Measurement of Health Outcomes

Montarat Thavorncharoensap BSc, PhD*,**

** Social and Administrative Pharmacy Excellence Research (SAPER) Unit, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand*
*** Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand*

Health outcomes are one of the most important components of health technology assessments (HTAs). All HTA outcomes should be measured from a relevant sample using a properly designed study and method. A number of recommendations on health outcome measurements are made in this second edition of Thailand's HTA guidelines. In particular, the use of final outcomes, rather than surrogate outcomes, in HTAs is stressed. Where surrogate outcomes are used, strong justification and evidence must be provided. Effectiveness is preferred over efficacy. The relative treatment effect (the difference between health outcome that would be experienced by patients receiving the technology and that experienced by the same group were they to receive an alternative technology) should be derived from a systematic review of head-to-head RCTs. Mixed treatment comparison (MTC) should be used only to provide supplementary data that cannot be obtained from a head-to-head comparison. Where no direct comparison evidence exists, indirect comparison and observational study data can be used.

Keywords: *Effectiveness, Efficacy, Outcome, Relative treatment effect*

J Med Assoc Thai 2014; 97 (Suppl. 5): S27-S32
Full text. e-Journal: *http://www.jmatonline.com*

One of the most basic aims of a health technology assessment (HTA) is to analyze whether the resources required to implement a given health technology (i.e. the costs) are worth the corresponding outcomes. As a result, ensuring that outcomes are accurately measured is an essential part of garnering valid, relevant, and reliable HTA results. In any HTA, the health outcome data should be taken for study and adhering to the recognized guidelines on study design, sample, and methodology.

This article explores three important issues related to health outcome measures: 1) a comparison of surrogate and final outcomes, 2) a comparison of efficacy and effectiveness, and 3) a discussion of the source of relative treatment effect. Finally, a number of key recommendations on health outcome measurements are given, as well as guidelines on the sources of baseline clinical data that are appropriate for accurate and valid HTAs.

Surrogate vs. Final outcome

The health outcomes of a given technology include both the health benefits and drawbacks that result from the technology. In general, these outcomes are then categorized as either surrogate outcomes or final outcome. A surrogate outcome is "a laboratory measurement or a physical sign that can be used as a substitute for a final outcome, which is a clinical meaningful outcome that measures directly how a patient feels, functions, or survives"[1]. Examples of surrogate outcomes include blood pressure, CD-4 cell count, and bone mineral density. Examples of final outcomes include cardiovascular death, fractures, Life Years Gained (LYGs), and Quality Adjusted Life Years (QALYs) gained. In many countries, the QALY is the recommended final outcome measurement for HTAs[2]. For further details on QALY and its calculation, please see the relevant article in this journal.

Surrogate outcomes are often, mistakenly, used in clinical research and HTAs. The rationale that has been used to support the use of surrogate outcomes is that improvement in surrogate outcomes can often serve as an indicator of final outcome improvement. For example, a reduction in blood pressure has been associated with reduced risk of cardiovascular disease related mortality, an increase in CD4 cell count in AIDS patientshas been associated with reduced mortality, and increased bone mineral density has been associated with a reduction in the rate of bone fractures. However, while surrogate outcomes can provide useful data on the benefits of a

Correspondence to:
Thavorncharoensup M, Social and Administrative Pharmacy Excellence Research (SAPER) Unit, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, 447 Sri-Ayudthaya Road, Payathai, Ratchathevee, Bangkok 10400, Thailand.
Phone: 0-2644-8677, Fax: 0-2644-8694
E-mail: montarat.t@hitap.net, montarat.tha@mahidol.ac.th

health technology, they are not appropriate for use as the sole indicator of the final outcome of a given technology. There have been many studies where surrogate outcome data has indicated the benefit of a health intervention, while final outcome data has shown no benefit at all or has even indicated that the intervention is harmful. For instance, in one study, a drug that was expected to reduce bone fractures because it increased bone density in patients was later found to be associated with an increase in bone fractures[3,4]. In another study, the level of CD4-cell count was not always found to be a good predictor of AIDS-related deaths[5].

