Systematic Review and Network Meta-Analysis in Health Technology Assessment

Nathorn Chaiyakunapruk PharmD, PhD*,**,***,****, Surasak Saokaew PharmD*,*****, Rosarin Sruamsiri BPharm, MBA*,******, Piyameth Dilokthornsakul PharmD*,*******

** Center of Pharmaceutical Outcomes Research, Department of Pharmacy Practice, Faculty of Pharmaceutical Sciences, Naresuan University, Phitsanulok, Thailand*
*** School of Population Health, Public Health Building, University of Queensland, Australia*
**** School of Pharmacy, Monash University Malaysia, Malaysia*
***** School of Pharmacy, University of Wisconsin, Madison, WI, USA*
****** Center of Health Outcomes Research and Therapeutic Safety (Cohorts), School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand*
******* Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Healthcare Institute, MA, USA*
******** Center for Pharmacoepidemiology and Pharmacoeconomic Research and Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, USA*

Conducting systematic review and meta-analysis (SR/MA) is a standard process for establishing evidences for health technology assessment. Quality assessment of studies included in SR/MA and SR/MA studies should be considered. This article provides recommendations on tools used for assessing the quality of studies included in each SR/MA and the quality of SR/MA. For assessing the quality of randomized controlled trial, we recommend a tool called "Risk of Bias", which focuses on random generation, allocation concealment, blinding and outcome reporting. For assessing the quality of observational study, the Newcastle Ottawa Scale (NOS) is recommended. The NOS consists of three different dimensions-selection, comparability, and outcomes or exposure. Another tool which is recommended is the Down and Black scale. It focuses on the quality of reporting, validity, bias and confounding, and power of study. For assessing the quality of SR/MA, we recommend to use a checklist developed by Klassen et al, covering well-defined question, inclusion criteria, comprehensiveness, quality of included studies, reproducibility, and external validity. This article also provides a fundamental of network meta-analysis that should be considered where no direct evidence exists or when there is a need to compare multiple interventions at the same time.

Keywords: *Systematic review, Meta-analysis, Network meta-analysis, Quality assessment, Health technology assessment*

J Med Assoc Thai 2014; 97 (Suppl. 5): S33-S42
Full text. e-Journal: *http://www.jmatonline.com*

An overview of the fundamentals of systematic review/meta-analysis (SR/MA) and an explanation of how they apply to health economic evaluation was provided in the first edition of Health Technology Assessment Guidelines for Thailand. In this article, the authors will expand on this by exploring a number of other topics related to quality assessment, including several recommendations for assessing both the quality of SR/MA and the quality of the studies that are included in each SR/MA. It is hoped that these recommendations will enable the users of these guidelines to gain a deeper understanding of the quality aspects of SR/MA. In addition, we will give a detailed description of a more advanced topic-network meta-analysis. The article will conclude with a set of clear recommendations for the second edition of Health Technology Assessment Guidelines for Thailand.

Correspondence to:
Chaiyakunapruk N, School of Pharmacy, Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor, Malaysia.
Phone: (+603) 5514-4413, Fax: (+603) 5514-6326
E-mail: nathorn.chaiyakunapruk@monash.edu

Quality assessment of studies for SR/MA

One of the most important processes when conducting a SR/MA is the undertaking of a quality assessment of the studies included in the review. While the first version of Health Technology Assessment Guideline of Thailand did refer to quality assessment,

only a brief overview of the fundamentals was provided, along with an introduction to the use of JADAD, a well-known method for evaluating the quality of randomised controlled trials (RCTs). In this article, the authors will provide an in-depth explanation of quality assessment for both RCTs and observational studies.

Quality assessment of RCTs

One tool that can be used to assess the quality of RCTs is the composite quality scale, which provides an overall quality score for the entire RCT. The JADAD scale is a very widely used composite quality scale, giving values ranging from 0 to 5, in which any study with a score of 3 or over is considered to be of good quality (full details of this approach are given in the first version of Health Technology Assessment Guidelines for Thailand). The JADAD approach suffers from several limitations such as unfair assessment on blinding and the lack of allocation concealment assessment, both of which lead to questionable results interpretation during the application process.

