (52% vs ~61%)
- Should these affect our analysis?

How Strong Were the Study’s Results?
- Did the investigators find any statistically significant results?
  - Roland Disability: at least 2 of the 3 groups differed significantly (p=0.04)
  - Pain Bothersome: at least 2 of the 3 groups differed significantly (p=0.01)
  - At 26 weeks, adjusted for age, sex, education, baseline score and pain duration
- How likely is it that results were due to chance or bias?

Primary Outcome (%)

| Follow-up Week | Usual Care | Mindfulness-Based Stress Reduction | Cognitive Behavioral Therapy | P-value for Overall
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Roland Disability Questionnaire Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>27.3 (20.3-36.6)</td>
<td>34.5 (26.8-44.3)</td>
<td>24.7 (18.1-33.8)</td>
<td>.23</td>
</tr>
<tr>
<td>8</td>
<td>35.4 (27.6-45.2)</td>
<td>47.4 (38.9-57.6)</td>
<td>51.9 (41.6-61.7)</td>
<td>.04*</td>
</tr>
<tr>
<td>26</td>
<td>44.1 (35.9-54.2)</td>
<td>50.5 (42.0-60.3)</td>
<td>57.7 (49.2-67.6)</td>
<td>.04*</td>
</tr>
<tr>
<td>Pain Bothersome Results</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>20.6 (14.6-26.9)</td>
<td>19.1 (13.1-27.4)</td>
<td>21.7 (15.3-30.6)</td>
<td>.88</td>
</tr>
<tr>
<td>8</td>
<td>24.7 (18.1-33.6)</td>
<td>36.1 (28.3-46.9)</td>
<td>33.8 (26.5-43.2)</td>
<td>.15</td>
</tr>
<tr>
<td>26</td>
<td>26.6 (19.8-35.9)</td>
<td>43.6 (35.6-53.3)</td>
<td>44.9 (36.7-55.1)</td>
<td>.01†</td>
</tr>
<tr>
<td>52</td>
<td>31.0 (23.8-40.3)</td>
<td>48.3 (40.3-58.3)</td>
<td>39.6 (31.7-49.5)</td>
<td>.02†</td>
</tr>
</tbody>
</table>

Chance & Bias
- More than 40 p-values presented in the tables
- Additional tests performed when one of these p-values<0.05
- Are results due to chance?
  - Only 50-60% of participants randomized to MBSR & CBT completed at least 6 classes
- Are results due to bias?

Limitations and Generalizability
Do the authors adequately address the study’s limitations and their implications?
- Highly educated & enrolled in a single health care system
- ~20% loss to followup

Who do the results of this study apply to?
“the generalizability of findings to other settings and populations is unknown”

Are the Conclusions Reasonable Based on the Study’s Aims and Results?
“Among adults with chronic low back pain, treatment with MBSR or CBT, compared with usual care, resulted in greater improvement in back pain and functional limitations at 26 weeks, with no significant differences in outcomes between MBSR and CBT.”

Is this conclusion compatible with the original study objective?
Do the results of the study justify the conclusions?
Are the Conclusions Reasonable Based on the Study’s Aims and Results?

“These findings suggest that MBSR may be an effective treatment option for patients with chronic low back pain.”

What do you think about this?

Superiority vs non-inferiority

How are these hypotheses different?

What is each testing?

MBSR is more effective than CBT in reducing back pain.

MBSR is non-inferior to CBT in reducing back pain

Learning from What’s Not There…

• If an article doesn’t mention a particular issue (e.g. blinding, randomization), it’s usually safe to assume that the study did not address that issue
• All studies have limitations. If none are mentioned, it probably means that issues with the study were not adequately addressed.

Overall Issues in Evaluation

Big picture:
• Quality of study
  • Appropriate hypothesis
  • Study design
  • Data quality
  • Plausible effect estimate or concern about biases
  • Generalizability

• Effect size and number of people
• Plausibility of biological/health mechanism
• Strength of the body of literature (on specific and larger related research questions)

Questions to ask yourself…

➢ What is the big picture?
  • Why is study important?
  • Plausible effect?
➢ Is story believable?
  • Any concerns about quality?
➢ All the important details:
  • Appropriate & specific hypothesis?
  • Design, subject selection, choice of variables, data quality, appropriate analysis, biases, etc.

Questions?
Thank you!