Clearly, the sole use of surrogate outcome data should be avoided in HTAs, as it is an inappropriate indicator of the benefits of a health technology. Moreover, surrogate outcomes are not appropriate for assessment of a given technology because, unlike final outcomes, they do not represent the real health benefit for the patient to whom the intervention is aimed, a fact which is often overlooked. For instance, a lower cholesterol level in and of itself is not the ultimate reason patients take antilipidemia. Rather, they take it to reduce their risk of death, by way of cholesterol reduction. This distinction, which is often overlooked, is a crucial one, and means that surrogate outcomes are unsuitable for cost per change calculations. For example, in an HTA assessing the health benefits of an antilipidemia drug, the cost per 1 mg/dL reduction in the level of LDL-C is not an appropriate measurement for assessing the benefit of the intervention, and thus should not be used to inform policy. Instead, the use of cost per life year gained or cost per QALY gained to measure the outcomes of a given technology is much more rigorous.

The International Society for Pharmaco-economics and Outcomes Research (ISPOR), which collates pharmaco-economic guidelines from 32 countries around the world found that 19 countries explicitly state that final outcome should be used in HTA, with surrogate out comes only to be cited as evidence where there is also a proven association between a surrogate and patient-important outcome[6]. The remaining guidelines do not state a preference, but none recommend the use of surrogate outcomes of over final outcomes.

Effectiveness vs. Efficacy

Health outcomes can be measured in terms of both efficacy and effectiveness. Efficacy assesses the outcomes of health technologies, as measured under ideal conditions[7], often in randomized control trials when strict selection process, randomization, blinding, and intensive monitoring/follow-up are being employed. The majority of these studies are conducted in institutions where the health care providers are specialists in the field and where all necessary equipment was in place. Given that these ideal conditions are often not replicated in real situations, effectiveness is measured to assess the outcomes of a technology in more realistic contexts (i.e. routine clinical practice)[7].

According to ISPOR, 19 pharmaco-economic guidelines around the world recommend that health outcomes in HTAs should be measured in terms of their effectiveness rather than their efficacy, as this is more indicative of the context within which an intervention is likely to deployed[6]. If efficacy data are to be used, they should be adjusted to reflect the effectiveness data by incorporating parameters related to adherence, sensitivity or specificity of diagnostic testing, coverage rate, and health professional skill etc. into the present study design[8].

Relative treatment effect

The relative treatment effect represents the difference between the health outcomes experienced by patients receiving the technology and those experienced by the same group were they to receive an alternative technology. Ideally, the relative treatment effect should be derived from systematic data review, based on high internal validity and external validity Randomized Control Trials (RCTs). However, if the available RCT data are inappropriate (for instance, if only studies with a short time horizon are available or if the population samples used in each study are non-comparable), data from high quality observational study can be also used as a supplement.

According to ISPOR[6], 28 pharmaco-economic guidelines suggest that relative treatment effects should be derived from systematic review data. Indeed, eleven national guidelines make specific recommendations that this systematic review should involve a comprehensive and systematic analysis of both published and unpublished studies. More details on the role of a systematic review and meta-analysis in HTA can be found in the first edition of the Thai HTA guidelines and in this volume.

In their 2005 study, Cooper et al[9] proposed a hierarchy of data sources that are appropriate for use in determining the relative treatment effect. In their study[9], meta-analysis of RCT with direct comparison

between comparator therapies, measuring final outcome was at the top of the hierarchy. On the other hand, expert opinion was at the lowest of the hierarchy.

While direct comparative data are preferred when determining relative treatment effect, this kind of data does not always exist. Where this is the case, indirect comparison meta-analyses may be used. Examples of how indirect comparison works are shown in Fig. 1A, where no direct comparison between treatment A and B exists and so treatment A is indirectly compared to treatment B by examining each of them with common comparator treatment C. However, if both direct and indirect evidence between treatment A and B exists, then a mixed treatment comparison (MTC) can be performed by using evidence from both direct and indirect comparison, as shown in Fig. 1B[10].

The benefits of using indirect comparison to estimate relative treatment effect have been widely recognized. This approach, while not as reliable as a meta-analysis of RCT with direct comparison between comparator therapies that measures final outcome, nonetheless allow the estimation of relative treatment effect when there is no head-to head comparison RCT data, without breaking randomization[10]. In addition, MTCs also allow the inclusion of all evidence, which reduces the uncertainty[10]. As a result, evidence from indirect comparisons is placed above that derived from observational studies but below head-to-head RCTs[11,12]. Despite this, the Indirect Comparisons Working Group (ICWG)[13] has warned that randomization may not be preserved in indirect comparison; this may mean that indirect comparison does not always offer a clear advantage over traditional observation, where known confounders can always be adjusted for (no confounder adjustment can be made in indirect comparison). Nevertheless, the current Pharmaceutical Benefits Advisory Committee (PBAC) submission guidelines still recommend indirect-comparison data over observational study data due to their pragmatic, rather than methodological, advantages[13].