Recently, the Cochrane Collaboration Working Group proposed an alternative approach to assess the quality of studies. Known as the 'risk of bias' approach[1], it is widely known for its high acceptance, coverage of issues for evaluation, and its clear process and criteria for evaluation. Risk of bias considers the following seven domains of a given study: 1) random sequence generation, 2) allocation concealment, 3) blinding of participants or personnel, 4) blinding of outcome assessment, 5) incomplete outcome data, 6) selective outcome reporting, and 7) other sources of bias. Each domain is assessed for risk of bias and categorized according to one of three levels: high risk of bias, unclear and low risk of bias.

The Cochrane Collaboration Working Group recommends that every study should present the level of risk of bias for each study domain (as shown in Table 1). In addition, the overall proportion of studies with high, unclear and low risk of bias should be graphically presented for each domain, as shown in Fig. 1.

To provide an overall assessment of study quality using a risk of bias tool, researchers have to identify the key area(s) for each study. In any one study, it is possible to have one or more key areas, depending on the subject and objective of study. For example, in studies of pain management, correctly implemented blinding of subjects is crucial, since this can affect the result of study and result in bias; in pain management studies, therefore, researchers are likely to deem

Table 1. Example of tabulation of studies using risk of bias tool (Adapted from Higgins et al[1])

Study	Random sequence allocation	Allocation concealment	Blinding of participants or personnel	Blind of outcome assessment	Incomplete outcome data	Selective outcome reporting	Other sources of bias
Study A (2010)	+	-	+	+	+	+	+
Study B (2011)	+	?	+	+	+	-	+
Study C (2011)	+	?	+	?	+	+	+
Study D (2011)	+	+	-	+	-	?	+
Study E (2009)	+	?	?	-	-	?	+

+ = denotes a study with low risk of bias in the domain
- = denotes a study with high risk of bias in the domain
? = denotes a study with unclear risk of bias in the domain

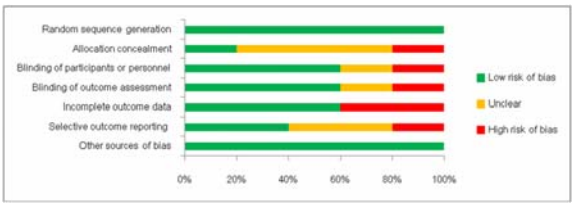

Fig. 1 Example of graphical presentation of results using risk of bias tool (Adapted from Higgins et al[1]).

blinding as a key domain. Generally speaking, the two domains that most affect the quality of RCTs in a meta-analysis are random sequence generation and allocation concealment.

A summary of key recommendations for evaluating overall quality of individual studies in an SR/MA are given below:

- Studies are considered to have a low risk of bias when all key domains are evaluated as low risk of bias. In this case, the SR/MAs are more likely to have low risk of bias.

- Studies are considered as an unclear risk of bias when the risk of bias of at least one part of the key domain is evaluated to be unclear. In this case, the meta-analysis might have bias that may be affected by the individual study.

- Studies are considered to have a high risk of bias when one of the key domains indicates a high risk of bias. The meta-analysis will be seriously affected by the inclusion of any studies that have a high risk of bias.

During the process of conducting a SR/MA, researchers must consider the results of the quality assessment of each individual study within the context of the overall analysis. For example, researchers might decide to include only studies with a low risk of bias or they may decide to include studies with both low and unclear risk of bias. Another option that researchers can use to assess the robustness of the pooled findings is a sensitivity analysis, which can be conducted to identify and exclude studies with a high level of bias risk. Researchers may choose to perform a subgroup analysis by pooling studies according to the risk of bias, or they may decide to pool all studies without considering the risk of bias and address the issue of risk of bias in the discussion section. Another important step that is used to minimize bias in an SR/MA is the incorporation of quality assessment into an analysis a priori. An assessment of the quality of individual studies is clearly an important part of the performance and interpretation of any SR/MA.

Quality assessment of observational studies

SR/MA can be used not only with RCTs, but also with observational studies. However, the process of quality assessment of observational studies is a somewhat different to that used to evaluate RCTs, due to the fact that observational studies tend to have a high risk of bias as a result of their low internal validity. This means that quality assessment is especially important for observational studies. The Cochrane Collaboration Working Group recommends two instruments to assess the quality of observational studies-the Newcastle Ottawa Scale (NOS)[2] and Downs and Black[3].

