****

Modul 8

FEB 326-Evidence-Based Practice Fisioterapi

Materi 10

Critical Appraisal Terkait Systematic Review

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2018

1. Pendahuluan

Penegakan diagnosis adalah komponen penting dalam aplikasi*evidence-based practice* fisioterapi. Untuk dapat menghasilkan diagnosis yang akurat, diperlukan pemahaman terkait jenis studi dan melakukan kritisi pada studi tersebut.

1. Kompetensi Dasar

Kompetensi dasar yang harus dimiliki oleh mahasiswa dan mahasiswi pada materi ini adalah pemahaman tentang proses kritisi studi terkait penegakan diagnosis

1. Kemampuan Akhir Yang Diharapkan

Setelah mempelajari materi ini, diharapkan semua mahasiswa dan mahasiswi mampu mengkritisi studi terkait penentuan diagnosis dan memahami pentingnya hal tersbut dalam aspek praktik klinis

1. Uraian Materi

What is a Systematic Review?

Systematic reviews are a special type of research study characterized as secondary research. Secondary research studies are “studies of studies” that summarize information from multiple primary studies. Primary studies are “original studies,” such as the randomized clinical trials and cohort studies described in previous chapters. Systematic reviews are the principal scientific method for secondary research. Systematic reviews are developed using a documented, systematic approach that minimizes bias.1 Authors of a systematic review define a specific purpose for the study. Methods that minimize bias are determined prior to the beginning of the study.

Unlike a randomized clinical trial, a systematic review does not include recruiting and enrolling participants. Specific inclusion and exclusion criteria are used to select appropriate studies for review. The sample size for a systematic review is the number of studies identified that meet the specific criteria. Systematic reviews of treatment interventions are most common; however, reviews can also appraise diagnostic tests, outcome measures, and prognostic factors. For example, Dessaur and Magarey2 conducted a systematic review of diagnostic studies to determine the diagnostic accuracy of clinical tests for superior labral anterior-posterior lesions in the shoulder. In another example, Blum and Korner-Bitensky3 conducted a systematic review of studies of the Berg Balance Scale to better understand its usefulness in stroke rehabilitation. In this chapter, we focus on systematic reviews of intervention studies; however, the same principles and appraisal processes can be applied to other types of systematic reviews (Fig. 7.1).



A meta-analysis is a statistical method used to summarize outcomes across multiple primary studies, usually as a part of a systematic review. The sample size of a meta-analysis is considered the total number of participants from all studies combined. Not all systematic reviews contain a meta-analysis. This analysis depends on the nature of the data included in the selected primary studies. In a meta-analysis, data are “normalized” across studies and the results are expressed as a quantitative value representing all included studies (e.g., effect size; see Chapter 4).

Narrative reviews provide an overview of literature and are commonly published in peer-reviewed research journals. These reports are sometimes confused with systematic reviews. A narrative review is not a systematic study and analysis of the literature.4,5 Rather, it is a description of the content of articles selected by an expert with the review expressing the expert’s perspective. As such, narrative reviews are subject to the bias associated with personal opinion. Narrative reviews are a useful resource of expert opinion, but they are not research. Hence, narrative reviews are represented at the lowest level of the research evidence pyramid, whereas systematic reviews are at the top of the pyramid. Clinical expertise is, however, one of the three pillars of evidence based practice (EBP). A narrative review is useful as an expression of clinical expertise. Table 7.1 includes a comparison of the methods and results typically included in systematic and narrative reviews.



Why Are Systematic Reviews at the Top of the Evidence Pyramid?

Systematic reviews are at the top of the Evidence Pyramid because they overcome many of the limitations of individual studies. A single research study rarely provides a definitive answer to a clinical or research question because it represents only one sample from the population. This is particularly true in rehabilitation literature because sample sizes are typically small. A high-quality systematic review combines all of the high-quality studies published on a given topic into one study. Systematic reviews provide a more comprehensive analysis and summary of the research available on a given topic than can be obtained from primary research studies. Systematic reviews also save time for the evidence based therapist. The number of articles published in peer-reviewed journals increases every year (Fig 7.2).