In the UK, the National Institute for Health and Clinical Excellence (NICE)[14] recommends the use of data derived from a synthesis of head-to-head RCTs. When head-to-head evidence does exist, evidence from mixed treatment comparison analyses may be presented as supplementary data, only where it adds information not available from the head-to-head comparison. However, they do suggest that MTCs may be useful where a number of treatments are being considered simultaneously and where no RCTs

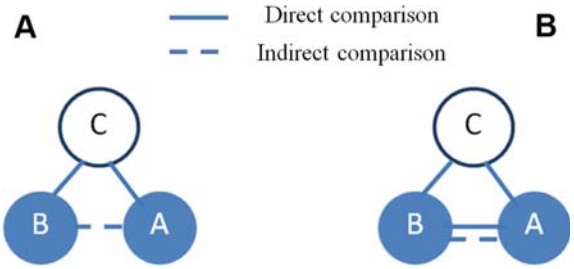

Fig. 1 A) Indirect comparison. B) Mixed treatment comparison.

comparing all of these treatments simultaneously "exist". In this case, NICE suggests the use of data from a series of pair-wise head-to-head RCTs and from MTCs may be helpful. They recommend indirect comparison when no head-to-head RCT exists, and suggest that a principle of good practice for standard meta-analysis be followed in all mixed and indirect treatment comparisons. NICE warns against the use of comparative results from single treatment arms from different trials (known as "naive indirect comparison"). More details on indirect comparisons and MTCs can be found in this volume.

In common with the UK's NICE guidelines, the guidelines of the Pharmaceutical Benefits Advisory Committee (PBAC)[15] of Australia and the Canadian Agency for Drugs and Technologies in Health (CADTH)[8] also recommend data derived from a synthesis of head-to-head comparative RCTs for determining relative treatment effect. The PBAC[15], suggests that, where no head-to-head RCTs exist, evidence derived from indirect comparison RCTs should be used, followed by data from observational studies. All indirect comparison analyses should be properly conducted and the details of the methodology should be clearly provided.

Baseline clinical data

Baseline clinical data are another important parameters used in HTAs. In their 2005 study, Cooper et al[9] proposed a hierarchy of data sources that are appropriate for baseline clinical data. Based on their recommendation[9], baseline clinical data obtained from case series of analysis of reliable administrative databases specifically conducted for the study covering patients solely from the jurisdiction of interest was at the top of the hierarchy, followed by those obtained from recent case series or analysis of reliable administrative databases covering patients solely from the jurisdiction of interest, and those obtained from

recent case series of analysis of reliable administrative databases covering patients solely from another jurisdiction, respectively. On the other hand, baseline clinical data that derived from expert opinion was placed at the lowest of the hierarchy.

Guidelines for health technology assessment in Thailand (second edition): Recommendations for measuring health outcomes

1. Where possible, relative treatment effect should be derived from a synthesis of evidence taken from head-to-head RCTs. In this case, MTC data can be used only as a supplement, to provide additional information that cannot be obtained through direct comparison. Where no head-to-head RCTs exist, evidence from indirect comparison studies, followed by data from observational studies is recommended. Where several different health technologies are being assessed simultaneously, and no RCTs that compare all of these treatments simultaneously exist, a series of pair-wise head-to-head RCTs should be used; this can be complemented by MTC data, where appropriate.

2. Final outcome is preferred. The use of surrogate outcome should be avoided. Where surrogate outcomes are used, justification should be provided along with clear evidence on the relationship between the surrogate outcome and the final outcome.

3. Effectiveness is preferred over efficacy.

4. The meta-analysis including indirect comparison and the MTC should be properly performed. Details should be clearly provided. Naive comparison is not acceptable.

5. The hierarchy of recommended sources for determining the relative treatment effect, recommended by the second edition of Thai HTA guidelines is shown in Table 1, while the hierarchy of evidence for baseline clinical data is similar to those recommended by Cooper et al[9].

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 author, I wish to acknowledge all individuals and related organizations that contributed throughout the process for guideline development. In addition, I 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. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. JAMA 1999; 282: 771-8.
2. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Pharmacoeconomic guidelines around the World [Internet]. 2012 [cited 2012 Sep 1]. Available from: http://www.ispor.org/peguidelines/index.asp
3. Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, Muhs JM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. N Engl J Med 1990; 322: 802-9.