The NOS[2] is a checklist that is used to assess the quality of observational studies; once applied, the checklist results in an overall score for each study. There are two versions of the NOS, according to the type of observational study under assessment-one for cohort study and one for case-control study. The NOS consists of eight items that can be categorised in three different dimensions-selection, comparability, and outcomes (for cohort study) or exposure (for case-control study). The quality assessment of NOS uses a star system to rate quality, based on a semi-quantitative assessment. An NOS score can range from 0 to 9 stars. The highest score for each individual topic is 1 star. Unlike other domains, 2 stars are given as the highest score for comparability domain.

Downs and Black is another tool that has been used in the quality assessment of observational studies. In addition, a checklist, the Downs and Black consists of 27 items that can be used to assess both randomised and non-randomised studies. The main focus areas that the Downs and Black assesses for quality are reporting, external validity, internal validity covering bias and confounding, and power of study. The highest possible score is 28. The Downs and Black tool suffers from several limitations, including that it is time consuming and can only be conducted by those with a solid background in epidemiology. Furthermore, it is difficult to use Downs and Black to evaluate case-control study. It is clear that both tools are different in terms of their assessment items and their application. To date, no clear guidelines have yet been provided by the Cochrane Working Group on how to use these tools. When conducting a meta-analysis, researchers may consider using the quality evaluation findings from these tools, as they would use quality assessment when conducting a meta-analysis of RCTs. In this case, researchers may include only observational studies with a high score (such as a score higher than four

when using NOS). Furthermore, researchers can conduct subgroup analyses categorised by the quality of the study or choose to present their overall findings and address the issue of quality in the discussion section.

In conclusion, assessing the quality of studies is a crucial part of any SR/MA. Each piece included in the present study should be evaluated with a standard, internationally recognized tool. Moreover, researchers should use the quality assessment component of the SR/MA, in conjunction with the results obtained during the analysis and interpretation process.

Quality assessment of SR/MA

The results obtained from an SR/MA are one of the most important input parameters for Health Technology Assessment (HTA); HTA researchers have to read and critically appraise all relevant SR/MA studies prior to using them in an HTA. By modifying the set of questions first developed by Oxman et al[4] to evaluate SR/MAs, Klassen et al[5] developed a total of six questions that HTA researchers should apply when deciding whether to use the data from a particular SR/MA.

Question 1: Did the SR/MA address a focused, well-defined question?

Research questions should be well defined in terms of the population, intervention, comparator, and outcome. All research questions should be defined as part of the protocol development process. To prevent possible bias, researchers must identify both primary and secondary questions. When evaluating the quality of a SR/MA, researchers should check whether the research question or hypothesis has been defined and has examined the primary question.

Question 2: Were the criteria used appropriate to select articles for inclusion?

A focused, well-defined question will lead to clear inclusion criteria. To reduce bias when deciding which studies to include in an SR/MA, it is essential that the characteristics of the studies be clearly defined. Researchers should define population, intervention, comparator, and outcome (PICO). In addition, SR/MA may specify inclusion criteria in terms of study design, study period, languages, and whether a study is published or unpublished (for unpublished studies, a clear justification of their inclusion should be provided). It is worth noting that, while broad specification of inclusion criteria for SR/MA increases the ability to generalize the findings of this study, this may also result in too much heterogeneity in the SR/MA. Therefore, the users of the SR/MA may need to critically evaluate whether their inclusion criteria are too broad, rendering their findings subject to heterogeneity.

Question 3: Is it likely that important and/or relevant studies were missed?

It is important to consider whether the study search has been comprehensive enough. A comprehensive search should use at least three databases: PubMed/Medline, EMBASE, and the Cochrane controlled trial registry. Other databases related to the specific topic of the SR/MA should also be included along with any of the studies referenced in the included studies. Other sources, identified while searching in other relevant databases or through individual searches might also be required in some situations.

Question 4: Was the quality of all included studies assessed?

It is essential that the quality of the studies that will be used in the SR/MA be assessed to ensure validity of the findings. Higgins et al[1] recommend that researchers report on the quality of all studies that will be included in the SR/MA and indicate their quality in the conclusion.

Question 5: Was the assessment of studies reproducible?

To ensure that the SR/MA findings are robust across all changes, such as method changes, it is important to demonstrate, by a means of sensitivity analyses, that the findings are robust.

Question 6: Were the study results similar across studies?