It is unrealistic to |expect to read and appraise all of the literature on the broad range of clinical questions encountered in your practice. Systematic reviews provide important literature for addressing this challenge. In a systematic review, Choi et al9 identified over 2000 potentially relevant articles that addressed the clinical question in Self-Test 7.1. Through a documented and systematic process, those 2000 articles were narrowed to 13 articles, all of which had high applicability to the specific question and rigorous research quality. The authors applied meta-analytic statistics to four of the studies to quantitatively express the effectiveness of post-treatment exercise programs. By combining four separate primary studies with a total of 407 participants we learn that injury recurrence decreased by 50% following a post-treatment exercise program for patients with low back pain. If we determine the review to be of high quality, it provides an efficient resource to guide our care for Mr. Green.

Appraising Systematic Reviews

How do You Determine if a Systematic Review is of Sufficient Quality to Inform Clinical Decisions?

The next two sections of this chapter detail the appraisal process for systematic reviews. Just as with intervention, diagnostic, or any other type of research study, the quality of the study must be appraised before it can be appropriately applied in the clinic. The sections that follow are organized into parts A through D to consider when appraising quality of a systematic review. The 12 questions within these parts form the checklist in Table 7.6.





Part A: Determining Applicability of a Systematic Review

Determining the applicability of a systematic review to your clinical question is similar to the process used for other study types. The questions about applicability presented in Chapter 3 have been modified for appraising systematic reviews.

QUESTION 1: Is the study’s purpose relevant to my clinical question?

As with other study types that we have reviewed, systematic reviews should provide a specific purpose, usually stated at the end of the Introduction. Because a systematic review is intended to appraise a body of literature, the purpose is likely to be broader than your clinical question. You may find that several clinical questions are addressed within one systematic review.

QUESTION 2: Are the inclusion and exclusion criteria clearly defined, and are studies that would answer my clinical question likely to be included?

Systematic reviews should provide specific inclusion and exclusion criteria for studies (not patients) to be considered for the review. It is important to appraise the appropriateness of the criteria used to select studies in the context of your question. Review the inclusion and exclusion criteria to determine if studies that address your patient characteristics and intervention of interest are included in the review. Also consider the study types included in the review. For example, a systematic review of constraint-induced movement therapy (CIMT) for children with hemiplegia due to cerebral palsy11 included only the three randomized clinical trial (RCT) studies on this topic. Recommendations were weakly supportive of CIMT for children and gave recommendation for future studies. A systematic review by Huang et al12 included all studies on this topic regardless of design type. This review gave stronger recommendations for CIMT and also gave recommendations for clinical application. You would need to appraise each review to determine which is more appropriate to guide your clinical practice regarding CIMT.

QUESTION 3: Are the types of interventions investigated relevant to my clinical question?

Because primary research studies investigate a broad variety of interventions, authors of a systematic review should clearly identify the specific interventions included in the systematic review. Assess if the included interventions are relevant to your clinical question and clinical practice. A systematic review could assist you in developing a treatment strategy that you have not used before, or it could provide evidence to support interventions that are already a part of your clinical practice.

QUESTION 4: Are the outcome measures relevant to my clinical question?

Consider if the systematic review addresses your outcome of interest (e.g., return to work, walking endurance, shoulder range of motion). The methods section should provide information about the outcome measures included in the review. Ultimately, however, the outcome measures assessed are dependent on the measures provided by the primary studies included in the review. A common challenge for authors conducting a systematic review is to draw conclusions from the diversity of outcome measures often used across primary studies.

QUESTION 5: Is the study population sufficiently similar to my patient to justify the expectation that my patient would respond similarly to the population?

The authors should provide a summary of the demographics of the participants in the primary studies included in the systematic review. Consider differences and similarities between the sample population and your patient. The study sample might be close enough, even if it does not exactly match your patient. Often the population will be broad (e.g., large age range, multiple diagnoses, different severities of disease). Use your clinical expertise to determine the relation of the information to your patient.

Part B: Determining Quality of a Systematic Review

The clinical usefulness of a systematic review is dependent on its applicability to the clinical case and to the rigor (quality) with which the study is conducted. To determine the quality of a systematic review, we appraise many of the same general concepts presented in Chapters 3 and 4. However, there are aspects of quality that are unique to systematic reviews.1,13 The gold standard method for conducting systematic reviews has been developed and maintained by the Cochrane Collaboration. Digging Deeper 8.1 describes this influential group.

QUESTION 6: Was the literature search comprehensive?