Table 1. Hierarchy of evidence for relative treatment effect for Thai's HTA guideline

1+	Meta-analysis of RCT with direct comparison between comparator therapies, measuring final outcome
1	Single RCT with direct comparison between comparator therapies, measuring final outcome
2+	Meta-analysis of RCT with direct comparison between comparator therapies, measuring surrogate outcome
2	Single RCT with direct comparison between comparator therapies, measuring surrogate outcome
3+	Indirect comparison, measuring final outcome
3	Indirect comparison, measuring surrogate outcome
4	Case control or cohort studies
5	Non-analytic studies, for example, case reports, case series
6	Expert opinion

4. Riggs BL, Seeman E, Hodgson SF, Taves DR, O'Fallon WM. Effect of the fluoride/calcium regimen on vertebral fracture occurrence in postmenopausal osteoporosis. Comparison with conventional therapy. N Engl J Med 1982; 306: 446-50.
5. Sande MA, Carpenter CC, Cobbs CG, Holmes KK, Sanford JP. Antiretroviral therapy for adult HIV-infected patients. Recommendations from a state-of-the-art conference. National Institute of Allergy and Infectious Diseases State-of-the-Art Panel on Anti-Retroviral Therapy for Adult HIV-Infected Patients. JAMA 1993; 270: 2583-9.
6. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. New York: Oxford University Press; 2005.
7. Cochrane AL. Effectiveness and efficiency: random reflection on health services. London: Nuffield Provincial Hospitals Trust; 1972.
8. The Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. 3rd ed. Ottawa: CADTH; 2006.
9. Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. J Health Serv Res Policy 2005; 10: 245-50.
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. Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. Health Technol Assess 2005; 9: 1-134, iii-iv.
12. McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. JAMA 1999; 282: 1371-7.
13. Report of the Indirect Comparisons Working Group (ICWG) to the Pharmaceutical benefits advisory committee: assessing indirect comparisons. Canberra: Department of Health and Ageing, Australian Government; 2008.
14. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: NICE; 2008.
15. Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the pharmaceutical benefits advisory committee (version 4.3). Canberra: Department of Health and Ageing, Australian Government; 2008.

การวัดผลลัพธ์ทางสุขภาพ

มนทรัตม์ ถาวรเจริญทรัพย์

ผลลัพธ์ทางสุขภาพเป็นองค์ประกอบที่สำคัญอย่างหนึ่งของการประเมินความคุ้มค่าทางสาธารณสุข ผลลัพธ์ทางสุขภาพที่นำมาใช้ในการประเมินคุณภาพควรได้มาจากการวัดในประชากรที่มีลักษณะใกล้เคียงกับบริบทของการประเมินด้วยรูปแบบและวิธีการที่เหมาะสม ข้อแนะนำสำหรับการวัดผลลัพธ์ทางสุขภาพของคู่มือการประเมินความคุ้มค่าทางสาธารณสุขในประเทศไทยสามารถสรุปย่อดังรายละเอียดต่อไปนี้ ผลลัพธ์ทางสุขภาพที่ใช้ควรเป็นผลลัพธ์สุดท้าย ควรหลีกเลี่ยงการใช้ผลลัพธ์ที่เป็นตัวแทน ในกรณีที่จำเป็นควรมีเหตุผลแสดงถึงความจำเป็นตลอดจนหลักฐานแสดงความสัมพันธ์ของผลลัพธ์ที่เป็นตัวแทนดังกล่าวกับผลลัพธ์สุดท้ายที่ต้องการ ผลลัพธ์ทางสุขภาพที่นำมาใช้ในการประเมินควรเป็นข้อมูลประสิทธิผลมากกว่าประสิทธิศักย์ ผลสัมพัทธ์ของการรักษาหรือความแตกต่างของผลลัพธ์ทางสุขภาพที่ผู้ป่วยได้รับจากทางเลือกต่าง ๆ ที่ต้องการเปรียบเทียบควรได้มาจากการวิจัยแบบทดลองโดยมีการสุ่มและการควบคุมซึ่งทำการ "เปรียบเทียบยา/มาตรการ" ที่ต้องการศึกษาโดยตรง ทั้งนี้ผลจากการเปรียบเทียบทางเลือกหลากหลายควรนำมาเสนอเพื่อประกอบการพิจารณาได้เฉพาะในกรณีที่พิจารณาแล้วว่าจะให้ข้อมูลเพิ่มเติมที่ไม่ได้จากการวิเคราะห์เปรียบเทียบทางตรงเท่านั้น ในกรณีที่ไม่มีการเปรียบเทียบทางตรง ควรพิจารณาใช้ข้อมูลจากการเปรียบเทียบทางอ้อม และการศึกษาเชิงสังเกตตามลำดับ