When studies with heterogeneity are pooled, the results may be imprecise and invalid. To identify whether there are high levels of heterogeneity in the pooled studies, it is necessary to determine whether the included studies are different. Two standard testing tools are used to conduct this assessment-the Cochrane Chi-square (Q-test) and (I^2). In general, studies are considered to have significant heterogeneity when the p-value according to the Q-test is less than 0.05 or when the I^2 is greater than 50%. The heterogeneity of the included studies may stem from differences in intervention, exposure, outcome, or population.

When heterogeneity exists, researchers

should investigate the potential causes of the difference. Specifically, researchers should look into the characteristics of the population, the nature of the intervention, the exposure, and the outcomes. When interpreting the findings of SR/MA, users should pay attention to whether SR/MA has been tested for heterogeneity. Where heterogeneity exists between studies, SR/MA researchers should report on how they dealt with those differences by, for example, performing a subgroup analysis of those studies.

All users of SR/MAs that include studies with heterogeneity should assess whether researchers have tried to search for the causes of that heterogeneity. Furthermore, researchers should report on how the heterogeneity of the included studies had been handled since results that include heterogeneous studies can be various and imprecise. In the situation where included studies are homogenous, the pooling method can use either a fixed-effects model or a random-effects model[6]. The decision regarding what method will be used to combine studies should be specified a priori. Researchers may decide to use a random-effects model as the primary analysis tool with a fixed-effect model as the secondary analysis tool.

Choosing an appropriate pooling method can be problematic in some situations. For some studies, researchers may decide to use a fixed-effects model when there is no heterogeneity and to use a random-effects model when heterogeneity exists, without attempting to discern the causes of heterogeneity. The use of fixed-effect models is acceptable with homogenous studies, but the use of random-effects model for heterogeneous studies can lead to biased findings. This issue has been described in details in the first version of HTA guidelines for Thailand.

Currently, there are some tools that have been developed for the quality assessment of SR/MA. One such tool is AMSTAR[7,8] (assessment of multiple systematic reviews), which consists of 11 questions that do not differ substantially from those in the SR/MA quality assessment checklist of Klassen et al[5].

In conclusion, it is clear that assessing the quality of studies that are to be included in an SR/MA is a vital part of assuring the quality of the findings. Where the findings of an SR/MA are to be used in an HTA, the SR/MA quality assessment should evaluate research questions, comprehensiveness of research, inclusion criteria, the quality of studies, the robustness of findings, and the level of heterogeneity among the included studies. All of these issues should be considered while using SR/MA for HTA as they might affect the accuracy of the SR/MA findings.

Network meta-analysis

Conducting an SR/MA of RCTs is an accepted standard process for establishing the evidence base for HTA. In some cases, meta-analyses have been used to combine quantitative data using statistical methodology.

The traditional meta-analysis, also known as the pairwise meta-analysis, is used to assess a new intervention by comparing it with standard care. This type of meta-analysis cannot be used to compare more than two interventions nor can it used to compare studies that have different characteristics. Generally, more than two choices of intervention are used in an HTA and comparison is usually made between the new intervention(s) and a placebo or standard care. Very few studies compare all choices in one study.

In 2002, a new tool, coined network meta-analysis, was developed by Lumley[9]. Network meta-analysis is based on the calculation of relative effect of multiple comparisons. Network meta-analysis is considered to be a good method to compare the effectiveness of standard and alternative interventions, especially in HTA.

Principle and types of network meta-analysis

Network meta-analysis applies the principle of a pairwise meta-analysis to create a network of multiple pairwise analyses by calculating the relative effects among different interventions through one or many common comparators. A network meta-analysis needs to be connected by establishing a network among interventions. This could be a simple closed loop (Fig. 2A) or a complicated connected network (Fig. 2B).

Network meta-analysis can be one of two types-indirect comparison (IC) or multiple treatment comparison, also called mixed treatment comparison (MTC) or multiple treatment meta-analyses (MTM). We give a brief summary of these two types below:

1) Indirect comparison: IC is a method that compares the relative effects of at least two

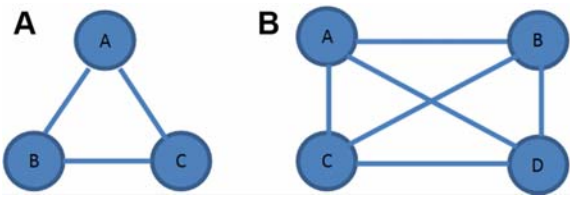

Fig. 2 A) Simple closed loop. B) Connected network.

interventions for which a head-to-head comparison is absent. The relative effects from indirect comparison are considered indirect evidence. For instance, where RCTs exist that compare the efficacy of drug A with placebo C on outcome X and drug B with placebo C on outcome X, but no RCTs exist that compare head-to-head the efficacy of drug A with drug B on outcome X, an indirect comparison could be used to compare the efficacy of drug A and drug B through placebo C, which is the common comparator (Fig. 3A).