Once a specific purpose has been defined for a systematic review, the next step is for the authors to conduct a comprehensive literature search to identify all articles that may be associated with the topic. You should consider several aspects of the search to determine if it was comprehensive.

* Did the authors use appropriate search terms to search for articles about the topic of interest?

Ideally, a systematic review should provide a detailed list of the terms and search strategies used to conduct the literature search. As you gain skills in searching you will be able to appraise the effectiveness of the authors’ strategy. Table 7.3 illustrates the search strategy reported in a systematic review provided by Alexander et al4 (from Table 7.1).



Was a comprehensive number of appropriate databases used?

The objective of the search methods described in Chapter 2 was to find the best available research to answer a clinical question. Authors of a systematic review attempt to find all of the available literature on a topic. Given this objective, a systematic review should describe numerous (usually three to seven) databases used to conduct the literature search. Chapter 2 emphasizes freely available databases (e.g., MEDLINE via PubMed; TRIP database; and PEDro). Table 7.4 illustrates a list (not exhaustive) of databases commonly used in systematic reviews of physical therapy literature.



Were articles from a wide range of languages included in the literature review?

Although the English language is predominant in physical therapy literature, it not the only language used for publication. Including articles in multiple languages may require the resources of a translator; as a consequence, language inclusion is commonly limited in systematic reviews. Language bias results when important study results are excluded from a systematic review because of language.

* Were efforts made to identify unpublished data relevant to the topic?

Authors of systematic reviews should make an effort to identify studies about the topic of interest that were conducted but have not been published. Publication bias is the tendency for studies with positive results to be published more often than studies with nonsignificant results. Clinical trial registries list clinical trials at their initiation and provide a method to monitor studies that are conducted but the results not yet published. Unpublished data are usually obtained by personally contacting the researcher who conducted the study.

Primary Study Appraisal

QUESTION 7: Was an objective, reproducible, and reliable method used to judge the quality of the studies in the systematic review?

Once the authors have identified studies that meet the preestablished inclusion and exclusion criteria, each primary research article included in the systematic review should be systematically appraised for quality. Ideally, two or more independent reviewers use a standardized study appraisal tool. Each reviewer conducts the appraisal independently and without knowledge of the other reviewer’s appraisal. After the appraisals are complete, discrepancies are addressed by discussion between the reviewers and, if needed, a third reviewer is consulted to resolve the discrepancy. Ideally, the authors should use a study appraisal tool with established reliability and validity in addition to reporting their own raters’ reliability.

There are several standardized appraisal tools used for appraisal. These tools assign a corresponding score for quality. The most commonly used tool in the physical therapy literature is the Physiotherapy Evidence Database (PEDro) scale.15 The PEDro scale ranges from 0 to 10, with 10 representing the highest score for quality. Figure 7.3 illustrates the PEDro scale and the use of the scale to appraise the quality of primary studies included in a systematic review.





Data Reduction and Analysis

QUESTION 8: Was a standardized method used to extract data from studies included in the systematic review?

Extraction of data from each primary study must be done carefully to prevent errors and omissions. For some studies, the authors of the primary studies must be contacted to obtain details beyond those in the published paper. Systematic reviews should provide an explicit description of data extraction. For example, the review by Alexander et al4 provides specific details of the extraction procedures used for the ultrasound parameters from eight primary studies.

QUESTION 9: Was clinical heterogeneity assessed to determine whether a meta-analysis was justified?

One of the most challenging elements of a systematic review, particularly in physical therapy literature, is determining if the data from multiple studies can be combined in a metaanalysis. 16 A meta-analysis is useful because it allows data from several studies to be combined. However, to produce valid results, combined studies must be similar in study population, intervention, and outcome measures. When authors conduct a qualitative analysis of these variables among the included studies, it is called an assessment of clinical heterogeneity.

Meta-analyses should only be used to pool study results when the studies can be justified as having clinical homogeneity in patient characteristics, interventions, and outcomes. In the systematic review of motor control exercises for patients with nonspecific low back pain by Macedo et al,17 14 RCTs met the inclusion and exclusion criteria. However, those studies varied widely with regard to the interventions studied and the outcomes measured. Therefore, several different meta-analyses were conducted on subsets of studies within the systematic review. For example, seven studies were included in a meta-analyses contrasting motor control exercise with minimal intervention. Among those studies, three studies were pooled to assess the impact of motor control exercises on short-term (<3 mo) pain relief. A fourth study was added to the meta-analysis to assess the impact of disability on intermediateterm (>3 and <12 mo) disability. You should appraise the process used for qualitative analysis of heterogeneity prior to pooling data for meta-analyses.