Suppose,

$D_{AC\text{-}direct}$ represents the relative effects of drug A compared to placebo C.

$D_{BC\text{-}direct}$ represents the relative effects of drug B compared to placebo C.

$D_{AB\text{-}indirect}$ represents the indirect effects of drug A compared to drug B.

$D_{AB\text{-}indirect} = D_{AC\text{-}direct} - D_{BC\text{-}direct}$ Equation 1

Fig. 3A illustrates the relationship among interventions in the network meta-analysis. Each arrow points to the intervention of interest while the comparator is located on the opposite side without an arrow sign. In this case, placebo C is the common comparator. This type of network is called anchored indirect comparison or adjusted indirect comparison.

2) Multiple treatment comparison: MTC combines the relative effects of interventions using both direct and indirect evidence. An example of a situation where a multiple treatment comparison might be used would be where there are RCTs that compare the efficacy of drug A with placebo C on outcome X and drug B with placebo C on outcome X as well as an RCT that compares the efficacy of drug A with drug B. A multiple treatment comparison will combine direct evidence ($D_{AB\text{-}direct}$) with indirect evidence ($D_{AB\text{-}indirect}$) using a meta-analysis, as shown in Fig. 3B.

Let us suppose the following statements hold: $D_{AB\text{-}direct}$ represents the direct relative effect of drug A compared to drug B and $D_{AB\text{-}indirect}$ represents the indirect relative effect of drug A compared to drug B. (Equation 1)

$W_{AB\text{-}direct}$ is the weighted relative effect of $D_{AB\text{-}direct}$ (calculated by an inverse variance of head-to head study)

$W_{AB\text{-}indirect}$ is the weighted relative effect of $D_{AB\text{-}indirect}$ (calculated by an inverse variance of indirect comparison)

$D_{AB\text{-}pooled}$ is the relative effect derived from multiple treatment comparison.

$$D_{AB\text{-}pooled} = \frac{(W_{AB\text{-}direct} * D_{AB\text{-}direct}) + (W_{AB\text{-}indirect} * D_{AB\text{-}indirect})}{(W_{AB\text{-}direct} + W_{AB\text{-}indirect})}$$ Equation 2

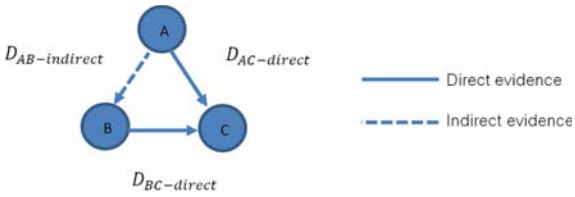

Fig. 3A The network diagram of indirect comparison.

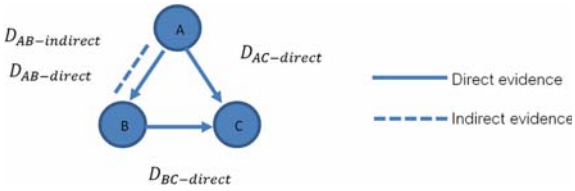

Fig. 3B The network diagram of multiple treatment comparison.

Since multiple treatment meta-analysis combines both direct and indirect evidence, it generates a more precise relative effect and can be used to compare directly the effectiveness of many interventions at the same time. Moreover, this kind of analysis allows all interventions of interest to be ranked.

General assumptions for network meta-analysis

Network meta-analysis is based on a number of assumptions. The first one is the similarity assumption. Studies included in network meta-analysis have to be similar in terms of study design, population, outcomes, outcome measurement, and quality of study. Studies included in network meta-analysis need to be homogenous in both within and between pair wise comparisons. For example, suppose we compare intervention A and B using a network meta-analysis. The patterns of concomitant interventions received in studies comparing A and C should be similar to that in the studies comparing B and C. If different, the patterns of concomitant interventions must not be an effect modifier of relative effects because that could lead to biased relative effects. This similarity assumption is required for both indirect comparison and multiple treatment comparison.