QUESTION 10: If a meta-analysis was conducted, was statistical heterogeneity assessed?

When a meta-analysis is conducted it should be paired with a test of statistical heterogeneity (also called a test of statistical homogeneity). Tests of statistical heterogeneity assess the likelihood that the variability between studies is due to chance. If the result indicates that differences are unlikely to be due to chance (i.e., p < 0.05), the meta-analysis result is called into question and the authors are expected to explore explanation for the differences by further analyzing subgroups of studies.

Different statistical methods are used for meta-analysis depending on the result of heterogeneity testing. For example, in the systematic review by Macedo et al,17 groups of studies with a positive test for heterogeneity were analyzed differently compared with those with a negative test. If the statistical heterogeneity test is positive, a random effects model might be used for the meta-analysis. If the statistical heterogeneity test is negative, a fixed effects model may be more appropriate. As a consumer of systematic reviews, it is important for you to recognize whether or not a test of statistical heterogeneity was conducted. Statistical tests commonly used to assess heterogeneity are described in Table 7.5.1 Interpretation of these tests (beyond their presence or absence) is complex, controversial, and exceeds the scope of this text. For more information we recommend the text by Egger et al.1



Part C: Interpreting Results of a Systematic Review Study

QUESTION 11: What methods were used to report the results of the systematic review?

Interpretation of Qualitative Results

At this point in your appraisal, you have already determined if a meta-analysis was conducted as part of the systematic review and if heterogeneity was considered as part of the analysis. If a meta-analysis was not conducted, the systematic review should provide a qualitative summary of the included studies. In addition, a “vote counting” analysis can also be included.1 Using vote counting, the authors count the number of studies favoring one intervention compared to the number of studies favoring an alternative intervention (or control). Your appraisal of a qualitative summary should include your assessment of the studies given the details provided by the authors. Figure 7.4 illustrates a qualitative summary of the eight studies included in the systematic review by Alexander et al4 regarding ultrasound therapy and soft tissue shoulder pathology.



Interpretation of Quantitative Results

If a meta-analysis was conducted, the results will include a statistical result that allows multiple studies to be expressed in the same units. If the outcome of interest is dichotomous, relative risks and odds ratios can be used to represent each study and the results compared. If the outcome of interest is continuous, the mean difference or effect size (weighted mean difference) between treatment and control groups is most commonly reported.

Forest plots are a graphical representation of the studies included in a systematic review. When a meta-analysis is included in the review, the forest plot is often used to illustrate individual studies and pooled results from the meta-analysis. Consider the forest plot in Figure 7.5 that illustrates three meta analyses comparing motor control exercises to minimal intervention for long-term impact on disability for persons with nonspecific low back pain.17 In the top third of the graph, results from five studies were pooled in a meta-analysis to determine the cumulative effect of motor control exercises versus minimal intervention on short-term disability. The effect size (boxes) and 95% confidence interval (lines) for each comparison are represented.

Data to the left of the vertical line, termed “Line of No effect,” favor motor control exercises. Data to the right of the Line of No Effect favor control interventions. The “Pooled” line represents the results of a meta-analysis, which indicates that overall there was a large effect size (9.6) in favor of motor control exercises. However, because the 95% confidence interval crosses zero (Line of No Effect), there is a greater than 5% chance that the difference in level of disability between patients who received motor control exercises and minimal intervention was due to chance.

In contrast, the meta-analysis for long-term recovery (>12 mo) demonstrates a statistically significant difference between groups. The effect size is –10.8 (the negative sign indicates that the difference favors motor control exercises) and the 95% confidence interval ranges from –18.7 to –2.8. We can determine that there was a statistically significant difference in the pooled effect because the confidence interval is to the left of the Line of No effect. Within the top box (outlined in blue) in Figure 7.5, the study authors illustrate the result of a meta-analysis for the impact of motor control exercise compared to minimal intervention in the short-term (<3 mo).

1. Each of six studies included in the analysis are listed, followed by “Pooled,” which represents the result of the meta-analysis.