The second assumption is the consistency assumption, also called the coherence assumption, which is required when conducting a multiple treatment comparison. The relative effects from the direct evidence should be similar to those obtained from the indirect evidence. It is important to test this assumption once researchers have performed the multiple treatment comparison. Nowadays, there is no standard method to test the consistency assumption. However, one test

that is widely accepted is the difference between the relative effects of direct and indirect evidence, an illustration of which is shown below.

Suppose the following statements hold:

$D_{AB\text{-direct}}$ represents the direct relative effect of drug A compared to drug B.

$D_{AB\text{-indirect}}$ represents the indirect relative effect of drug A compared to drug B (Equation 1).

Diff is the difference of relative effect between direct and indirect evidences (Equation 3).

δ^2_{Diff} is variance of the difference of relative effect between direct and indirect evidences (Equation 4).

δ^2_{AB} is variance of the relative effect of drug A compared to drug B.

δ^2_{AC} is variance of the relative effect of drug A compared to drug C.

δ^2_{BC} is variance of the relative effect of drug B compared to drug C.

$Diff = D_{AB\text{-direct}} - D_{AB\text{-indirect}}$ Equation 3

$\delta^2_{Diff} = \delta^2_{AB} + \delta^2_{AB} + \delta^2_{AB}$ Equation 4

$Z_{statistic} = \frac{Diff}{\sqrt{\delta^2_{Diff}}}$ Equation 5

Analytical techniques used in network meta-analysis

Both fixed-effects and random-effects models can be used in network meta-analysis. The process for deciding which method is more appropriate is similar to the pairwise meta-analysis outlined in the first edition of HTA guidelines for Thailand. Furthermore, network meta-analysis can be analyzed with either frequentist or Bayesian approaches. With a frequentist approach, the results will be reported as a point estimate with a 95% confidence interval. The 95% confidence interval indicates that if study is repeated 100 times, 95 of the 100 results will fall in the range; frequentist approach findings should not be regarded as representing direct probability.

The Bayesian approach, which combines prior information (prior probability distribution) with current observed data, can also be applied to network meta-analysis, to show the posterior probability distribution. The results generated by application of the Bayesian approach can be interpreted as direct probability, and can be used for ranking the interventions. For these reasons, the Bayesian approach is recommended for use in HTA to inform the decision-making process.

Critical appraisal of network meta-analysis

Three key questions that need to be considered when performing network meta-analysis. The first is whether the studies included for each pairwise meta-analysis are heterogeneous. The second is whether studies across pairs are similar in terms of population, outcomes, and quality of the studies. The third, which is necessary when conducting a multiple treatment comparison, examines whether there is consistency between the direct and indirect evidence.

These three questions capture the key assumptions of network meta-analysis, and represent the key areas of evaluation. Moreover, critical appraisal of network meta-analysis should include standard assessment issues in pairwise meta-analysis, such as publication bias.

Current situation: the usage of network meta-analysis around the world

According to a report published by the Agency for Healthcare Research and Quality in 2012[10], only 25 publications have made recommendations or provided guidance on the use of network meta-analysis. Most of these publications focused on the importance and application of network meta-analysis, and only two provided a detailed methodology and interpretation of their network meta-analysis findings.

In terms of the use of network meta-analysis in the HTA process, in 2008 NICE (the National Institute of Clinical Excellence) recommended that researchers use evidence from head-to-head RCTs as the reference-case analysis in HTA. However, NICE also outlined that researchers may use evidence from network meta-analysis in cases where network meta-analysis findings provide additional important relevant information or where there is no direct evidence that compares multiple interventions in one study.

The Pharmaceutical Benefits Advisory Committee (PBAC)[11] also recommends using evidence from head-to-head RCTs where possible, and evidence from network meta-analysis where no direct evidence exists. CADTH also recommends using indirect evidence where there is no direct evidence[12]. In all cases, guidelines are clear that justification should be provided when network meta-analysis data are used, together with a clear explanation of the methodology. In 2009, CADTH[13] published a set of recommendations for HTA on cancer treatment, recommending that indirect evidence could be used for HTA where no direct comparative evidence existed on the new intervention and appropriate comparators.

In 2011 CADTH[14] published the Common Drug Review Submission Guidelines for Manufacturers,

in which it stipulated that pharmaceutical companies that believe that their products have unique clinical data (including efficacy, effectiveness, and safety) should submit all evidence-based information, full details of their methodology, and all results of the indirect comparison that support this claim.

Cooperd network meta-analysis in HTA revealed that the differences between the relative effects of pairwise and network meta-analyses is unpredictable. That is, there is no current consensus as to whether the results of the pairwise meta-analysis are more precise than those of the network meta-analysis. Evidence from network meta-analysis is increasingly being used for HTA and can often be interpreted as providing an overall summary. It is worth noting, however, that network meta-analysis requires more assumptions and is more complicated than pairwise meta-analysis.

Guidelines for health technology assessment in Thailand (second edition): Recommendations for systematic review and meta-analysis

1. Researchers should consider whether quality assessment for each study in the SR/MA has been performed, and interpret the results accordingly.

2. Researchers should conduct a quality assessment of all SR/MAs, according to the recommendations provided herein.

3. Researchers may wish to consider evidence from network meta-analysis where no direct evidence exists or when there is a need to compare many interventions at the same time.

4. Researchers should continue to comply with the guidelines already provided in the first edition of HTA guideline for Thailand, which are summarized below for ease of reference:

- As a minimum, evidence should be gathered from the following three electronic databases: Medline, EMBASE, and the Cochrane controlled trial registry

- Searching should not be restricted by language

- The inclusion criteria for SR/MA should explicitly specify whether reports, proceedings, or abstract are included; specific justification of the inclusion criteria should also be included.

Acknowledgement

The development of these guidelines would not have been possible without the technical support, challenging criticism and encouragement of many colleagues and institutions. As authors, we wish to acknowledge all individuals and related organizations that contributed throughout the process for guideline development. In addition, we would like to give particular thanks to the funding support through the Health Intervention and Technology Assessment Program (HITAP) from the National Health Security Office, the Thailand Research Fund under the Senior Research Scholar on Health Technology Assessment (RTA5580010) and Thai Health Global Link Initiative Program (TGLIP), supported by Thai Health Promotion Foundation.

Potential conflicts of interest

None.

References

1. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions. Version 5.0.2 [updated September 2009]. Oxford: The Cochrane Collaboration; 2009: Chapter 8.1-17.
2. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses [Internet]. 2012 [cited 2012 Oct 19]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
3. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998; 52: 377-84.
4. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994; 272: 1367-71.
5. Klassen TP, Jadad AR, Moher D. Guides for reading and interpreting systematic reviews: I. Getting started. Arch Pediatr Adolesc Med 1998; 152: 700-4.
6. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. Lancet 1998; 351: 123-7.
7. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007; 7:10.: 10.
8. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the

methodological quality of systematic reviews. J Clin Epidemiol 2009; 62: 1013-20.
9. Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002; 21: 2313-24.
10. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. Pharmacoeconomics 2008; 26: 753-67.
11. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50: 683-91.
12. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Assoc 2006; 101: 447-59.
13. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29: 932-44.
14. Mills EJ, Ioannidis JP, Thorlund K, Schunemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. JAMA 2012; 308: 1246-53.
15. Coleman CI, Phung OJ, Cappelleri JC, Baker WL, Kluger J, White CM, et al. Use of mixed treatment comparisons in systematic reviews. AHRQ Publication No. Rockville, MD: Agency for Healthcare Research and Quality; 2012.
16. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011; 14: 417-28.
17. Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value Health 2011; 14: 429-37.
18. Gaitanis A, Staal S. Liposomal doxorubicin and nab-paclitaxel: nanoparticle cancer chemotherapy in current clinical use. Methods Mol Biol 2010; 624:385-92.
19. Desai N. Challenges in development of nanoparticle-based therapeutics. AAPS J 2012; 14: 282-95.
20. Patel TA, Perez EA. Metastatic breast cancer. In: Jatoi I, Kaufmann M, editors. Management of breast diseases. New York: Springer; 2010: 425-49.
21. Dranitsaris G, Coleman R, Gradishar W. nab-Paclitaxel weekly or every 3 weeks compared to standard docetaxel as first-line therapy in patients with metastatic breast cancer: an economic analysis of a prospective randomized trial. Breast Cancer Res Treat 2010; 119: 717-24.
22. Laino C. Paclitaxel outperforms newer agents for metastatic breast cancer. Oncology Times [Internet]. 2012 [cited 2013 Jan 13]; 34: 32-5. Available from: http://journals.lww.com/oncology-times/Citation/2012/10100/Paclitaxel_ Outperforms_Newer_Agents_for_Metastatic.7.aspx
23. Cooper NJ, Peters J, Lai MC, Juni P, Wandel S, Palmer S, et al. How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? Value Health 2011; 14: 371-80.