2. The small squares represent the mean effect size for each individual study. The squares on one side (left, in this case) represent results that favor motor control exercises; squares on the other side (right, in this case) represent results that favor minimal intervention.

3. The horizontal lines represent the confidence interval for each study’s effect sizes.

4. The diamond represents the mean effect size from the meta-analysis.

5. The vertical line is called the Line of No Effect and represents the point on the scale at which neither treatment is favored over the other. If the horizontal we know that there was not a statistically significant difference between groups.

6. We identify that the result is statistically significant because the 95% Confidence Interval line does not cross the Line of No Effect. Confidence Interval lines (C) cross the Line of No Effect.

Interpreting Forest Plots



Use Figure 7.5 to answer the following questions:

1. Which studies did not find a statistically significant benefit for motor control compared to minimal intervention at the intermediate time point?

2. What was the mean effect size identified by the metaanalysis at the intermediate time point? Is this a large effect size? Was this a statistically significant result?

3. Review the long-term time period. Which primary study had the greatest benefit of motor control exercises compared to minimal intervention?

4. Does this graph support that motor control exercises are preferable to minimal intervention in the Intermediate recovery period? Was this a statistically significant result?

Part D: Summarizing the Clinical Bottom Line of a Systematic Review

QUESTION 12: Does the systematic review inform clinical practice related to my

clinical question?

A high-quality, clinically applicable systematic review provides a comprehensive literature review, appraisal, and synthesis of study results in one document. The culmination of many studies can be more powerful for informing practice than single primary studies. However, because systematic reviews combine studies, there are limitations that must be considered. There is wide variation in intervention and outcome measurements. Pooling data from studies in which interventions were applied with disparate frequency, intensity, duration, and methods of application can be misleading. Likewise, pooling data from studies in which the outcome measures used are widely disparate can also be misleading. You must carefully appraise the

characteristics of the individual studies included in a systematic

review (a summary table of individual studies should be provided) before accepting and applying the results. Many systematic reviews do not provide the necessary details to guide specifics for patient treatment. These reviews can be used to gain insight into the body of literature that relates to your question and to identify primary studies that provide more detail to guide practice. The literature review incorporated in the systematic review can be used to identify high-quality studies and to compare their results. Consider the forest plot in Figure 7.6. The graph illustrates the pooled results of 15 studies that assessed the efficacy of physical therapy interventions on motor and functional outcomes among persons with chronic stroke.18 The pooled result favors physical therapy. However, most of the studies have 95% Confidence Intervals that cross the Line of No Effect (indicating no statistically significant difference between groups). To inform your clinical practice, it might be helpful to read the article by Yang et al,19 published in 2006, to learn more about the methods that produced the largest effect favoring physical therapy over a control. The 12 questions that inform Parts A through D are included in the checklist in Table 7.6.



Ringkasan

Systematic reviews are the principal scientific method for secondary research. There is a difference between narrative reviews and systematic reviews. Narrative reviews are not research studies and provide a nonsystematic summary of research that is at risk for author bias. A systematic review uses a documented and systematic approach to obtain and analyze studies that meet set criteria for answering a clinical question. High-quality systematic reviews provide a transparent and reproducible process for identifying studies, judging their quality, extracting data, and conducting qualitative or quantitative analyses. Systematic reviews reduce the risk of bias inherent in informing practice based on single primary studies. However, as with any study, systematic reviews must be appraised for applicability and quality before they can be used to inform clinical practice.

1. Pertanyaan
2. Jelaskan apa yang dimaksud dengan sensitivity!
3. Jelaskan apa yang dimaksud dengan specificity!
4. Jelaskan apa yang dimaksud dengan receiver operating curve!
5. Jelaskan apa yang dimaksud dengan positive and negative predictive values!
6. Jelaskan apa yang dimaksud dengan positive and negative likelihood ratios!
7. Daftar Pustaka

Linda Fetters, Julie Tilson, Evidence Based Physical Therapy, F.A Davis Company, Philadelphia, 2012), pp 86-100.

Cook CE, Hegedus EJ. Orthopedic Physical Examination Techniques: An Evidence Based Approach. Upper Saddle River, NJ: Pearson/Prentice Hall; 2008.

Campo M, Shiyko MP, Lichtman SW. Sensitivity and specificity: a review of related statistics and controversies in the context of physical therapist education. J Phys Ther. Educ. 2010;24:69-78.