การทบทวนวรรณกรรมอย่างเป็นระบบและการวิเคราะห์อภิมานเครือข่ายในกระบวนการประเมินเทคโนโลยีด้านสุขภาพ

ณธร ชัยญาคุณาพฤกษ์, สุรศักดิ์ เสาแก้ว, รสรินทร์ สรวมศิริ, ปิยะเมธ ดิลกธรสกุล

การทบทวนวรรณกรรมอย่างเป็นระบบและการวิเคราะห์เชิงอภิมานนับว่าเป็นกระบวนการมาตรฐาน ในการสร้างหลักฐานเชิงประจักษ์สำหรับการประเมินเทคโนโลยีด้านสุขภาพการประเมินคุณภาพของการศึกษา ที่ได้รับการคัดเข้าในการทบทวนวรรณกรรมอย่างเป็นระบบ และการวิเคราะห์เชิงอภิมานและการทบทวนวรรณกรรม อย่างเป็นระบบและการวิเคราะห์เชิงอภิมานเองควรได้รับการพิจารณา บทความนี้นำเสนอแนวทางการประเมินคุณภาพของการศึกษาที่ได้รับการคัดเข้าในการทบทวนวรรณกรรมอย่างเป็นระบบ และการวิเคราะห์เชิงอภิมานและการประเมินการทบทวนวรรณกรรมอย่างเป็นระบบและการวิเคราะห์เชิงอภิมาน สำหรับการศึกษาที่ใช้รูปแบบทดลองโดยมีการสุ่มและควบคุม ผู้นิพนธ์แนะนำให้ใช้ “Risk of Bias” ซึ่งเป็นเครื่องมือสำหรับการประเมินคุณภาพที่ครอบคลุมการประเมิน ด้านกระบวนการสุ่ม กระบวนการปกปิดผลการสุ่ม การปกปิดการได้รับสิ่งทดลองและการรายงานผลการทดลอง สำหรับการศึกษาเชิงสังเกต ผู้นิพนธ์แนะนำให้ใช้ “Newcastle-Ottawa scale” ซึ่งเป็นเครื่องมือที่ประเมินคุณภาพจากข้อมูล 3 ส่วนหลักๆ ได้แก่ การคัดเลือกผู้ป่วยเข้าการศึกษา การเปรียบเทียบของการรักษาที่ได้รับผลลัพธ์ของการศึกษาหรือการรักษาที่ได้รับ อีกเครื่องมือหนึ่งที่ผู้นิพนธ์แนะนำให้ใช้สำหรับการประเมินการศึกษาเชิงสังเกตคือ “Down and Black scale” ซึ่งเป็นเครื่องมือที่ประเมินคุณภาพของการรายงานผลการศึกษา ความเที่ยงตรงของผลการศึกษา อคติ ตัวแปรกวน และอำนาจการจำแนกของการศึกษา ส่วนการประเมินคุณภาพของการทบทวนวรรณกรรมอย่างเป็นระบบและการวิเคราะห์เชิงอภิมานนั้น ผู้นิพนธ์แนะนำให้ใช้เครื่องที่นำเสนอโดย Klassen และคณะ ซึ่งครอบคลุมในส่วนของ การตั้งคำถามที่เหมาะสม การคัดเลือกการศึกษา ความครอบคลุมของการสืบค้นคุณภาพของการศึกษาที่คัดเข้าในการทบทวนวรรณกรรม ความสามารถในการทำซ้ำ และความเที่ยงตรงภายนอก นอกจากนี้บทความนี้ยังนำเสนอพื้นฐานของ การวิเคราะห์อภิมานเครือข่าย ซึ่งควรถูกพิจารณานำมาใช้ โดยเฉพาะอย่างยิ่งในกรณีที่ไม่มีหลักฐานเชิงประจักษ์โดยตรง หรือว่ามีความต้องการที่จะเปรียบเทียบทางเลือกหลายๆ ทางพร้อมๆ กัน